

**A Psychometric Assessment of Wechsler Short Forms and the Shipley Institute of Living  
Scale in the Estimation of Wechsler Adult Intelligence Scale-III IQ Scores With An  
Aging Sample**

**by**

**Ian Clara**

**A thesis submitted to the University of Manitoba in partial fulfillment of the requirements**

**for the degree of**

**Master of Arts**

**Department of Psychology**

**University of Manitoba**

**Winnipeg, Manitoba**

**(c) September, 2000**



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-56114-3

Canada

**THE UNIVERSITY OF MANITOBA  
FACULTY OF GRADUATE STUDIES  
\*\*\*\*\*  
COPYRIGHT PERMISSION PAGE**

**A Psychometric Assessment of Wechsler Short Forms and the Shipley Institute of Living Scale in the Estimation of Wechsler Adult Intelligence Scale-III IQ Scores with an Aging Sample**

**BY**

**Ian Clara**

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University  
of Manitoba in partial fulfillment of the requirements of the degree  
of  
Master of Arts**

**IAN CLARA © 2000**

**Permission has been granted to the Library of The University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis/practicum and to lend or sell copies of the film, and to Dissertations Abstracts International to publish an abstract of this thesis/practicum.**

**The author reserves other publication rights, and neither this thesis/practicum nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.**

### Acknowledgments

I would like to thank my advisor, Dr. Cam-Loi Huynh, for his generous help and input on this study. I would also like to thank my Committee members, Dr. E. Schludermann and Dr. B. D'Entremont for their insights and comments. I would also like to thank my family, for believing in me and for giving me the chance to pursue my interest in Psychology. Thanks go out to Rene "Mulder" Luna, whose phone messages relieved many tiring days of work. A special thanks goes out to the participants at all of the personal care homes and hospitals, without whose help I could not have completed this study.

### Abstract

Short forms of extensive assessment batteries are desired by clinicians in order to reduce testing time, as well as to reduce the fatigue and frustration that examinees may encounter during a long testing session. Four short-forms commonly associated with the WAIS-R and two conversion formulae commonly used with the Shipley Institute of Living Scale (SILS) were evaluated on their accuracy in estimating full-scale WAIS-III IQ scores on various criteria with a sample of elderly participants. All of the estimation procedures produced mean estimates of Full-Scale IQ (FSIQ) that were comparable to the full-form FSIQ, and all evidenced highly significant correlations with FSIQ. Consistent with previous literature, qualitative classification rates were not very high, compromising the utility of all of the short forms to some degree. When considering non- psychometric factors (e.g., testing time, ease of administration and scoring) and psychometric factors (e.g, reliability, validity), a 4-subtest short form by Silverstein (1982) emerged as the most psychometrically valid, and also as the most clinically useful short form. A 4-subtest short form by Ward et al. (1987) was comparable, but evidenced poor validity. The SILS, while providing accurate mean estimates, was not as highly correlated with FSIQ, produced many estimation and classification errors, and was found to be difficult by many participants. All of the Wechsler-based short forms could mimic the full-form in mapping cognition over the elderly lifespan, but the SILS deviated to a large degree. Trend analysis of the data did not find the linear decrease in cognition that is commonly found in cross-sectional studies with the elderly (Botwinick, 1977). Future research in short forms for the WAIS-III need to be carried out, especially with regards to improving

classification accuracy. Specific questions regarding the WAIS-III and the elderly also need to be addressed (e.g., item difficulty).

**Table of Contents**

Acknowledgments	ii
Abstract	iii
Table of Contents	v
List of Tables	ix
List of Figures	xi
INTRODUCTION	1
Literature Review	3
A History of the Wechsler Batteries of Intelligence	3
Methodological Concerns with Cognitive Testing That Led to Short Form Development	6
Short Forms For The Wechsler Batteries	7
Specific Short Forms of the Wechsler Batteries	16
Comparisons and Criticisms of Some of the Short Forms Based on Subtests	22
Short Forms in the Elderly	27
General Criticisms of Short Forms	30
General Advantages of Short Forms	33
The Shipley Institute of Living Scale (SILS)	35
Cognitive Patterns In The Elderly	40
PRESENT STUDY	44
Research Objectives	44

	<b>WAIS-III Short Forms</b>	vi
Hypotheses		49
<b>METHODOLOGY</b>		<b>54</b>
Participants		54
Research Protocol		55
Statistical Protocol		58
<b>RESULTS</b>		<b>63</b>
Descriptive Statistics of IQ Scores, Administration Times of Forms, and Psychometric Properties of Forms		63
Full-Form and Full-Scale IQ		63
WAIS-R-Based Short-Form Estimates of WAIS-III IQ		63
Linear-Equated Short Form Estimates of WAIS-III IQ		65
SILS Conversion Estimates of WAIS-III IQ		66
Statistical Inferences		67
Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 1		67
Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 2		68
Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 3		71
Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 3a.		74

<b>Comparison of Linear-Equated Short-Form Full-Scale IQ With WAIS-R-Based- Conversion Full-Scale IQ -- Hypotheses 4 and 5</b>	75
<b>General Cognitive Trends in the WAIS-III</b>	78
<b>Age Trends in Cognition -- Hypothesis 6</b>	78
<b>VIQ - PIQ Differences for Full-Form and 4-Subtest Short Forms - Hypothesis 7</b>	91
<b>DISCUSSION</b>	95
<b>The Flynn Effect</b>	108
<b>Education as a Covariate</b>	109
<b>Summation and Recommendations</b>	111
<b>References</b>	116
<b>Appendix A (Description of the Subtests of the Wechsler-Bellevue (W-B), the Wechsler Adult Intelligence Scale and the Wechsler Adult Intelligence Scale - Revised)</b>	132
<b>Appendix B (Description of the new subtests of the Wechsler Adult Intelligence Scale — Version III)</b>	133
<b>Appendix C (Copy of the Shipley Institute of Living Scale)</b>	134
<b>Appendix D (Introductory Letter to Centres and Hospitals)</b>	136
<b>Appendix E (Copy of the Mini-Mental State Examination)</b>	138
<b>Appendix F (Instructions for Administration of the Mini-Mental State Examination)</b>	140

<b>Appendix G (Introductory Letter to Participants)</b>	<b>142</b>
<b>Appendix H (Consent Form)</b>	<b>143</b>
<b>Appendix I (Debriefing Form)</b>	<b>144</b>
<b>Appendix J (Feedback Letter for the Participants)</b>	<b>145</b>
<b>Appendix K (Feedback From for Administrators and Institutions)</b>	<b>146</b>

## List of Tables

Table 1	Means and Standard Deviations, and Administration Times for Full-Form WAIS-III IQ and Wechsler-Based Short Form IQ Estimates	63
Table 2	Means, Standard Deviations, and Classification Percentages of Linear-Equated Short Forms for Estimating WAIS-III Full-Form IQ	66
Table 3	Means, Standard Deviations, Administration Times, and Classification Rates for Shipley IQ Estimates	67
Table 4	Repeated Measures ANCOVA for Wechsler-Based Short Forms With the WAIS-R Conversion Tables	69
Table 5	Repeated Measures ANCOVA for the Linear-Equated Wechsler-Based WAIS-III IQ Estimates with Education Level as Covariate	70
Table 6	Repeated Measures ANCOVA for SILS-Conversion IQ Estimates with Education Level as Covariate	71
Table 7	Classification Accuracy and Misclassification of the WAIS-R-Based IQ Estimation Procedures	72
Table 8	Frequency Table of Absolute Differences Between Full-Form Full-Scale IQ and SILS Conversion Full-Scale IQ Estimates	74

Table 9	<b>T-Tests of the Difference Between the Linear-Equated Full-Scale IQ Estimates and the WAIS-R Conversion Full-Scale IQ Estimates</b>	77
Table 10	<b>Frequency Distribution of Absolute Differences of Linear-Equated Full-Scale IQ Estimates from the Full-Form Full-Scale IQ</b>	78
Table 11	<b>Sample Sizes, Means, and Standard Deviations of Full-Scale IQ, Verbal IQ, Performance IQ, Linear-Equated Full-Scale IQ Estimates, and SILS-Conversion Full-Scale IQ Estimates for Each Age Group</b>	79
Table 12	<b>Verbal and Performance IQ Differences with the Full-Form IQ and Short- Form Wechsler Scales</b>	92
Table 13	<b>Frequency Counts of Absolute VIQ-PIQ Differences for Full-Form and Four-Subtest Short Form Estimates</b>	93
Table 14	<b>Frequency Distribution of Absolute Differences of 4-Subtest Linear-Equated Verbal and Performance IQ Estimates from the Full-Form Verbal and Performance IQ</b>	94

List of Figures

<u>Figure 1.</u>	Plot of Full-Scale Wechsler IQ for each age group.	81
<u>Figure 2.</u>	Plot of Verbal Wechsler IQ for each age group.	82
<u>Figure 3.</u>	Plot of Performance Wechsler IQ for each age group.	83
<u>Figure 4.</u>	Plot of linear-equated IQ estimates for Silverstein's (1982) 2-subtest short-form for each age group.	85
<u>Figure 5.</u>	Plot of linear-equated IQ estimates for Silverstein's (1982) 4-subtest short-form for each age group.	86
<u>Figure 6.</u>	Plot of linear-equated IQ estimates for Ward et al.'s (1987) 4-subtest short-form for each age group.	87
<u>Figure 7.</u>	Plot of linear-equated IQ estimates for Reynolds et al.'s (1983) 4-subtest short-form for each age group.	88
<u>Figure 8.</u>	Plot of IQ estimates for Paulson and Lin's (1970) SILS conversion for each age group.	89
<u>Figure 9.</u>	Plot of IQ estimates for Zachary et al.'s (1985) SILS conversion for each age group.	90

## INTRODUCTION

Wechsler (1958, cited in Groth-Marnat, 1990, p.110; Matarazzo, 1972, p.79) considered intelligence to be a global concept that “involved an individual’s ability to act purposefully, think rationally, and deal effectively with the environment.” As such a global concept, intelligence could only be assessed by sampling abilities from several cognitive domains (Wechsler, 1981). This multifaceted view of intelligence and its assessment has been supported by others (e.g., Schaie & Geiwitz, 1982). Horn and Cattell (1967, cited in Botwinick, 1977, p. 589) and Cohen (1968, cited in Matarazzo, 1972, p.56) have both made the distinction between two types of intelligence, calling them “fluid” and “crystallized” intelligence. Fluid intelligence is “conceived as a general (independent of sensory area), relation-perceiving capacity which is determined by each individual’s unique endowment in cortical, neurological connection count development”; crystallized intelligence “reflects material normally taught in school and which manifests itself in ability tests of vocabulary, synonyms, numerical skills, mechanical knowledge, a well-stocked memory, and even habits of logical reasoning” (Matarazzo, 1972, p. 56). Troll (1982) expands on these definitions, adding that fluid intelligence includes the active processing of information and that crystallized intelligence is knowledge based on experiences. These functional definitions of the facets of intelligence are congruent with the more theoretical orientations of “pragmatic” and “mechanical” cognition, as outlined by Lindenberger and Reischies (1999).

The intelligence and cognitive batteries that have been developed through the years have always aimed towards the goal of giving the assessor an accurate description

of the level of cognitive ability for an individual. The determination of cognitive level plays a key role in psychological assessment. Results from a cognitive battery are used to assess strengths and weaknesses of mental functioning, and can also act as screening criteria for diagnostic purposes. The test instrument represents a method for the assessment and appraisal of an individual, and this assessment can influence the future development of the individual. Conrad (1931) has stated that a general test of intelligence should: (1) be a point scale with a definite zero point, (2) provide scores on scales made up of equal units or should be scales so as to be converted into equal parts, (3) include an adequate number of sample tasks, (4) measure general intelligence as well as specific aspects of intelligence, (5) be reliable and valid, and (6) measure verbal and non-verbal factors separately but equally. This description implies the goal of assessment with psychological tests as being to determine the individual's cognitive potential, and that proper use of an intellectual assessment instrument will ensure that potential is accurately and validly measured.

The Wechsler intelligence batteries for adults (Wechsler-Bellevue, WAIS, WAIS-R, and WAIS-III; Wechsler, 1939, cited in Matarazzo, 1972; Wechsler, 1955, 1981, 1997) are standardized test-batteries composed of a number of subtests, each of which samples a different cognitive domain. The aggregate of scores on these subtests enables an overall IQ score to be generated, which gives an estimate of the general level of the test taker's cognitive functioning (Caruso & Cliff, 1999). The two specific facets of intelligence, traditionally categorized as crystallized and fluid components, can also be derived from the Wechsler batteries. The Verbal scales of the Wechsler test batteries are

associated with crystallized intelligence and the Performance scales of the Wechsler test batteries are associated with fluid intelligence. Given these characterizations of facets of intelligence, the Wechsler test batteries are well suited to the assessment of cognitive functioning, and fulfill most of the criteria set by Conrad (1931).

### Literature Review

#### A History of the Wechsler Batteries of Intelligence.

#### The Wechsler-Bellevue (W-B) and the Wechsler Adult Intelligence Scale

(WAIS). The Wechsler-Bellevue I (W-B) Scale was developed in 1939 by David Wechsler to measure general intelligence (Matarazzo, 1972). The W-B was normed on a nationally- representative US sample of 1750 participants of both genders, with ages ranging from 17 to 69 years. The Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) is a substantial revision of the W-B in terms of item content, and was normed on a US nationally-representative stratified sample of 1700 participants in the same age range as the W-B standardization sample.

Both the W-B and the WAIS are composed of eleven subtests, six of them Verbal and five of them nonverbal (Performance) in nature. The Verbal subtests are: Information, Comprehension, Similarities, Arithmetic, Digits (Forward and Backwards), and Vocabulary. The five Performance subtests are: Picture Completion, Picture Arrangement, Object Assembly, Block Design, and Digit Symbol. The eleven subtests are described in Appendix A. These verbal and nonverbal groups can be administered to yield, respectively, a Verbal, Performance, and Full Scale IQ score. The items for each subtest are arranged from the easiest to the most difficult, with the exception of the

Object Assembly and Digit Symbol subtests. Discontinue rules are used to allow items that would certainly be failed to be skipped, and to thus maintain confidence and motivation for the test taker. For some of the subtests, the very easy items at the beginning of the test can be omitted for those test takers who are assumed to be able to answer these items correctly.

The Wechsler Adult Intelligence Scale - Revised (WAIS-R). The Wechsler Adult Intelligence Scale - Revised (Wechsler, 1981) is a revision of the WAIS, and used a standardization sample of 1880 individuals who were considered to be representative of the 1970 U.S. Census. The WAIS-R contains the same eleven subtests as previous Wechsler batteries, with the replacement of dated items with newer items. The ordering of administration for the items within each of the subtests was modified from the original WAIS to reflect the newer items, as was the scoring of the items. The procedure of ordering items on a subtest from easiest to most difficult was retained, as was the use of discontinue rules. As the subtests of the WAIS-R are different from the WAIS in terms of items and not in terms of overall nature, the description of the subtests for the WAIS given in Appendix A applies.

The WAIS-R provides the three traditional IQ scores: an overall or Full Scale IQ, a Verbal IQ, and a Performance IQ. Split-half reliabilities for the three IQ scales in the WAIS-R are all above .90. The individual subtests have a median reliability of .83, with the highest being for Vocabulary ( $r = .96$ ) and the lowest being for Object Assembly ( $r = .52$ ) (Groth-Mamat, 1990; Wechsler, 1981). Previous factor analytic studies (e.g., Enns & Reddon, 1998) have supported the presence of both single- and two-factor

representations, thus supporting the presence of a general factor of intelligence (the Full-Scale IQ), and the two facets of intelligence (the Verbal and Performance IQ components).

The Wechsler Adult Intelligence Scale - III (WAIS-III). The WAIS-R has recently been revised to give the Wechsler Adult Intelligence Scale III (WAIS-III), released in the summer of 1997. The WAIS-III was standardized using a sample of 2450 individuals that were representative of the 1995 United States Census, ranging from 16 to 89 years of age (Wechsler, 1997). The WAIS-III is comprised of fourteen subtests, with the eleven subtests from the previous Wechsler batteries being retained, and three new subtests added: the Symbol Search subtest adapted from the Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1997), Matrix Reasoning and Letter-Number Sequencing. A description of the original eleven subtests of the WAIS-R are given in Appendix A, with descriptions of the three new subtests given in Appendix B. Similar to its predecessor, the WAIS-R, the items on the subtests of the WAIS-III are arranged in order of increasing difficulty. Discontinue rules are used to avoid examinees experiencing frustration with overly-difficult items.

Administration of the complete WAIS-III takes approximately 75 minutes, with a range of 60-90 minutes. Administration of the first eleven subtests on the WAIS-III allows the test administrator to construct Full Scale, Verbal, and Performance IQ scores. A change from previous Wechsler batteries is the placement of the Object Assembly subtest. It is not used to compute any of the IQ scores, and has been replaced by the Matrix Reasoning subtest. The Object Assembly subtest can be used as a replacement for

a Performance subtest should that subtest become spoiled (e.g., the examinee refuses to complete the subtest), and it is not recommended for use with examinees older than 74 years of age.

Methodological Concerns with Cognitive Testing that Led To Short Form Development.

Criticisms of Full-Form Administration: Time and Assessment Issues. A concern with many clinicians and practitioners is the amount of time required to administer a test that adequately measures the level of cognitive functioning in an individual. Mumpower (1964) reports the amount of case overload and the limitation of available time (also stated by Boone, 1990; Compton, Bachman, & Logan, 1997; Silverstein, 1985, 1990; Ward, Selby, & Clark, 1987) as two driving forces behind the search for shorter tests of intellectual functioning. The concern over testing time may be more relevant for the Wechsler batteries, where complete administration time averages between 60 and 90 minutes (Thompson, 1987).

An attempt to address these concerns has been the development of shorter test forms for measuring intelligence, and most notably for the Wechsler batteries. These shorter tests can have many appearances, being a selected number of subtests of the Wechsler scales or being a completely different test from the Wechsler batteries. Levy (1968) states the main purpose of a short form is to “save testing time” (p. 410), and this sentiment has been echoed by other authors (e.g., Ward et al., 1987). This had been noted even earlier by Cotzin and Gallagher (1950), who desired short forms that had a reduced administration time, could measure intelligence accurately, were easy to score while

remaining objective in that scoring, and which also retained items that had diagnostic significance to the clinician.

Boone (1990) states that, from a clinician's perspective, a short form is acceptable for use when there is only time for a brief administration, when only an estimate of intellectual functioning is required, or when a cognitive deficit needs to be ruled out by the tester. Silverstein (1990a), Hoffman and Nelson (1988), and Groth-Marnat (1990) further state that a short form of an intelligence test is useful as a screening criterion for acceptance into some experimental group. Short forms can also be used with those participants who are unable to tolerate the amount of time needed for the full-scale testing, or unable to tolerate the demands needed to perform the full-scale test (Boone, 1990, 1991).

The Wechsler batteries have become the standard against which many of these short form or alternate test measures have been compared (e.g., Brooker & Cyr, 1986; Crompton, et al., 1997; Reynolds, Wilson, & Clark, 1983; Silverstein, 1985a; Zarantonella, Munley, & Milanovich, 1993). The common goal of all of these alternate measures, and of the Wechsler scales themselves, has been the accurate measuring of intellectual functioning in the population in an efficient and time-effective manner.

#### Short Forms For The Wechsler Batteries.

As the Wechsler batteries are well-standardized and well-researched, attention has naturally turned to the determination of short forms of the Wechsler batteries themselves. Short forms have found a considerable amount of usage in clinical settings. For example, Holmes, Armstrong, Johnson, and Ries (1965), in a study of a neuropsychiatric hospital,

found that over a 12-year period 90% of the Wechsler Adult Intelligence Scales administered utilized a short-form of the Wechsler battery rather than an administration of all the subtests.

Doppelt (1956) has said that a short form seeks to find a “compromise between time and effort and accuracy of prediction” (p. 63). Attaining such a compromise lies in finding an appropriate abbreviated test form. Wechsler (1958, cited in Silverstein, 1990, p.3) has stated that the use of combinations of three or even two subtests from the Wechsler batteries can be used to obtain a measure of IQ for screening purposes and for obtaining estimates of cognitive ability. This sentiment has been echoed by other authors throughout the investigation of numerous short forms for the Wechsler batteries (e.g., Hoffman & Nelson, 1988; Silverstein, 1990a; Watkins, 1986), and has been further extended to detecting cognitive disability. Cotzin and Gallagher (1950) desired a short form that would be applicable to the determination of cognitive capacities in mental defectives, noting that some of the existing short forms were unable to meet the measurement criteria for this population.

How Short Forms for the Wechsler Batteries Can Be Created. Maxwell (1957) has stated that the selection of an abbreviated scale “depends upon the situation in which it must be administered and the purpose for which it is being given” (p. 125). This statement has a much broader meaning than simply choosing random sections of a test. Considerations such as testing time and clinical intent can play a role in the selection of which subtests are chosen (Ryan, Lopez, & Werth, 1998). As the Wechsler batteries are composed of subtests that sample several cognitive domains, the creator of potential short

forms must, along with concerns over testing time and clinical relevance, take into consideration psychometric properties such as validity, reliability, and accuracy of IQ estimation.

Levy (1968) lists five different approaches to the development of a short form within the Wechsler batteries. One is by finding the most valid set of subtests that a clinician is willing to administer which gives the highest correlation with the Full-Scale IQ score. This method may not allow one to sample subtests from both the Verbal and Performance domains of the Wechsler batteries, as it is possible to have only Verbal or only Performance subtests as the most valid subset. A second approach is to use a stratified scale-sampling procedure, where the short form is composed of subtests from both the Verbal and Performance domains having the highest correlation with the Full-Scale IQ score (an approach used by Doppelt (1956) in constructing his four-subtest short form). A third approach is to use item sampling, where items from all of the subtests are sampled to produce a short form that is representative of the full scale but does not contain all of the items of the full scale, a method used by Satz and Mogel (1962), and is sometimes considered to be a split-half method of forming a short form (Vincent, 1979). A fourth approach is to use factor sampling, where selected subtests are used to represent factors established by a factor analysis (for example, a set of subtests that can be shown to load on a memory factor can be used as a test of memory function). A fifth method is to sample subtests according to the needs of the population being tested. For example, if testing an elderly population, considerations for administration time, relevancy of terminology, and physical abilities of the examinee may dictate which subtests are

chosen.

Two more methods, not mentioned by Levy, also exist for constructing a short form of the Wechsler batteries. The sixth method requires looking at the administration times for each subtest, and to choose the most valid short form that takes the lowest amount of time. This method, which has a more psychometric focus, has been used by Ward et al. (1987) to construct their four-subtest short form. These considerations have also been used by Kaufman, Ishikuma, and Kaufman-Packer (1991) in selecting two-, three-, and four-subtest short forms, who additionally considered the amount of time it would take to score each selected subtest, and the subtests placement within the standardized presentation order.

A seventh method of developing a short form is to use a selection procedure to choose the subtests for inclusion into the short form. This procedure was utilized by Jones (1962), who used the Wherry-Doolittle Selection Method (Garrett & Woodworth, 1966) to design a short form. With this procedure subtests are added until the correlation between the short form and the criterion (here the criterion is full scale Wechsler IQ score) reaches a maximum, and the addition of further subtests results in a decrease in the correlation (Garrett & Woodworth, 1966). Jones (1962) notes that this decrease in the correlation cannot occur when parts of a test are being used to predict the whole test, as is the case with the Wechsler battery. The use of this selection method was advocated by Jones due to its analytical nature and freedom from subjectivity.

Criteria for Evaluating Short Forms. In a rather extensive review of the literature on short-forms for the Wechsler scales, Silverstein (1985a, 1990a) reported three criteria

that are generally used when evaluating a short-form. These three criteria were stated originally by Rabin (1943), later by Resnick and Entin (1971), and reviewed further by Silverstein (1985a, 1990a). The first criterion states that the correlation between the IQ scores from the short-form and the IQ scores on the original scale should be high, significant, or both. Silverstein (1985a, 1990a) points out that the inclusion of a significance test of the correlation as part of the first criterion can lead to the instance of a large enough sample making any non-zero correlation statistically significant. Other authors (Carver, 1978; Frick, 1995; Oakes, 1975; Swoyer & Monson, 1975) have recognized the similar fact that, when looking for any difference of a population parameter from the value of zero, a large-enough sample size will virtually guarantee that the difference will reach significance, even if that difference is small. As a result, the strength of corroboration that comes from the use of a significance test is a false strength since the significance can be manufactured merely by increasing the sample size.

The second criterion states that the mean IQ-score estimate from the short-form should not differ significantly from the mean IQ score on the original scale. This can cause similar problems to those cited earlier, since a large-enough sample of participants can make even the smallest difference in scores significant, and this difference does not need to be a large one, if the correlation between the short-form and the original scale is high (Silverstein, 1990a). The intent of devising a short form is so that it will be able to accurately reproduce the Full-Scale IQ score, and an obvious method of attaining this is to construct a short-form that has a high correlation with the original scale. This was one of the methods of short form construction advocated by Levy (1968). As a result, the

second criterion is simply a recapitulation of the first criterion. The first criterion of a high correlation is sure to have already been met if the mean IQ estimates from the short form are significantly close to the Full-Scale IQ scores (Silverstein, 1990a). It is inconceivable that a short form which is either uncorrelated or has a low correlation with the original scale could lead to the mean IQ scores on both forms being close to each other. On the other hand, a high correlation between the short form and the original form could lead to the instance of even small differences between the short form and original scale mean IQs becoming significant, and as such the second criterion has been failed by the short-form. The final conclusion from Silverstein, (1990a) is that, if the first criterion is met by the short-form, then the evaluation of the second criterion becomes meaningless, especially when using large samples.

The third criterion states that the percentage of disagreements between the short-form and the original scale in classifying participants into qualitative IQ categories should not be so high that the usefulness of the short-form is compromised. This criterion was also used by Rabin (1943) as part of the evaluation of his three-subtest short form for the Wechsler-Bellevue Scale. This third criterion can lead to difficulty, since a difference between the estimated and actual IQ scores of a single IQ point may lead to a disagreement in some cases, and in other cases a much larger difference may not (Silverstein, 1985a). One other difficulty is that there is no definitive standard for when classification crosses from being inadequate to adequate.

In his summary evaluation of these three criteria for short-form evaluation, Silverstein (1985a, 1990a) concluded that the first criterion is usually certain to be met, it

makes little difference if the second criterion is met or failed, and the third criterion is almost guaranteed to be failed. In the light of these conclusions, of the three proposed criteria only the first appears to be of any use to researchers, and the second and third criterion need not be considered. Other researchers have echoed these sentiments, but contend that they should still be used so as to provide a conceptual framework for evaluating short forms (Thompson, 1987).

Reliability and Validity of the Short Forms. A different way to assess the accuracy of a short-form is to investigate its reliability and validity in predicting full-form WAIS-R IQ scores, a point raised by Ryan et al. (1998). "Reliability" is the consistency to which a test form can replicate or be consistent in its results, and is measured by a ratio of the variance of the true scores on a test to the total variance of the test scores (Gregory, 1992). Tellegen and Briggs (1967) have provided a formula for computing the reliability of a short form derived from the Wechsler battery, which has been used by Silverstein (1990b) to determine the most reliable subtest-type short forms for the Wechsler batteries.

"Validity" is the extent to which a test measures what it claims to measure, and that inferences made from the test have meaning and are useful (Gregory, 1992). These types of statistics are normally evaluated through the assessment of the correlation coefficient. Silverstein (1965, 1967a) reports that it is fairly common to assume the validation of a short form of an intelligence test like the Wechsler batteries through the evidence of a high correlation between the short form estimate of IQ score and the Full-Scale IQ score. Other authors (e.g., Levy, 1968; McCormack, 1956; Mumpower, 1964),

along with Silverstein (1965), have indicated that the correlation coefficient between the short-form and the full-form IQ scores is not the relevant statistic for deciding the worth of a short form. The correlation between the short-form IQ estimate and the Full-Scale IQ is considered to be spuriously high due to the fact that it is a correlation of a part with a whole. Doppelt (1956) provided an early counter-argument by reporting that this reasoning does not apply since the goal of creating the short form is to predict the whole from the part. Levy (1967, cited in Levy, 1968, p. 412) has stated that the spuriousness in the correlation comes not from the inclusion of the part from the whole, but in the inclusion of the error variance in the estimates in both the short-form and the full scale IQ scores. From these arguments have risen two types of corrections — one for if the short form is to be used in conjunction with the rest of the test battery (McNemar, 1950), and one for if the short form is to be used as a stand-alone measure (Silverstein, 1985a, 1990a). Depending on how the researcher or clinician is intending to use the short form will determine how the correlation coefficient is to be computed. From a research point of view, the only way to determine if the short form is performing adequately is to compare it against some standardized and well-recognized measure, which in this case is the Wechsler battery from which the short forms are constructed, thus implying the promotion of McNemar's (1950) formula. Tellegen and Briggs (1967) proposed a different version of the formula that corrected for the common error variance shared by the short form and the full form.

Scoring Issues for Short Forms. Levy (1968) indicates two ways in which scores from the short forms are used to estimate the IQ score on the full Wechsler battery. One

is by a proration method, where scale scores are multiplied according to the proportion of items taken from that scale. This method was initially used by Rabin (1943) in his development of a short-form for the Wechsler-Bellevue Scale. A second way to attain Full-Scale IQ estimates is through the use of regression equations, which transform the scaled scores from the short-form into a least-squares estimate of the Full-Scale IQ score. Himelstein (1957) compared the scoring method of regression equations against a proration scoring method for the estimation of short-form estimates of WAIS IQ scores. It was found that both methods were acceptable for obtaining estimates in a sample of hospitalized psychiatric patients, and the estimated scores were virtually equally correlated with the actual Full-Scale IQ score. Similar results were found by Whitmyre and Pishkin (1958) in a sample of male psychiatric patients. Crawford, Allan, and Jack (1992) advocated for the use of regression equations due to their increased accuracy, the more appropriate estimation obtained for both extreme scores and for those individuals who fall outside the standardized age range, and also that they are easier for clinicians to use.

Tellegen and Briggs (1967) have suggested the use of deviation quotients that have a mean of 100 and a standard deviation of 15, which Silverstein (1984) referred to as linear scaling, and which is also known as linear equating, and have a form similar to regression equations. Silverstein (1984) has determined that the use of linear scaling methods to obtain estimates of IQ scores are as accurate and valid as using regression equations, and has also presented equations for determining the standard error of estimate for deviation quotients on short forms (Silverstein, 1985b). Engelhart, Eisenstein,

Johnson, and Losonczy (1999) compared linear-equating with proration methods for obtaining IQ estimates, and found that both methods were comparable, but that the linear-equating method performed better when scores were in the extreme ranges.

Specific Short Forms of the Wechsler Batteries.

Silverstein (1990a) reports the first study of a short-form for a Wechsler test battery as being done by Rabin (1943) for the Wechsler-Bellevue scale. Rabin chose particular subtests based on a clinical impression of how well the particular subtests correlated with the original scale (no indication is given of statistics being used in the selection process, and the impression that the reader gets is that it was a subjective judgement) and ease of administration in terms of testing materials required for the subtests. The three subtests chosen by Rabin were all from the Verbal scale. With the continued development of the Wechsler Adult Intelligence battery, as well as developments on the psychometric front regarding validity, reliability, and IQ estimation with short forms, newer short forms have emerged that satisfy both clinical and statistical properties.

Silverstein's (1982) 2-subtest Short Form. Silverstein (1967) presented a 2-subtest short form based on the standardization data of the WAIS, composed of the Vocabulary and Block Design subtests, and provided a table for converting the sum of the age-scaled scores on the two subtests to estimates of full-scale WAIS IQ. Silverstein (1970) later corrected his earlier tables by recognizing that the validity of a short form is spuriously high (Levy, 1968) since "an item which is passed (or failed) on the short form is assumed to have been passed (or failed) on the short form" (Silverstein, 1970, p. 559),

implying that the subtests were perfectly reliable. The Wechsler manual for the WAIS (Wechsler, 1955) indicates that the validities for the two subtests used by Silverstein (1967) actually vary with the particular age group, and Silverstein (1970) subsequently developed corrected regression equations for estimating WAIS IQ from the Vocabulary-Block Design short-form and presented a resulting table giving estimates for each age group in the WAIS. The resulting corrected validity for the two-subtest short-form was .91 and the reliability was .93.

Silverstein (1982, 1985c) redefined this two-subtest short form with the standardization data from the WAIS-R and found that, when this short form was used as a replacement for the full form, reliability and validity values (calculated from the average subtest reliabilities and intercorrelations across age groups given in the WAIS-R manual), were quite high. Cyr and Brooker (1984) also derived this short-form as the most psychometrically valid 2-subtest short-form for the WAIS-R. For the two-subtest short-form, the reliability values ranged from .91 to .95, and validity ranged from .87 to .92, and these values were higher than any other two-subtest combination. A different set of conversion tables were created by Brooker and Cyr (1986) for this same short form, which differ slightly from the conversion estimates provided by Silverstein (1982). No psychometric properties regarding the use of this different conversion table are given by Brooker and Cyr (1986).

Silverstein (1967) found that this short form correctly classified 68% of individuals according to their full-form IQ classification with the WAIS. This short form has been used by Crawford et al. (1992) in a British version of the WAIS-R with a normal

population, and evidenced slightly lower psychometric properties when compared to short forms which were composed of more subtests. A criticism against this short form (and against the four-subtest short-form to be discussed next) is that the Vocabulary subtest takes a considerable amount of time to administer, and also that the scoring of this subtest is somewhat subjective (Boone, 1992). This short form has been studied with Veteran patients (Ryan et al., 1998), elderly samples (Hoffman & Nelson, 1988; Paolo & Ryan, 1991), psychiatric inpatients (Benedict, Schretlen, & Bobholz, 1992; Boone, 1990; Roth, Hughes, Monkowski, & Crosson, 1984; Ryan, Larsen, & Prifitera, 1983), and non-psychiatric samples (Haynes, 1985; Robiner, Dossa, & O'Dowd, 1988).

Silverstein's (1982) 4-subtest Short Form. This short form, originally chosen by Doppelt (1956), is composed of the Vocabulary, Block Design, Arithmetic, and Picture Arrangement subtests. It produced high correlations between the full-form and estimated IQ among the age groups in the standardization sample, ranging from .85 to .95. Based on this short form, Doppelt (1956) produced regression equations for estimating the WAIS IQ for each age group in the standardization sample for the WAIS, and reduced them to a single regression equation for each age group, introducing an age-corrected intercept for each regression equation. The standard error of estimate for Doppelt's (1956) short form was 7 scaled score points, or about 4.2 IQ points, and 98% of the differences between the estimated and actual full scale IQ scores were contained within two standard errors. This short form took between 35 and 40 minutes of testing time. Silverstein (1982) further refined this short form using standardization data from the WAIS-R, providing newer conversion tables and psychometric data. Across all of the age

groups the reliability values ranged from .92 to .95, and the validity ranged from .90 to .95.

Crawford et al. (1992) found that this short form performed well with a normal British population, and was recommended over other four-subtest short forms. Boone (1990) found this short form to outperform five other short forms in a sample of psychiatric patients by virtue of its having a smaller standard error of estimate and providing accurate prediction of WAIS-R IQ (within one standard error of measurement) 90% of the time. This short form has also been studied with veterans (Ryan et al., 1998), elderly samples (Hoffman & Nelson, 1988; Margolis, Taylor, & Greenlief, 1986; Paolo & Ryan, 1991), psychiatric inpatients (Benedict et al., 1992; Boone, 1990, 1992; Roth et al., 1984; Ryan et al., 1983), individuals suspected of brain damage (Ryan, Georgemiller, & McKinney, 1984), and non-psychiatric samples (Haynes, 1985; Robiner et al., 1988; Ward, 1990).

Ward, Selby, and Clark's (1987) 4-subtest Short Form. Ward et al. (1987) investigated the claim that the purpose of a short-form is to save administration time, observing that many studies have only looked at length of time in terms of the number of subtests and related validities and not the actual administration times. They contended that there was more to selecting subtests for a short-form than simply looking at the validity of combinations of a number of subtests. A four-subtest short-form from the WAIS-R (Digit Span, Similarities, Picture Completion, and Digit Symbol) had a validity of .93, and took the same amount of time as a two-subtest short-form of Vocabulary and Block Design, which had a validity of .91. The authors do not provide any conversion

tables for these proposed short forms. Very little research has been done with this short form beyond the initial postulation of its existence.

Reynolds, Wilson, and Clark's (1983) 4-subtest Short Form. Reynolds et al. (1983) proposed a four-subtest short-form, consisting of the Information, Arithmetic, Picture Completion, and Block Design subtests from the WAIS-R. The two Verbal scale subtests had a correlation of .84 with the total Verbal scale, and the two Performance scale subtests had a correlation of .67 with the total Performance scale. Reynolds et al. (1983) also presented a reliability of .88 and a standard error of measurement of 5.26 for their short form. Silverstein (1990c) has presented a correction for the short form proposed by Reynolds et al. (1983), pointing out that a correction used by Reynolds for the spuriousness in the correlations between the subtests and the full scale was actually an overcorrection. The correction provided by Silverstein resulted in higher correlations for the Verbal and Performance dyads (.93 and .89, respectively) than those reported by Reynolds et al. (1983). Silverstein (1990c) also recalculated the reliability and standard error of estimate, achieving values of .94 and 5.26 respectively. Silverstein (1990c) then used a linear equating method to develop separate equations for each age group for the purposes of converting scaled scores on the Reynolds et al. (1983) short form to estimates of the Full Scale IQ, and presented a single regression equation that was used to construct a conversion table for the sum of scaled scores on the short form to Full Scale IQ estimates. Crawford et al. (1992) found this short form to perform slightly more poorly than the Silverstein (1982) four-subtest short form in a normal British population. This short form has also been studied in a veteran population (Ryan et al., 1998), elderly

samples (Hoffman & Nelson, 1988; Paolo & Ryan, 1991), and with psychiatric inpatients (Boone, 1992).

Alternative Short Forms (not studied in this project). Satz and Mogel (1962) proposed a shortening of the Wechsler battery by omitting several items from each subtest, rather than reducing the number of subtests which are administered. This allowed the short form to assess the full range of intellectual functions, as well as to produce many of the diagnostic features of the entire Wechsler battery. Mogel and Satz (1963) found high correlations of the short-form IQ estimates with the actual Full-Scale IQ score ( $r = .99$ ) and of the Verbal and Performance IQ estimates with their actual Full Scale counterparts ( $r = .99$  and  $r = .97$ , respectively). Holmes, Armstrong, Johnson, and Reis (1966) have also found support for this short form, but noted that the use of a scatter analysis derived from the short form could be compromised due to the reduction in administered items.

Silverstein (1968) developed a split-half version of the WAIS, and found that the estimated IQ scores were highly correlated with the actual Full-Scale IQ. Vincent (1979) developed a modified version of the WAIS (WAIS-M), using scores on a single subtest to determine the starting points on many of the remaining subtests. This method only reduced total administration time by about 25%, but showed incredible accuracy with regards to estimated IQ scores, was able to provide clinically relevant subscale patterns, and was extremely accurate with regards to IQ classification. Such findings have been replicated by Cargnello and Gurekas (1987, 1988) with the WAIS, and with the WAIS-R by Cella (1984).

Kaufman et al. (1991) considered both psychometric and non-psychometric issues in the development of their short forms. Besides considering validity and reliability, they looked at ease of administration and ease of scoring the chosen subtests. They also took into account the placement of the subtests within the overall administration procedure, as most short forms are developed from retrospective administrations of the complete Wechsler battery. This was intended to compensate for fatigue, boredom, and distractibility factors that may have resulted from subtests that are later in the administration. They further contended that the two-subtest short form by Silverstein (1982) was very time consuming, requiring over thirty minutes to administer in the standardization sample, and was very difficult to score. Using the WAIS-R standardization data, they devised three short forms. Their two-subtest short form is Information and Picture Completion; the three-subtest is Information, Picture Completion, and Digit Span; the four-subtest short form is composed of Arithmetic, Similarities, Picture Completion, and Digit Symbol. All of these short forms have administration times that are less than 20 minutes, with average reliabilities that are over .90 and average validities that range from .88 to .95. While these short forms have shown impressive psychometric qualities, one consideration is that some of the chosen subtests are somewhat unreliable and susceptible to situational factors (Boone, 1992).

#### Comparisons and Criticisms of Some of the Short Forms Based on Subtests.

Mogel and Satz (1963) have noted that short forms composed of selected Wechsler subtests do not give a lot of information, since these types of short forms place an emphasis on a single IQ index (the Full-Scale IQ score) and are generally unable to

yield other types of cognitive indices (e.g., Verbal and Performance IQ). The Wechsler tests are intended to give more than just a clinical score and can give insights into the individual work habits and behaviours of the test taker (Zimmerman & Woo-Sam, 1973).

Cyr and Brooker (1982) and Brooker and Cyr (1984) have pointed out that both reliability and validity need to be considered in the selection of subtests for a short form, reporting that Silverstein (1968, 1970) and possibly many other short form developers only considered validity in the selection of their subtests from the Wechsler batteries. They stipulate that “the consideration of only one psychometric property and not the other in the selection of subtest combinations would deter from capitalizing on the contribution of two basic properties of empirical measurements” (Cyr & Brooker, 1984, p. 903). Using the standardization data for the WAIS-R (Wechsler, 1981), the subtest intercorrelations and reliabilities were corrected for attenuation according to formulas outlined by Tellegen and Briggs (1967). For each possible combination of dyads, triads, tetrads, and pentads the average validity and reliability coefficients yielded composite measures that were used to define the theoretical overall “psychometric effectiveness” (Cyr & Brooker, 1984) of the combinations (the psychometric effectiveness rating will be denoted as  $r_{vr}$ ). Results showed that there were single combinations of dyads, triads, and tetrads that emerged, while the 32 best pentads were no better than the best tetrad. The best dyad was composed of Vocabulary and Block Design ( $r_{vr} = .93$ ); the best triad of Vocabulary, Block Design, and Information ( $r_{vr} = .95$ ); the best tetrad of Vocabulary, Block Design, Arithmetic, and Similarities ( $r_{vr} = .96$ ). The dyad combination is identical to the one reported by Silverstein (1968) for the WAIS and by Silverstein (1982) for the

WAIS-R. By considering both validity and reliability in the selection of subtests, Cyr and Brooker (1984) have opened the doorway for relevant clinical considerations (Doppelt, 1956) in the selection of a short form that best suits the needs of the clinician — the clinician can now choose a short form that will most adequately fit the functional ability of the test taker, and be assured that the short form will give an accurate measure of intelligence. Brooker and Cyr (1986) have provided tables for converting age-scaled scores from their three short forms (Cyr & Brooker, 1984) into Full-Scale WAIS-R IQ estimates.

Levy (1967), Tellegen and Briggs (1967), Bashaw and Anderson (1967, cited in Silverstein, 1990a, p. 4), and McNemar (1950) have maintained that the reported correlations between the short form and the Full Scale from which they have been derived are spuriously high, since the abbreviated tests are part of the original test. This has led to the developed short forms being subject to the assumption that the subtests chosen are equally reliable. If this assumption is not true, then the validities for the short forms may not be related to the reliability of the individual subtests which compose them. This can lead to the problem of different short forms giving different estimates when directly compared with each other. Silverstein (1985a, 1985c) has presented solutions and considerations on this point, some of which have been addressed at an earlier part of this paper (see Reliability and Validity of the Short Forms.).

Silverstein (1965) has investigated the issue of classification, as Mumpower (1964) has pointed out that the correct classification of an individual is more important than a high correlation between the short form and the full scale. Silverstein (1965)

argues that the degree of agreement of classification between a short form and the full scale will depend on two things: the degree of correlation between the short form and the full scale, and also the width of the classification interval. It was determined that as the width of the classification interval decreases and the correlation between the two forms decreased, the percentage of agreement in classification between the two forms decreased dramatically. For example, a correlation of .90 and a classification width of 8 IQ points gives agreement in IQ classification 85.6% of the time. While the classification widths of the Wechsler batteries are generally either ten- or twenty-point intervals, classification becomes a relevant issue for those individuals who are at the borderline of classifications, especially with the knowledge that the standard error of measurement of the Wechsler batteries is roughly 3 IQ points (Wechsler, 1997). This standard error of measurement is the amount of variance a given score will have around its true score (Kaplan & Saccuzzo, 1993), and thus an estimated IQ score that can fall within this standard error of measurement constitutes an accurate estimation of cognitive ability.

Thompson (1987) noted that a rescoring of existing protocols, a common procedure in the development of a potential short form, can neglect the effects that motivation and attention have on performance with the short forms. He found that the 2-subtest short form by Silverstein (1982) resulted in overestimation of IQ scores and greater misclassifications when the short form was administered separately from the rest of the Wechsler battery. These effects were not seen with the 4-subtest short form, indicating that as the number of subtests included in a short form increases the extent of misclassification decreases. A second issue raised by Thompson is concerned with the

ease of scoring the short form, since a short form should be just as easy to score as it is to administer. A careful choice of subtests could reduce the scoring time and possibly reduce the risk of scoring errors (Levy, 1968).

Silverstein (1990a) has indicated that the use of regression equations to provide estimates of IQ scores can give systematic over- or under-estimation of the actual IQ score. It has also been shown that there are two ways to score an individual's responses on a short form when using regression equations to obtain estimates — one way is to use the procedure outlined by McNemar (1950) if the short form is to be followed by the administration of the rest of the battery, or the procedure given by Silverstein (1970) where the short form is used as a replacement for the battery.

As noted earlier (see Scoring Issues for Short Forms), Tellegen and Briggs (1967) have proposed the use of Deviation Quotients (DQ), which have the same mean and standard deviation as the Wechsler batteries (Silverstein, 1985b). These DQs are not IQ scores, but are a standard score of a particular type, and they circumvent the problem of using either the McNemar (1950) or the Silverstein (1970) formulas as a basis for the regression estimates. Silverstein (1985b) has provided equations for the standard error of the estimate of the DQs, showing that the increase in error with the DQ method is negligible. Silverstein (1984) has also shown that the use of DQs yield accurate estimates of the full scale IQ, and are a viable alternative to linear regression. Engelhart et al. (1999) have shown that there is no difference between prorating and using linear equations to obtain IQ estimates from short forms, although the prorating method tended to inflate the standard deviation of the estimated scores and could generate IQ estimates

that were rather extreme. These issues were lessened as more subtests were included in the short form.

As mentioned previously, some of the short forms pose a problem to clinicians who use a profile analysis on the observed responses on the Wechsler batteries as an aid to their diagnosis. Short forms composed of selected subtests are unable to give the test examiner a profile, and short forms similar to the Satz and Mogel (1962) form are very unreliable (Watkins & Kinzie, 1970, cited in Silverstein, 1990, p. 5). This problem can be overcome through the use of special profile-analysis techniques (Goebel & Satz, 1975).

#### Short Forms in The Elderly.

The use of short-forms has been advocated within the elderly population for several reasons, mainly for issues related to time of administration (Ward et al., 1987). Additional concerns have been raised by Paolo and Ryan (1991) that are specific to the elderly, namely the reduction in both fatigue and motivational problems through the use of a short form. Hoffman and Nelson (1988) have also extended the issue of testing time to the case of requiring many tests for differential diagnosis. An extended or lengthy testing session can lead to a lack of concentration which could result in a test score that does not accurately reflect the true cognitive ability of the elderly person. Paolo and Ryan (1993) investigated the claim by Kendrick (1982, cited in Paolo & Ryan, 1993, p. 720) that elderly test takers find the WAIS-R to be “childish” and “largely irrelevant to the examinee’s everyday behaviors”. It was found that these claims were unsupported, and that most of the elderly in their sample found the WAIS-R to be “interesting,

challenging, motivating, and within their endurance” (Kendrick, 1982, cited in Paolo & Ryan, 1993, p. 721). Of all the subtests, only Picture Arrangement and Block Design were considered as being “tricky” (Kendrick, 1982, cited in Paolo & Ryan, 1993, p. 722), although positive and negative ratings of these two subtests showed no statistically significant differences.

Margolis et al. (1986) evaluated the two- and four-subtest short forms devised by Silverstein (1982) in a geriatric sample of patients suspected of dementia. The two-subtest short form gave a correlation of .93 with the full scale WAIS-R IQ, but the mean estimated IQ was higher than the mean actual IQ, and the IQ classification of the short form only matched that of the actual IQ 67% of the time. The four-subtest short form gave a correlation of .96 with the full scale WAIS-R IQ, the mean estimated IQ was not significantly different from the mean actual WAIS-R IQ, and estimated IQ classification was correct 83% of the time. These findings led Margolis et al. (1986) to contend that the short forms should not be used with the elderly when a precise estimate of IQ is required, but are appropriate for screening purposes and for monitoring the cognitive functioning of an elderly person over time.

Hoffman and Nelson (1988) compared six of the WAIS-R short-forms mentioned previously (Brooker & Cyr, 1986; Reynolds et al., 1983; Silverstein, 1982) in a sample of healthy geriatric participants. The results were analysed according to the three criteria postulated by Rabin (1943) and Resnick and Entin (1971), and evaluated by Silverstein (1985a, 1990a). The first criterion of high correlation between the short-forms and the original scale was met. The second criterion, that the mean IQ score on the short-forms

and the mean IQ of the original scale should not differ, was met by all of the short-forms except the four-subtest short-form proposed by Reynolds et al. (1983). The third criterion, regarding the percentage of disagreements between classifications, was not satisfied by any of the short-forms used, with the most accurate short-form correctly classifying 64.3% of the participants in accordance with their original-scale IQ classification. Performance on this third criterion improved greatly with a comparison of the short-form mean IQ scores that fell within a single standard error of measurement ( $\pm 6$  IQ points) of the original scale IQ score, resulting in classification accuracies as high as 93%.

Paolo and Ryan (1991) investigated six different short forms using the elderly standardization sample of the WAIS-R. They investigated the two- and four-subtest short forms of Silverstein (1982), the four-subtest short form of Reynolds et al. (1983), as well as the three short forms generated by Kaufman et al. (1991). It was determined that the short forms yielded similar mean IQ estimates, with the exception of the Reynolds et al. (1983) short form. Another finding was that the short forms performed poorly with respect to IQ classification, and this performance dropped even further when a sample of clinical elderly were used. Finally, Paolo and Ryan (1991) found that, as the number of subtests within a short form increased, the accuracy in IQ estimation increased.

As yet there has not been a great deal of research on the use of short forms with the WAIS-III. The most notable research has been conducted by Ryan and colleagues (Ryan, 1999; Ryan & Ward, 1999; Ryan et al., 1998) with two seven-subtest short forms. Iverson, Myers, Bengston, and Adams (1996) have found that both types of these short

forms provide identical results with regards to IQ estimation, and the seven-subtest short forms also provide estimates of Verbal and Performance IQ scores. This appears to be the only short form that has been systematically studied with the WAIS-III in terms of estimation of IQ scores.

#### General Criticisms of Short Forms.

McCormack (1956) has pointed out that the common practice of validating a test by correlating it with an already extensively validated test falsely implies that the new test will have similar validity properties to the old test. He has shown that two alternate scoring keys, which are similar to having a short form and a long form of a test, may be highly correlated but the new key may have a wide range of validities. As the correlation between two alternative keys decreases and the validity of one key (i.e., the full-form Wechsler) decreases, the range in possible validities for the alternate key (i.e., the short form) grows larger. Even at a correlation of .99 and a validity of .6 for the full-form, the range of possible validities for the short form is from .48 to .71, which are not very respectable values for a test of intelligence.

Even if the validities of two short forms are similar, it does not mean that both short forms save the same amount of administration time. Many studies do not give estimates of the total time saved by administration of the short form (Levy, 1968). Since test reliability is a function of test length, and many of the short forms are composed of different Wechsler subtests, a possible solution is to agree on a total short form time and vary the subtests within this time frame (Woodbury & Lord, 1956, cited in Levy, 1968, p. 413).

Levy (1968), Tellegen and Briggs (1967), Bashaw and Anderson (1967, cited in Silverstein, 1990), and McNemar (1950) have maintained that the reported correlations between the short form and the full scale from which they have been derived are spuriously high since the abbreviated tests are part of the original test, and as a result the errors of measurement in the short form are not independent of those in the original form. Doppelt (1956) argued against this criticism, saying that the whole point of devising a short form was to be able to predict the whole from the part. Other authors (Silverstein, 1970; Tellegen & Briggs, 1967) have produced correction formulae to address this criticism.

Kramer and Francis (1965) discovered that the use of short forms yielded seriously inaccurate estimates of Full-Scale WAIS IQ, even when the correlation between the short-form and the full-form was high, a point also noticed by Holmes et al. (1965). Kramer and Francis (1965) utilized the Doppelt (1956) short form and the prorating of scores on a three-subtest short form (Information, Similarities, Block Design) with a sample of psychiatric patients, both of which had correlations of .94 with the Full Scale WAIS IQ. It was found that the Doppelt short form misclassified 56% of the participants, and the three-subtest short form misclassified 71% of the participants. When the scores from both short forms were transformed to have the same mean and standard deviation as the Full Scale WAIS IQ scores (similar to linear equating), the misclassification dropped to 27%. Such a transformation is assumed by the correlation given previously (Kramer & Francis, 1965), but this cannot be done in regular clinical settings since generally the full WAIS is not given for the test examiner to compare the short form estimated IQ score

with. Kramer and Francis (1965) computed the effective correlation of the short form with the Full Scale WAIS as being equal to .60. It should be noted that the argument presented by these authors does not actually address the precise estimate that the short form yields, but instead questions the classification of the individual according to an IQ level. This same criticism has been pointed out by other authors (Groth-Marnat, 1990; Mumpower, 1964). Silverstein (1965) reached the same conclusion regarding the transformation of the scores as Kramer and Francis (1965) have, and the linear equating procedure by Tellegen and Briggs (1967) addresses this issue.

Groth-Marnat (1990) has stated that short forms have a higher magnitude of error when estimating the IQ of an individual. As most short forms of any worth give accurate estimates of the full scale IQ score, the real issue may be the amount of error that is associated with the estimate, which may be elevated due to the use of fewer subtests. Silverstein (1984) addressed the issue of the standard error of the estimate when a short form of a Wechsler test is used with a deviant participant (a participant who is very high- or low-scoring) and who was not part of the standardization sample for the Wechsler battery, a criticism that is similar to McNemar's (1950) regarding the use of atypical samples when selecting the subtests or items to be included in a short form for the Wechsler batteries. Results indicated that the standard error of estimate for a high- or low-scoring individual is not much different from that of an average-scoring participant (differing by .01 when the correlation between the short form and the full scale is .95, and differing by .02 when the correlation between the short form and the full scale is .90). Silverstein concludes that the standard error of the estimate for the standardization

sample is appropriate to be used as an approximation of the standard error of estimate for a high- or low-scoring participant, and as a result the criticism pointed out by Groth-Marnat (1990) is not as serious as it appears. Ryan and Rosenberg (1984, cited in Spandel, 1985) have stated that if only a single IQ index is available from the administration of a short form, then the test administrator should indicate a precision range for the IQ index which takes into account the standard error of estimate.

McNemar (1950) and Doppelt (1956) have stated that most validation studies for short forms have used atypical samples, and the results have been compared with the standardization sample for accuracy. With the atypical samples, fewer items of the short form will have been given, and this is compounded, if one is to assume that performance by such atypical samples would have been identical, if they had been given the full scale (Tipton & Stroud, 1973). This is related to the methodological concern of rescored Wechsler protocols that was raised previously (see Scoring Issues for Short Forms). Some of the scoring methods used in creating short forms (e.g., proration, linear equating) assume that the test taker would have answered consistently on all subtests, and with atypical samples this assumption stands a good chance of being false, and so estimates of IQ scores based on them may not be entirely accurate.

#### General Advantages of Short Forms.

One of the obvious advantages of a short form is the reduced administration time (Cargnello & Gurekas, 1987; Doppelt, 1956; Levy, 1968). Depending on the type of short form chosen, reductions in administration time can range from 25 to 50%. The type of short-form also dictates how the estimates of IQ scores are generated, either by

prorating scores or through the use of regression equations. Whitmyre and Pishkin (1958), Himmelstein (1957), and Engelhart et al. (1999) have shown that both of these methods give similar results, with regards to estimation accuracy, for both the WAIS and the WAIS-R.

Mogel and Satz (1963) have found that factors such as the discontinuity in item difficulty have a negligible effect on performance of the test taker, which is especially relevant when one is using a short-form composed of selected items from all of the subtests. An advantage that is specific to those short forms that use a selection of items from all of the subtests (e.g., Mogel & Satz, 1963) that make up the entire scale is that they save time and do not sacrifice some of the features that are assessed with the complete scale (Holmes et al., 1965), namely the availability of other cognitive indices (VIQ and PIQ), and the observation of the individual in a number of problem-solving situations. This advantage is also seen in the short form that was constructed by Vincent (1979).

As noted above in the section on criticisms of the short forms (General Criticisms of Short Forms), McNemar (1950) and Doppelt (1956) have said that the determination of the usefulness of short forms has typically been carried out using atypical samples. This had led to a number of the short forms that have been developed being based on the standardization data from the respective Wechsler battery from which they have been made. McNemar (1950) has stated that using the standardization data in the formation of a short form results in more dependable validities.

The Shipley Institute of Living Scale (SILS).

The Shipley Institute of Living Scale (SILS), developed by Walter Shipley (1940), was originally devised to provide a quick measure of mental deterioration, and can be administered either individually or to a group. A copy of the SILS is given in Appendix C. The SILS was standardized using 1046 individuals (217 college students, 257 high school students, 572 grade school children), with mental age equivalents (the average mental ability displayed by people of a given age; Coon, 1995) generated from the total score on the SILS. The SILS consists of two tests, a Vocabulary test and an Abstraction test, designed to assess the differential between vocabulary and abstract thinking levels as an index of mental deterioration. This index of deterioration is also known as the Conceptual Quotient (CQ). The two halves of the scale have a ten-minute time limit each, and are both pencil-and-paper tests allowing the test taker to indicate responses by filling in a blank or by circling a response. For those test takers who are unable to physically complete the SILS, it is possible for answers to be verbalized and recorded by the examiner in an individual testing session. Shipley (1940), comparing the mental ages of a sample of 322 army recruits against the mental ages of the standardization sample, reported a reliability coefficient of .92 when the scale was used as a measure of intelligence.

The SILS has been found to be resistant to the effects of gender and race (Corotto, 1966; Dalton & Dubnicki, 1981). Age effects have been seen with the SILS when comparisons are made across the lifespan (e.g., Corotto, 1966; Dalton & Dubnicki, 1981; Harnish, Beatty, Nixon, & Parsons, 1994; Heinemann, Harper, Friedman, & Whitney,

1985), with most trends showing decreasing scores as age increased. Paulson and Lin (1970a) have found that Shipley Vocabulary scores increase while Shipley Abstraction scores decrease with increasing age (the decline in Abstraction scores occurred after age 44). Morgan and Hatsukami (1986) obtained SILS data for a highly functioning elderly sample (indicated by performance on WAIS Block Design and Digit Span tests), covering the ages from 60 to 85 years. Their goal was to determine the efficacy of the SILS as a screening device for cognitive ability in the elderly. There was no effect of age on the SILS scores, indicating that, for this age group as a whole, the changes in cognitive functioning are small (this can also be interpreted as a stable level of functioning across the age group). Further investigation of the scoring patterns on the components of the SILS revealed that participants in this age group scored low on the number of correct Abstraction items (38% of the sample answered four or fewer items correct), even though participants were considered to be highly functioning. As well, only 5% of the sample's SILS scores were within the CQ "normal" range, although this was most likely due to the low Abstraction scores for the majority of the participants. The authors suggested that age-normed data may be insufficient to make the SILS a useful screening measure for the elderly. This reflects a concern mentioned by Powell and Whitla (1994), where some tests that are used for cognitive screening in the elderly have a low ceiling. The healthy elderly test taker will achieve a score on a test that is similar to the score that would be achieved by an unhealthy elderly test taker, and so the ability of the test to discriminate the two classes of elderly test takers is weak at best, even with normative data from a set of healthy elderly test takers.

The effects of education on the SILS have been shown by Harnish et al. (1994), with higher scores being associated with increasing years of education. Prado and Taub (1966) had earlier indicated that the SILS is an accurate predictor of intelligence regardless of age, which speaks towards the advocacy of normalization data that is not based on the age of the test taker. A closer inspection of the participant demographics in the Morgan and Hatsukami (1986) study showed that 72% of the sample had only high school or lower education. Morgan and Hatsukami (1986) also found a significant effect for education level on the achieved score on the SILS, as have others (e.g., Dalton & Dubnicki, 1981; Fowles & Tunick, 1986; Paulson & Lin, 1970a; Prado & Taub, 1966). Kish and Ball (1969) have further demonstrated the influence of low education on Abstraction scores for the SILS, and concluded that this factor could lead to misclassification of the participant if it has not been taken into account.

Given the low academic level of the Morgan and Hatsukami (1986) sample, and the influence that education has towards performance on the SILS, the real issue may be one of education-normed data and not age-normed data, a conclusion that is in support of the one found by Morgan and Hatsukami (1986). This lends further support to the general finding that higher levels of education lend themselves towards higher intelligence scores (Powell & Whitla, 1994; Schaie & Geiwitz, 1982; Troll, 1982), and also reinforces the need to consider education in the standardization of any test that is used to assess cognitive ability.

Estimation of WAIS IQ Using the SILS. Prado and Taub (1966) proposed to use the SILS as a screening test for estimating intellectual functioning, and they compared

obtained scores on the SILS with obtained scores on the WAIS (Wechsler, 1955). The SILS was found to rapidly and accurately estimate average or better intellectual functioning on the WAIS. Bartz and Loy (1970), Paulson and Lin (1970b), and Sines and Simmons (1959) have provided SILS-to-WAIS conversion tables to facilitate the use of the SILS for estimating WAIS IQ. Dennis (1973) performed a comparison of these conversion formulae used in estimating full-scale WAIS IQ score from the SILS. The conversion procedure outlined by Paulson and Lin (1970b), which includes an age correction, led to the best estimations of Full-Scale WAIS IQ, as well as having the lowest standard error of estimate. The age-corrected tables of Paulson and Lin (1970b) were also shown to work well at the higher and lower ends of total SILS scores.

Estimation of WAIS-R IQ Using the SILS. Zachary, Crumpton, and Spiegel (1985) developed an updated procedure for estimating the Full-Scale WAIS-R IQ score from the SILS using a sample of male psychiatric inpatients. The authors produced an age-corrected procedure which gave accurate WAIS-R IQ estimates and a correlation of .85 with the estimated WAIS-R IQ and actual WAIS-R IQ scores. In a comparison with five other conversion systems, the age-corrected formula provided by the authors was the only procedure that produced IQ estimates which did not over-estimate the actual IQ of the test taker. The authors also presented a table for the conversion of SILS scores, stratified by age group. A limitation in their procedure is that it produces a low ceiling on the maximum obtainable WAIS-R estimate. As a result, those individuals who score high on the SILS may receive a converted WAIS-R score that is not representative of their true IQ score. This restriction is most noticeable with test takers in the 30-50 age

range.

Retzlaff, Slicner, and Gibertini (1986) tested the procedure outlined by Zachary et al. (1985) in a homogeneous sample of nonpsychiatric military personnel. A low correlation between the estimated and actual WAIS-R IQ score was obtained ( $r = .45, p < .01$ ) which was attributed to the homogeneity of the sample. The authors found a small measure of validity given to the Zachary et al. (1985) procedure, since the estimates of full-scale WAIS-R IQ scores were comparable to the actual WAIS-R IQ score (105.03 and 101.63, respectively). Retzlaff et al. (1986) also found that the addition of a constant value to the SILS-to-WAIS conversion table of Paulson and Lin (1970b) resulted in accurate estimations of the WAIS-R IQ score. Weiss and Schell (1991) replicated the findings of the Retzlaff et al. (1986) study with a sample of psychiatric inpatients.

Dalton, Pederson, and McEntyre (1987) used the Zachary et al. (1985) procedure on a comparison of the SILS and various short-forms of the WAIS-R for giving estimates of WAIS-R IQ scores. The authors used the Vocabulary and Abstraction component scores rather than the total SILS score and added race and education variables in an attempt to devise a more accurate formula than the Zachary et al. (1985) procedure. The Zachary et al. (1985) procedure gave a correlation of .73 with the obtained WAIS-R IQ score, and the new formula devised by the authors did not yield a significantly higher correlation ( $r = .75$ ). Contrary to previous research, education level was not a significant moderating variable in the resulting regression formula used by the authors. The authors also determined that a number of single-subtest and two-subtest short-forms of the WAIS-R gave equal or better correlations than the Zachary et al. (1985) procedure. Three

two-subtest short-forms which had correlations ranging from .85 to .87 were Information and Picture Completion, Arithmetic and Picture Completion, and Similarities and Picture Completion.

The over-estimation of WAIS-R IQ scores evidenced by the many of the SILS conversion procedures (e.g., Paulson & Lin, 1970b; Sines & Simmons, 1959) could be due to what is known as the “Flynn effect” (Neisser, 1997). The Flynn effect states that scores on current intelligence tests appear to be increasing at a rate of about 0.3 IQ points per year when compared to results from previous test batteries. As a result, individuals who are administered the WAIS-III today and who land in the “average intelligence” range will land in the “high average” or “superior” IQ ranges on the WAIS-R. As the Sines and Simmons (1959) and Paulson and Lin (1970b) conversions are based on the WAIS, their overestimation of predicted WAIS-R IQ is most likely a reflection of the Flynn effect. The SILS itself may not be subject to the Flynn effect due to its nature — it was never intended to be an intelligence test, but instead a test to indicate mental deterioration. The Flynn effect may become evident when a researcher is making an estimate of IQ score based on the SILS, since the estimation procedure is achieved by relating a SILS score to a score on an intelligence test. Although previous research has not documented the Flynn effect with other conversion methods for the SILS, this effect may be present when using the Paulson and Lin (1970b) or the Zachary et al. (1985) procedure to estimate the Full-Scale IQ on the WAIS-III.

#### Cognitive Patterns In The Elderly.

As noted earlier in this paper, there are certain cognitive patterns that emerge

across the lifespan. There are three facets of cognition that are generally observed: general cognition, pragmatic cognition, and mechanic cognition (Lindenberger & Reischies, 1999). Also noted earlier were the parallels between these three areas and the concepts of the Full-Scale IQ, the Verbal IQ, and Performance IQ as defined within the Wechsler test batteries. Thus, cognitive patterns over the lifespan can be investigated in two areas: the functioning of general cognition, and the relative functioning of pragmatic to mechanic cognition.

Both Botwinick (1977) and Labouvie-Vief (1990) have reviewed literature regarding the course of general cognitive functioning in the elderly, which Botwinick has termed the "classical aging pattern." When compared to younger adults, there is a decline in all cognitive areas, which can begin as early as age 50 but does not become severe until the mid-70s, a conclusion reported by Botwinick in 1977. Eisdorfer, Busse, and Cohen (1959, cited in Botwinick, 1977, p. 584) have shown that this pattern holds across gender, race, and socioeconomic status using the WAIS. Lindenberger and Reischies (1999) have found decline as age increases in a sample of individuals aged 70 and above on five cognitive tasks (reasoning, knowledge, perceptual speed, memory, and fluency), and that this pattern was seen in both healthy elderly and those diagnosed with dementia. A similar result was found even earlier by Schaie and Strother (1968, cited in Matarazzo, 1972, p. 113) using the Primary Mental Abilities Test, in which performance was stable between the ages of 30 and 60, at which point a decline in ability began to appear. This pattern is even seen in a comparison of elderly and young adults who were matched for Wechsler IQ score (Harwood & Naylor, 1971, cited in Botwinick, 1977, p. 585).

The picture changes somewhat in longitudinal studies of intellectual performance. Botwinick (1977) states that most longitudinal studies show less decline in intelligence than cross-sectional studies, and that the decline begins later in life than the cross-sectional designs would have us believe. Botwinick also states that any maintenance of abilities that are seen in cross-sectional studies are also preserved in longitudinal studies, a finding that has been supported by Labouvie-Vief (1990) and by Schaie and Geiwitz (1982). Schaie and Geiwitz (1982) also point out that many of the differences observed in cross-sectional studies are due primarily to generational differences and not actual age changes. These findings are supported by recent research with the WAIS-R (Hopp, Dixon, Grut, & Backman, 1997), which showed a stable pattern of performance over a 2-year period in a sample of healthy elderly. Willis and Baltes (1982) also state that longitudinal studies show a great deal of inter-individual differences, generally manifesting as a large amount of variability in performance as age increases. Thus, any study that focuses on the elderly will most likely discover that, even while there is a great deal of variability in the observed scores, the overall picture will be one of stability rather than of decline.

The investigation into the relative performance of pragmatic and mechanic cognition, more commonly known as crystallized and fluid intelligence, shows a slightly different pattern than that seen with general cognition. Matarazzo (1972) has stated that, with regards to certain intellectual tasks, performance ceases to increase with age due to the abilities which are tapped by the tasks being at their limits of development. Botwinick (1977) called those tasks which did not decline as being “age-insensitive”, and

that these were generally the crystallized tasks. In contrast, the performance of fluid intelligence declines in the elderly. The concept of a “classic aging pattern” with respect to these abilities was highlighted by Botwinick (1977), in which performance on tasks that are nonverbal show greater age differences across the lifespan than do those which are verbal. This was also supported by Labouvie-Vief (1990). Research specifically with the Wechsler battery (e.g., Boone, 1995) has shown that Verbal IQ scores remained constant over the age span but that Performance IQ scores showed a decrease.

Many clinicians utilize the Verbal and Performance IQ scores to assess the presence of potential cognitive deficit (Groth-Marnat, 1990). As the mere presence of a discrepancy does not indicate a neurological deficit, the meaning of any observed discrepancy in the Verbal and Performance scales must be made from their association with other measures (O’Donnell & Leicht, 1990). While the WAIS-III - WMS Technical Manual (Psychological Corporation, 1997) posits a difference of 9 IQ points as being a significant difference at the .05 level, they cite some evidence that such differences may not be rare in the general population. Many of the short forms that have been studied with the Wechsler batteries have not addressed their ability to produce a VIQ-PIQ difference, having instead focused on providing an accurate estimate of global cognitive functioning, embodied by the FSIQ score. Thus, at this point, it is unknown how the short-forms perform with regards to generating VIQ-PIQ differences.

## PRESENT STUDY

### Research Objectives

The aim of this study is to investigate the psychometric properties of several procedures for estimating the Full-Scale IQ score on the Wechsler Adult Intelligence Scale III. There will be a total of six estimation procedures evaluated. The first four procedures will involve short-forms of the Wechsler scale that were derived from the WAIS-R. The four short forms are of the subtest-only format, and were chosen for their prominence in the psychological literature compared to other short forms. The two remaining methods are based on a conversion procedure that utilizes the Shipley Institute of Living Scale (SILS), a test battery that is both quantitatively and qualitatively different from the Wechsler battery.

The elderly were chosen as a sample due to a rather pragmatic observation that these individuals would be one of the most likely set of candidates for cognitive testing. This is due to an observation that there is a difficulty in separating changes in behaviour and functioning from the results of normal or pathological aging (Hayslip, 1983). Further, these individuals would generally require a differential diagnoses to be made regarding cognitive functioning, which requires many test batteries to be administered (Hoffman & Nelson, 1988) and a desire for compact, reliable testing procedures (Paolo & Ryan, 1991; Ward et al., 1987). The determination of short forms for assessing cognition would be of relevance to clinicians in this regard.

This study seeks to address several specific questions regarding the psychometric properties of the six chosen short forms, the findings of which may be extended to other

short forms. Firstly, and most importantly, can the short forms yield accurate estimations of WAIS-III Full-Scale IQ scores? This question shall be addressed by examining each of the short forms in relation to the three criteria stated by Resnick and Entin (1971). Namely, (a) is the correlation between the short form estimated IQ and the full-form IQ high and significant, (b) is the short form estimated IQ statistically different from the full-form IQ, and (c) is the qualitative classification achieved by the short form estimated IQ similar to that denoted by the full-form IQ?

Several other considerations for adequacy of the short forms will also be examined. Since the short forms were developed for the WAIS-R, several potential problems present themselves. The first is the type of procedure used to obtain the estimated IQ from the short form. For those four short forms which are based on the WAIS-R, some researchers or clinicians may continue to use these short forms in regular practice. As a result, they may also utilize the conversion tables that were derived from the WAIS-R with these WAIS-III subtests. As Neisser (1997) has shown, this may lead to inaccuracy in IQ estimation, usually manifesting as an overestimation of IQ scores. Further, those WAIS-R conversion tables may contain a cohort effect, due to the type of standardization sample used, which could add to the inaccuracy of the IQ estimates when carried over to the WAIS-III. Thus, two sets of scores will be computed. The first set of IQ estimates will be computed using the WAIS-R conversion tables for the elderly samples. The second set of IQ estimates will be computed using the linear equating procedure outlined by Tellegen and Briggs (1967). This will allow for the comparison of the two sets of estimations in their ability to estimate the WAIS-III IQ scores. It may be

that the IQ estimates from both types of conversion are statistically similar according to the three criteria outlined by Resnick and Entin (1971), and this finding would preclude the need to develop newer scale conversions. Yet, two sets of scores could meet all of the criteria, but one set could be closer to the evaluative criterion (in this case, full-form Full-Scale IQ) than the other. As a result, the two estimation procedures need to be evaluated with respect to each other. By omitting those short forms that do not provide similar mean estimates to the full-form WAIS-III score, an examination of the remaining short forms on other psychometric properties (e.g., correct IQ classification, administration time) will allow for the determination of the best short form, and will provide added evidence towards a decision to create newer conversion tables for use with the currently investigated short forms. Similar considerations as those noted above exist for the SILS, although some previous authors have reported that a simple addition or subtraction of a numerical constant is all that is required to make the formulas compatible with different Wechsler batteries.

A concern noted by several authors, with respect to the Wechsler-based short forms, is the lack of information that a short form can provide. This is especially true if the short form is composed of a limited number of subtests from the entire battery. A comparison of the short forms with the Full-Scale Wechsler battery with regards to some of the previous theoretical and empirical findings regarding cognition and ageing are necessary. Specifically, can the short forms exhibit the same age-related trends in overall, pragmatic, and mechanic cognition, operationally defined as Full-Scale, Verbal, and Performance IQ, respectively? A singular interest in this regard is whether or not the

short forms can evidence a pattern of responding across the age groups that is similar to that of the full-form Wechsler battery. A second area of interest is related to one of the clinically useful properties of the Wechsler batteries, and that is the use of the difference between Verbal and Performance IQ scores to detect potential cognitive deficit. If short forms are to be used for screening purposes, their ability to detect potential deficit needs to be examined.

Four short-forms have been chosen due to the frequency of their investigation in previous studies and their reported ability in producing accurate estimates of WAIS-R IQ scores. The four short forms were discussed at some length in the Literature Review section (Short Forms For The Wechsler Batteries), and are reiterated here.

- Short-Form #1: Vocabulary, Block Design (Cyr & Brooker, 1984; Silverstein, 1982, 1985)
- Short-Form #2: Vocabulary, Block Design, Arithmetic, Picture Arrangement (Silverstein, 1982, 1985)
- Short Form #3: Digit Span, Similarities, Picture Completion, Digit Symbol (Ward et al., 1987)
- Short-Form #4: Information, Arithmetic, Picture Arrangement, Block Design (Reynolds et al., 1983; Silverstein, 1990)

The third short form, devised by Ward et al. (1987), was chosen to explore their observation that this particular short form takes the same amount of time to administer as does the two-subtest short form of Vocabulary and Block Design (Cyr & Brooker, 1984; Silverstein, 1982, 1985), as well as possessing equivalent psychometric properties. These

four short forms are all of the selected-subtest variety as this will facilitate the research methodology to be used in this study, namely the administration of the subtests that make up the short forms initially, followed by the remaining subtests in the prescribed sequence in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997). Thompson (1987) has shown that altering the administration sequence of the Wechsler batteries does not affect performance on the battery.

For the Shipley Institute of Living Scale, there are two main methods that yield accurate estimations of Wechsler IQ scores:

(1) Wechsler IQ score =  $0.92 (\text{SILS}) + 63.10$  (Paulson & Lin, 1970b)

(2) procedure outlined by Zachary et al., (1985)

(a) estimated Wechsler IQ mean =  $85.859 + (1.196 \times \text{AGE}) - (.017 \times \text{AGE}^2)$

(b) estimated Wechsler SD =  $12.481 + (.581 \times \text{AGE}) - (.006 \times \text{AGE}^2)$

(c) Wechsler scaled score =  $27.650 + (1.320 \times \text{SHIPLEY}_{\text{total}})$

or

Wechsler scaled score =  $9.659 + (1.680 \times \text{SHIPLEY}_{\text{total}})$

(d) Wechsler estimated IQ =  $\frac{((\text{WAIS scaled score} - \text{estimated WAIS mean}) / \text{estimated WAIS SD}) \times 15 + 100$

In part (c) of the Zachary et al. (1985) procedure, the first equation was based on a least-squares regression procedure, and the second equation was derived using a rescaling method so that the estimated and observed scores have the same mean and standard deviation. Zachary et al. (1985) stipulate that either equation can be used at the discretion of the researcher. In the present study the second equation in part (c) will be utilized as it

provides estimates which are closer approximations to the range of actual WAIS-R scores (Zachary et al., 1985). The first conversion procedure is based on the WAIS, and the second based on the WAIS-R.

### Hypotheses.

The specific hypotheses to be evaluated are:

H1: The correlation between the short form IQ estimates and the full-form IQ scores will be high, positive, and significant for both the Wechsler-based short forms and the SILS.

H2: The distributions of the IQ estimates from both the Wechsler-based short forms and the SILS will not be significantly different from that of the full-form WAIS-III IQ scores.

H3: That the short forms will be able to accurately classify individuals into qualitative categories that are similar to those indicated by the full-form WAIS-III IQ score.

A side issue concerns the number of subtests which comprise the short forms. Wechsler (1981) has reported that assessment of intelligence can be achieved by assessing several cognitive areas. Of the four short forms chosen for study, one of them is comprised of only two subtests while the remaining three are comprised of four subtests. With respect to the Wechsler-based short forms, it is expected that the short forms which utilize four subtests will give higher correlations, IQ estimates which are more accurate, and a greater amount of agreement in qualitative IQ classification than short forms which utilize three or fewer subtests due to the greater number of subtests

sampling a larger domain of cognitive abilities. More formally,

H3a: Those short forms composed of four subtests will provide higher correlations with Full-Scale IQ, and be more accurate with regards to mean Full-Scale IQ estimates and with qualitative IQ classification.

A fourth hypothesis comes from the recognition that the Flynn effect may be present in the conversion procedures being currently investigated. As two methods of estimation are being used with the Wechsler-based short forms, namely the WAIS-R-based conversion tables and a linear equating procedure, the Flynn effect stipulates that the WAIS-R-based conversions should provide overestimation of the WAIS-III IQ score, and that this overestimation will be over and above any seen by the use of the linear equating procedure. A similar line of reasoning holds for the SILS conversion protocols, in that they should provide overestimations of the Full-Scale IQ which are over and above that seen by the linear equating procedure. Thus, while all of the short forms may find support for Hypotheses 1 through 3 (and 3a for the Wechsler-based short forms), the use of the linear equating procedure should provide estimates which are more accurate than the other estimation methods. This can be statistically tested by the following two hypotheses:

H4: The correlation between the linear equated Full-Scale IQ estimates and the actual Full-Scale IQ scores will be higher and significantly different than the correlation between the analogous WAIS-R-based Full-Scale IQ estimate and the actual Full-Scale IQ score.

H5: The difference between the linear-equated and the WAIS-R-based Full-Scale

IQ estimates will show the linear-equated estimate to be significantly lower than the WAIS-R-based estimate.

As the linear-equated estimates are expected to perform in a superior manner compared to their WAIS-R-based counterpart estimations, the accuracy of the linear-equated estimates of Full-Scale IQ will be investigated by computing the absolute difference in estimated Full-Scale IQ score from the full-form Full-Scale IQ. As there are no linear-equated analogs for the SILS estimates, statistical comparisons such as those outlined above are difficult to carry out. As such, a more qualitative assessment will be utilized. Specifically, it is expected that the Zachary et al. (1985) method for converting SILS scores will give both closer estimates of WAIS-III IQ scores and will be more congruent with the IQ classification than the Paulson and Lin (1970b) SILS conversion tables for several reasons: (1) the Zachary et al. (1985) equations being based on the WAIS-R versus the Paulson and Lin (1970b) tables being based on the WAIS, and (2) that the Zachary et al. (1985) scores will show less upwards drift in IQ scores than will the Paulson and Lin (1970b) tables due to the relative recentness of the WAIS-R compared to the WAIS.

As noted earlier, it is expected that the linear-equated short forms will provide the most accurate estimation of IQ scores. If they are supported through Hypotheses 1 through 5, it would be informative to determine if the linear-equated short forms can evidence some of the more theoretical aspects of cognition in the elderly. There is a consistent pattern to cognitive functioning in the elderly, as discussed in the Cognitive Patterns In The Elderly section, and the ability of the short forms to capture this pattern

needs to be assessed if they are to pass muster on theoretical grounds. A trend analysis (Keppel, 1991) shall be carried out to determine if there is the presence of a significant trend within either the full-form or the linear-equated Full-Scale IQ scores. The cross-sectional research to date has indicated a slightly decreasing trend as age increases. Thus, a trend analysis should be able to detect a decreasing linear trend in Full-Scale IQ estimates as age increases. A sixth hypothesis is concerned with the ability of the short forms to mimic the mapping of the general path of cognitive ability in the elderly, and can be formally stated as:

H6: As age increases, the estimate of Full-Scale IQ from the linear-equated short-forms composed of four subtests will produce a decreasing linear trend.

A seventh area of investigation is the ability of the short forms to detect indicators of potential cognitive deficit. One indicator that is commonly used with the Wechsler batteries is the difference between the Verbal and Performance IQ scores (Groth-Marnat, 1990; O'Donnell & Leicht, 1990). Using the Verbal and Performance scores computed for each of the four-subtest short forms, the absolute difference between Verbal and Performance scores will be computed. A seventh hypothesis can then be stated as:

H7: The number of significant absolute VIQ-PIQ differences in the 4-subtest short forms will be similar to the number of absolute VIQ-PIQ differences of the full form.

An investigation into the accuracy of the estimation procedures for each individual will also be carried out. The absolute difference between the full-form and estimated VIQ-PIQ difference will be calculated for the full-form battery and for the four-subtest-only short forms, as the single subtests which make up the two-subtest short

forms can be considered unrepresentative of the Verbal and Performance cognitive domains. As well, no conversion formulae have been presented to convert the SILS scores into VIQ or PIQ scores. A frequency table will be constructed to examine the distributions of the estimated differences. As this procedure is exploratory in nature, no specific hypotheses as to how the short forms will perform will be made.

The influence of education on intellectual test performance has been noted by Matarazzo (1972) with the Wechsler batteries, and also with the SILS (e.g., Harnish et al., 1994; Morgan & Hatsukami, 1986). The amount of attained education will be assessed to determine if it has an effect on both the full-form IQ scores and on the short-form IQ estimates.

## METHODOLOGY

### Participants

Participant recruitment started with an initial meeting with the Director of Nursing at each of the personal care homes and day hospitals. Each centre was given an introductory letter and statement, given in Appendix D. The nature and aim of the study was stated in this document, and an outline of several particular eligibility criteria was included in this statement. The first criterion was that the participant fell within the appropriate age range. A second criterion was that the individual not be diagnosed with dementia or some other neurological deficit (e.g., Alzheimer's). This second criterion was not diagnosed directly by the principal investigator, but was instead part of the initial selection procedure carried out by the nursing staff at each of the personal care homes and day hospitals. A third criterion was that the individual not have any serious hearing, vision, or motor problems. Individuals with less than 20% vision were excluded from the study. These first three criteria were not directly assessed by the principal investigator, as this information was readily available from the nursing staff at the respective personal care homes.

To test for motor and cognitive soundness, a final criterion was an evaluation with the Mini Mental State Examination (MMSE; Folstein, Folstein & McHugh, 1975, cited in Pangman, 1996). This was given individually to each participant if an MMSE had not been administered by the participating centre in the past two months. The MMSE provides information about orientation in time and space, short-term memory and mental skills (e.g., calculation), and language function (Edwards, 1993). The MMSE consists of

11 questions and the range of the total score is from 0-30. High scores (24-30) indicate intact cognitive functioning, and low scores (0-23) indicate varying degrees of impaired cognitive functioning (severe to mild) (Pangman, 1996). The MMSE has shown resistance to age and education effects in older samples (Hopp et al., 1997). Pangman (1996) reports that a cut-off score of 23/24 out of 30 on the MMSE is generally used in the elderly population to ensure adequate sensitivity. The mean MMSE score for the current sample was 27 ( $SD = 3$ ), with a range from 23 to 30. A copy of the MMSE is given in Appendix E, with instructions for administration given in Appendix F. No further screening for dementia or other cognitive deficits was performed.

A total of 109 elderly individuals (36 male, 73 female) were approached and signed consent forms for participation in this study. Eight of these individuals (3 male, 5 female) had to be dropped from the study due to scores below 23 on the MMSE. One female had to be dropped due to a failure to see the stimuli of the WAIS-III, even though this individual obtained a passing score (23 out of 30) on the MMSE. The final sample was composed of 100 elderly individuals (33 male, 67 female) over the age of 65 who were volunteers from personal care homes and day hospitals in the Winnipeg area. The mean age of the sample was 80.6 years ( $SD = 7.6$ ), with a range from 65 to 97 years of age. The mean education level was 10 years ( $SD = 3.14$ ), with a range from 3 years to Masters level education.

#### Research Protocol

All participants were given a copy of an introductory letter, given in Appendix G, which outlined the nature of the entire study and the extent of their participation. These

points were also reiterated verbally by the principal investigator in a face-to-face meeting.

Each participant signed a consent form which stated principally that (a) the information obtained from the experiment would remain confidential and anonymous, (b) participants would receive a full debriefing with regards to the nature of the study, and (c) that participants were free to stop the experiment at any time. A copy of this consent form is given in Appendix H. No deceptions of any sort were employed in recruitment. As well, each introductory letter and consent form had a unique identification code (ID number) on it, so that the identification of the participants (e.g., name) was not required on the test forms. This ID number was also used to identify the participants on the test forms, in place of their names, to ensure confidentiality.

After the screening procedure, commencement of the experimental session began. Every session began with the administration of the Wechsler Adult Intelligence Scale III (WAIS-III; Wechsler, 1997), a fourteen-subtest battery designed to assess levels of cognitive functioning. The Object Assembly subtest was not administered to any of the participants, as the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997) cites this subtest as being optional for the elderly. To facilitate administration of the test battery to this particular sample, each administration was performed in two sessions. This allowed for a more flexible schedule to be had that would not interfere with any existing programs or appointments that the participants may have had.

The order of subtest presentation was modified to provide a direct assessment of the short forms. This allowed the removal of the influence of factors such as fatigue

being introduced as a confounding factor on short form performance, caused potentially by the presentation of intervening subtests. There have been findings that changing the test sequence has no effect on the validity of the test scores (Thompson, 1987). Three of the evaluated short forms (short forms #1, 2, and 4) were grouped together due to an overlap in the specific subtests which are contained within these abbreviated forms. The remaining short form (short form #3) was administered separately as it has no overlap with the other three short forms. This yielded two possible orderings of administration for the entire battery. In order to avoid order effects, these two presentation orders were split evenly between the sample, with 50 participants receiving Order A and 50 participants receiving Order B. The final presentation orderings were:

Order A: Vocabulary, Block Design, Arithmetic, Picture Arrangement, Information, Digit Span, Similarities, Picture Completion, Digit-Symbol Coding, and the five remaining subtests.

Order B: Digit Span, Similarities, Picture Completion, Digit-Symbol Coding, Vocabulary, Block Design, Arithmetic, Picture Arrangement, Information, and the five remaining subtests.

Upon completion of the WAIS-III, participants then completed the Shipley Institute of Living Scale (SILS; Shipley, 1940). The SILS is a two-part pencil-and-paper test, composed of a Vocabulary test and an Abstraction test, designed to assess the differential between vocabulary and abstract thinking levels as an index of mental deterioration. The Vocabulary section has the participant choose a single word from four options that has a similar meaning to a stimulus word. The Abstraction section has the

participant fill in blanks to abstract thinking problems. Both halves of the test have a ten-minute time limit each. For those individuals who were physically unable to complete the SILS, their answers were verbalized and marked by the principal investigator.

A modification was made to some of the visual stimuli used in both the Wechsler and the SILS batteries. Pictures and some text were enlarged to facilitate viewing by individuals who reported minor vision problems.

After completion of both batteries, each participant was debriefed and informed that they would be receiving a letter that would outline the general findings of the study. A copy of the debriefing form is given in Appendix I. The tests were then scored by the principal investigator and entered into an ASCII text file in preparation for data analysis. A letter given in Appendix J will be sent to each participant, outlining the general findings of the study when it is complete. A letter will also be sent to the participating centre outlining the findings in a more detailed manner, and this is given in Appendix K.

#### Statistical Protocol

All data was analysed with the SAS statistical computer package v 6.13 (SAS Institute Inc., 1997). For the Wechsler battery each of the subtest scores were converted to age-corrected scale scores, which were then summed to produce the Full-Scale, Verbal, and Performance scores. These scores were then converted to IQ scores for each individual, according to the conversion tables provided in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997). Each of the estimation procedures utilized the age-corrected scale scores in all of their calculations. The Wechsler short-form IQ estimates were formed according to two methods. The first

method was based on the tables found for conversion based on the WAIS-R by the respective authors of the short forms. The only exception was with the Ward et al. (1987) short form, for which no conversion tables were provided by the original authors. For this particular short form the linear equating procedure outlined by Tellegen and Briggs (1967) was used to provide IQ estimates. There are two separate conversion tables for the first short-form (Vocabulary and Block Design; Cyr & Brooker, 1984; Silverstein, 1982), so separate IQ estimates were obtained using both tables to assess their comparability in providing IQ estimates.

The second method of IQ estimation arose both from the use of the linear equating procedure outlined by Tellegen and Briggs (1967) for the Ward et al. (1987) short form and from there being no existing conversion tables for the remaining short-forms with regards to the WAIS-III. In the first scoring procedure mentioned previously, the Ward et al. (1987) estimates are not directly comparable to the WAIS-R-based estimates, as they are derived from the actual WAIS-III scores. Thus, to ensure comparability between all of the short forms, the linear equating procedure was used on all four of the short forms to obtain IQ estimates. With all of the Wechsler-based short forms, linear equating procedures were also used to determine an estimate of Verbal and Performance IQ scores. The IQ estimates from the SILS were formed according to the two formulae outlined previously by Paulson and Lin (1970b) and Zachary et al. (1985).

To satisfy the three evaluation criteria proposed by Resnick and Entin (1971), the estimated IQ scores for each of the four Wechsler-based short forms and the SILS-conversion IQ estimates were correlated with the full-form IQ, and descriptive statistics

(e.g., mean, standard deviation) were obtained for all IQ scores, both full-form and estimated. A classification table was also constructed, which compared the IQ classification (e.g., normal / average, above average, below average) achieved by the full-form IQ with the IQ classification achieved with the short form IQ estimate for each individual participant. The classification system used was the system suggested in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997). A correct classification is one in which the individual is classified identically by both types of form (full-form and short-form), and an incorrect classification occurs when the estimated IQ misclassifies an individual by placing them within a different category than that indicated by their full-form IQ score. No discrimination was made between over- and under-estimation for individuals, although an inspection of the classification tables can indicate if a short form predominantly over- or under-estimates the full-form IQ.

Separate repeated-measure ANCOVAs were performed for both the WAIS-R-based and the linear scaling IQ estimates, with education level acting as the covariate and the repeated factor being the IQ score (actual full form IQ and short form estimated IQ). To investigate any significant main effects, contrasts between the full-form IQ and each of the short-form IQ estimates were evaluated. It should be noted that these contrasts also represent t-tests of the differences between the means of the estimated IQ score and the full-form IQ score. In order to maintain a Type I error rate, the overall alpha of .05 was divided equally among the three repeated-measure ANCOVAs. Thus, each of the ANCOVAs were evaluated at an alpha level of .016. The contrasts for the subtest-only ANCOVAs were evaluated at an alpha level of .004 for each contrast (the alpha level for

the individual ANCOVA divided by the number of contrasts, which is four in this case). For each repeated-measures ANCOVA, the analysis strategy outlined by Keppel (1991) was followed, whereby analysis of significant effects went through three steps. The first step was an investigation of the conventional F-statistic. If this was not significant, the null hypothesis of no difference between the forms was retained. If this F-value was significant, the Geisser-Greenhouse correction was applied. If the F-value was significant under the Geisser-Greenhouse correction, the null hypothesis was rejected. If it was not significant, then the Box correction was applied to the F-statistic. If this Box correction was significant, the null hypothesis was rejected. If this Box correction was not significant, then the null hypothesis was retained.

Reliabilities of each of the short forms were computed according to the formula outlined by Tellegen and Briggs (1967). The correlations among each of the subtests were computed and used in this calculation. The formula also requires the reliability of each of the individual subtests which make up the short form. These reliabilities were taken from the WAIS-III - WMS-III Technical Manual (Psychological Corporation, 1997), which are based on the standardization sample. Validities were computed according to the method outlined by Tellegen and Briggs (1967) for controlling for the common error shared by the short-form estimate and the full-form actual score.

A third repeated-measure ANCOVA was also computed for the Shipley-based IQ estimates, with education again acting as a covariate and the type of SILS estimation method acting as a repeated factor. The alpha level for this ANCOVA was held at .016 to control for Type I error. In the event of a significant effect of estimation method,

contrasts between the SILS IQ estimate and the full-form IQ score were to be computed, with the alpha level being .008 for each contrast (the original alpha of .016 for the overall ANCOVA divided by the number of contrasts, which in this case was two).

To investigate patterns of cognition in this ageing sample, the age range of the sample was divided into five sections: 65 - 69, 70 - 74, 75 - 79, 80 - 85, and 85 and over. The mean full-form Full-Scale IQ, Verbal IQ, Performance IQ, and estimated Full-Scale IQ for each short form were plotted against these groups to determine if the short forms could mimic the ageing pattern in cognition that would be evidenced by the full WAIS-III. A trend analysis was performed on the Full-Scale IQ estimates only, following the procedure of Keppel (1991), using orthogonal contrasts to define the linear trend to be investigated.

The difference between the Verbal IQ and the Performance IQ scores were computed for the full-form Wechsler battery as well as the three short-forms which were comprised of four subtests. The Verbal and Performance scores were computed using the linear equating procedure that was used for the Full-Scale IQ estimation. It was felt that the two-subtest short-form by Silverstein (1982) and Cyr and Brooker (1984) did not contain enough subtests to fully represent the two domains. With regards to the SILS, no conversion formulas exist for calculating separate Verbal and Performance scores. For the three chosen short-forms, the linear equating procedure was used separately to obtain estimates of Verbal and Performance IQ. A frequency table of the absolute differences between the full-form VIQ-PIQ difference and the short-form estimated VIQ-PIQ difference was constructed to examine the accuracy of estimations by the short forms.

## RESULTS

Descriptive Statistics of IQ Scores, Administration Times of Forms, and PsychometricProperties of FormsFull-Form and Full-Scale IQ.

The means and standard deviations for the full-form WAIS-III IQ are given in Table 1, as well as the total administration time for the entire Wechsler battery. The full-form Full-Scale IQ scores were normally distributed, with no evidence of large effects of skewness or kurtosis. The scores ranged from a minimum of 67 to a maximum of 142.

Table 1.

Means and Standard Deviations and Administration Times for Full-Form WAIS-III IQ and Wechsler-Based Short form IQ Estimates.

Form	Mean IQ Score	Standard Dev. IQ	Absolute Difference	Mean Admin. Time	SD Time	% Time Saved	Reliability (Validity) [Correlation]
Full FSIQ	88.78	13.14		79.87	15.21		
SF1aIQ	90.21	12.69	1.43	16.98	4.04	79	.90 (.88) [.90*]
SF1bIQ	102.85	13.79	14.07	16.98	4.04	79	.90 (.88) [.90*]
SF2IQ	93.57	14.01	4.79	31.97	6.14	60	.94 (.94) [.85*]
SF3IQ	87.34	13.93	1.44	19.07	4.72	76	.94 (.74) [.94*]
SF4IQ	90.91	13.48	2.13	30.79	6.44	62	.93 (.92) [.95*]

Note: SF1aIQ = Cyr & Brooker (1984) 2-subtest short form, SF1bIQ = Silverstein (1982) 2-subtest short form, SF2IQ = Silverstein (1982) 4-subtest short form, SF3IQ = Ward et al. (1987) 4-subtest short form, and SF4IQ = Reynolds et al. (1983) 4-subtest short form.

\*  $p < .001$ .

WAIS-R-Based Short-Form Conversion.

The descriptive statistics for all of the conversions based on the WAIS-R short-forms are also given in Table 1. The estimates based on the 2-subtest short-form tables

by Cyr and Brooker (1984) and Silverstein (1982), the 4-subtest short-form by Silverstein (1982), the 4-subtest short-form by Ward et al. (1987) and the 4-subtest short-form by Reynolds et al. (1983) appear to provide accurate estimates of the full-form IQ score. The conversions for Silverstein's (1982) two-subtest short-form appear to grossly overestimate the actual IQ score. Table 1 also lists the absolute difference of the mean estimated IQ from the full-form IQ score. All of the short forms appear to reduce the administration time to less than 50% of the administration time needed for the entire battery. The sixth column in Table 1 provides the percentage of time that each form saves, with respect to the full-form administration. All of the short forms had some difficulty covering the range of the full-form Full-Scale IQ, most notably at the upper range of scores. Both the Silverstein (1982) and the Brooker and Cyr (1986) conversions produced ceiling effects, with maximum obtainable IQ scores being 120 for the Silverstein (1982) 2-subtest short form, and 136 for the Brooker and Cyr (1986) conversions. No noticeable ceiling effect was seen for any of the other Wechsler-based short forms. All of the WAIS-R-based short form Full-Scale IQ estimates were able to match the full-form Full-Scale IQ estimate at the bottom end of the scoring range.

The last column in Table 1 contains the reliabilities and validities for each of the Wechsler-based short forms. These values were calculated using the formula provided by Tellegen and Briggs (1967). To compute the reliability for each short form, test-retest reliabilities were taken from the WAIS-III - WMS-III Technical Manual (Psychological Corporation, 1997). The test-retest reliabilities were averaged across the elderly age groups (i.e., 65-69, 70-74, 75-79, 80-84, 85 and over), and this average was used in the

calculations for the short form reliabilities. Each short form exhibited excellent reliability, with all values ranging between .85 and .94. The validities ranged from .73 for the Ward et al. (1987) short form to .94 for the Silverstein (1982) 4-subtest short form, and are given in Table 1.

#### Linear-Equated Short-Forms.

The means and standard deviations of each of the linear-equated short-form IQ estimates are given in Table 2. All of the linear-equated estimates were normally distributed, with no significant effects of skewness or kurtosis evident. All of these short forms are providing IQ estimates which are quite close to the full-form IQ, with none of the mean differences exceeding two IQ points. As with the WAIS-R-based conversion, the linear-equated estimates were able to match the full-form with regards to capturing the bottom range of scoring, but were unable to do so at the upper end, usually falling short by 20 IQ points. As the formulas for reliability and validity of the short forms do not depend on the estimation method, the reliability and validity values reported in Table 1 apply equally to the linear-equated short forms.

**Table 2.**  
**Means, Standard Deviations, and Classification Percentages of the Linear-Equated Short Forms for Estimating WAIS-III Full-form IQ.**

Form	Mean IQ	SD	Absolute Difference	% Correct	% Incorrect	Correlation with full-form
Full-Form IQ	88.78	13.1				
SF1IQ	89.27	13.8	0.49	58	42	0.90*
SF2IQ	90.02	13.7	1.24	77	23	0.96*
SF3IQ	87.34	13.9	1.44	72	28	0.94*
SF4IQ	90.62	13.9	1.84	70	30	0.95*

Note: SF1IQ = Silverstein (1982) 2-subtest short form, SF2IQ = Silverstein (1982) 4-subtest short form, SF3IQ = Ward et al. (1987) 4-subtest short form, and SF4IQ = Reynolds et al. (1983) 4-subtest short form. \*  $p < .001$ .

### SILS Conversion Estimates of WAIS-III IQ

The mean SILS Vocabulary score was 21.91 ( $SD = 10.60$ ), with a range of scores from 0 to 38. The average administration time for this scale was 7.43 minutes ( $SD = 2.57$ ). For the SILS Abstraction scale the mean score was 10.26 ( $SD = 6.85$ ), with a range of scores from 0 to 34. The mean administration time for this scale was 5.74 minutes ( $SD = 2.66$ ). The SILS showed poor reliability, with split-half alpha for the Vocabulary section being .25, and that for the Abstraction section being .83.

Table 3 provides basic descriptive statistics, administration times, and classification rates for the SILS-conversion IQ estimates. The means of the estimated IQ scores are close to the full-form IQ score, with the Paulson and Lin (1970b) conversion procedure providing a mean estimate of just under 4 IQ points from the full-form Full-Scale IQ. The conversion procedure by Paulson and Lin (1970b) follows the pattern of

the previous short-forms with regards to matching the full form for range of scoring. The Zachary et al. (1985) procedure provided a range of scoring that was similar to that of the full-form, with a minimum of 60 and a maximum of 140.

**Table 3.**  
**Means, Standard Deviations, Administration Times, and Classification Rates for the Shipley IQ Estimates.**

Form	Mean IQ Score	Standard Dev. IQ	Absolute Difference	Mean Admin. Time	SD Time	% Correct Classification	% Incorrect Classification
Full FSIQ	88.78	13.14		79.87	15.2		
SFIQ5	92.43	13.9	3.65	13.27	4.46	38	62
SFIQ6	94.89	19.22	6.11	13.27	4.46	39	61

Note: SFIQ5 = Paulson & Lin (1970b) Shipley Conversion, and SFIQ6 = Zachary et al. (1985) Shipley Conversion.

### Statistical Inferences

#### Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 1.

The first hypothesis was concerned with the correlation of the short-form IQ estimates with the full-form IQ scores, postulating that these correlations would be both high and significant for all of the short forms. The correlation between each of the WAIS-R-based short-form Full-Scale IQ estimates and the full-form Full-Scale IQ are given in the last column of Table 1. As can be seen, each of these short forms pass the first hypothesis, that of a high and significant correlation between the short form estimate and the full-form IQ.

The correlations for each of the linear-equated FSIQ estimates with the full-form Full-Scale IQ are given in Table 2. They are all above .90 in magnitude, and were all

significant at the .05 alpha level. Thus, all of the short-form estimates using linear equating procedures also show support for the first hypothesis.

The SILS estimates showed moderate correlations with the full-form IQ, with a value of .71 for the Paulson and Lin (1970b) conversion and a value of .68 for the Zachary et al. (1985) conversion, both of which were significant at an alpha level of .05. Thus, these conversion estimates for the SILS also found support for the first hypothesis.

#### Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 2.

The second hypothesis was concerned with the comparison of the full-form IQ scores with the short-form estimates, postulating that the IQ estimates provided by the short forms would not be significantly different from the full-form Full-Scale IQ. This was investigated through the use of a repeated-measures ANCOVA, with education entered as a covariate. The results of the repeated-measures ANCOVA are given in Table 4 for the WAIS-R-based IQ estimates. This analysis indicated that education was a significant covariate,  $F(1, 98) = 9.60, p < .01$ . There was a significant effect for type of short form used,  $F(5,490) = 11.54, p < .0001$ , and as a result single-df contrasts of each of the short-form IQ estimates against the full-form IQ were carried out. These contrasts showed that the 2-subtest short-form advocated by Silverstein (1982) and the 4-subtest short-form advocated by Reynolds et al. (1983) were significantly different from the full-form IQ. The  $F$ -value for the Silverstein 2-subtest contrast was 39.59 at (1, 98) degrees of freedom, and the  $F$ -value for the Reynolds et al. (1983) short form was 9.85 at (1,98) degrees of freedom. Thus, with regards to the second hypothesis regarding accuracy in

mean IQ estimate, only the 4-subtest short-forms by Silverstein (1982) and by Ward et al. (1987) were supported. The remaining WAIS-R-based short form estimates, those from the 2-subtest short-form by Silverstein (1982) and the 4-subtest short-form by Reynolds et al. (1983) were not supported.

**Table 4.**  
**Repeated Measures ANCOVA for the Wechsler-Based Short Forms With the WAIS-R Conversion Tables.**

Source	df	Type III SS	Mean Square	F-Value	p-value
Education	1	8696.48	8696.48	9.6	0.002
Error	98	88796.89	906.09		
Form	5	1280.7	256.14	11.54	< .0001
Error	490	10877.74	22.2		
Contrast 1	1	9.63	9.63	0.28	0.6
Contrast 2	1	1428.25	1428.25	39.59	< .0001
Contrast 3	1	177.82	177.82	3.13	0.08
Contrast 4	1	70.25	70.25	3.03	0.08
Contrast 5	1	180.56	180.56	9.85	0.002

Note: Contrast 1: Full-Form IQ vs Cyr and Brooker (1984) 2-subtest short form, Contrast 2: Full-Form IQ vs. Silverstein (1982) 2-subtest short form, Contrast 3: Full-Form IQ vs Silverstein (1982) 4-subtest short form, Contrast 4: Full-Form IQ vs Ward et al. (1987) 4-subtest short form, and Contrast 5: Full-Form IQ vs Reynolds et al. (1983) 4-subtest short form.

The repeated-measures ANCOVA for the full-form IQ scores and the linear-

equated IQ estimates are provided in Table 5. This analysis again had education emerge as a significant covariate,  $F(1, 98) = 8.71, p < .01$ . The  $F$ -value for the main effect of type of test form was not significant,  $F(4,392) = 3.48$ , at the adjusted alpha level according to the adjusted Greenhouse-Geisser probability level. As there was no significant main effect for type of test form, no single-df contrasts were performed. This is indicative of the linear-equating procedure being able to provide more consistent estimates with all of the Wechsler-based short forms than the older WAIS-R-based conversion tables.

**Table 5.**  
Repeated Measures ANCOVA for the Linear-Equated Wechsler-Based WAIS-III IQ Estimates with Education Level as Covariate.

Source	df	Type III SS	Mean Square	F-Value	p-value
Education	1	7019.43	7019.43	8.71	0.004
Error	98	78961.47	805.73		
Form	4	244.36	61.09	3.48	0.03
Error	392	6878.68	17.55		

A final repeated-measures ANCOVA was performed for the SILS-conversion estimates of WAIS-III IQ, with education again entered as a covariate. This factor emerged as a significant covariate,  $F(1, 98) = 9.52, p < .01$ . There was no significant effect for the type of short form, indicating that both of the estimation methods for the

SILS provided mean estimates that were similar to the mean full-form WAIS-III IQ. The results of the ANCOVA are given in Table 6.

**Table 6.**  
**Repeated-Measures ANCOVA for the SILS-Conversion IQ Estimates with Education Level as Covariate.**

Source	df	Type III SS	Mean Square	F-Value	p-value
Education	1	5312.65	5312.65	9.52	0.003
Error	98	54687.82	558.04		
Form	2	45.64	22.81	0.35	0.7
Error	196	12739.51	64.99		

### Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 3.

The third hypothesis was concerned with the classification accuracy of each of the short forms. A classification table was constructed by converting the individual IQ scores (full-form and short-form estimates) into qualitative categories as defined by the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997). A correct classification is one where the individual is placed within the same qualitative category by both the full-form and the short-form. The classification accuracy for each of the WAIS-R-based estimates of full-form IQ are reported in Table 7, given as the percentage of correct and incorrect classifications in the sample. As can be seen, for most of the WAIS-R-based short-form estimates, the misclassification rate is routinely over 50%.

The exceptions were the 2-subtest short-form conversion by Brooker and Cyr (1986), which misclassified only 43% of the participants, and the Ward et al. (1987) 4-subtest short-form which misclassified 28% of the participants. The superiority of the Ward et al. (1987) short-form should not be taken as definitive, since it was noted earlier that conversion tables for this short form are not available for the WAIS-R and, as such, a linear equating procedure was used on the data from the current sample to generate the IQ estimates. The classification table also indicated that the Brooker and Cyr (1986) 2-subtest conversion predominantly underestimated FSIQ, the Silverstein (1982) 2-subtest conversion showed fairly even under- and over-estimation, and the Silverstein (1982) and Reynolds et al. (1983) 4-subtest conversions showed predominantly overestimation. As the Ward et al. (1987) estimates were obtained through a linear equating procedure on the current data, a comment on their estimation accuracy in relation to the conversions based on the WAIS-R is unwarranted. On the whole, the WAIS-R-based conversions can be regarded as having failed the third hypothesis.

**Table 7.**

**Classification Accuracy and Misclassification of the WAIS-R-Based IQ Estimation Procedures.**

Short Form	Correct Classification	Incorrect Classification
SF1a	57	43
SF1b	17	83
SF2	7	93
SF3	72	28
SF4	67	33

Note: SF1a = Cyr and Brooker (1984) 2-subtest, SF1b = Silverstein (1982) 2-subtest, SF2 = Silverstein (1982) 4-subtest, SF3 = Ward et al. (1987) 4-subtest, and SF4 = Reynolds et al. (1983) 4-subtest.

Table 2 shows the percentage of correct and incorrect IQ classifications, using the same classification system as with the WAIS-R-based short-form estimates, for the linear-equated short-form IQ estimates. The percentage of correct classifications has greatly increased for many of the short forms in comparison to those for their respective WAIS-R-conversion counterparts. The only exception is the 4-subtest short-form by Ward et al. (1987), for which no WAIS-R conversion tables were able to be obtained.

The majority of short forms using the linear-equating procedure tended to overestimate the actual IQ score, the exception being the 4-subtest short-form by Ward et al. (1987) which showed an underestimation. Given the high amount of misclassification that is still present, evidenced in Table 2, the third hypothesis can be regarded as being failed by the linear-equated short forms.

The SILS IQ estimates showed a high degree of misclassification, presented in Table 3. An inspection of the classification matrix showed that the SILS conversion formulae tended to overestimate the actual IQ level of the individual. To investigate the accuracy of the Shipley estimation procedures on an individual level, the frequencies of absolute differences of the full-form Full-Scale IQ from each of the SILS-conversion procedures were calculated and are presented in Table 8. The SILS estimates performed very poorly on an individual level, with the majority of estimated scores being different from the actual IQ scores by at least 8 IQ points.

**Table 8.**  
**Frequency Table of Absolute Differences Between Full-Form Full-Scale IQ and**  
**SILS Conversion Full-Scale IQ Estimates.**

Difference	Paulson and Lin (1970b)	Zachary et al. (1985)
0	5	5
1	6	3
2	7	4
3	4	9
4	8	2
5	9	2
6	2	9
7	6	4
8	7	6
9 or more	46	56

**Comparison of Short-Form IQ Estimates with Full-Form IQ -- Hypothesis 3a.**

This hypothesis was concerned with the relative efficiency of the four-subtest short-forms versus the two-subtest short-forms, and postulated that those short forms composed of four subtests would outperform the short form composed of two subtests. From the results of the previous sections it appears that all of the Wechsler-based short forms, both 2- and 4-subtest versions, are providing accurate Full-Scale IQ estimates. As a result the assessment of these short forms in relation to Hypothesis 3a cannot be answered based on these findings alone. The reliability values of the 4-subtest short forms are generally higher than the 2-subtest short forms, as are the validity values. The exception for the validity values is the Ward et al. (1987) 4-subtest short form, and with

the correlations it is the WAIS-R-based Silverstein (1982) 4-subtest short-form. Looking at the linear-equated estimates, the 2-subtest short-form by Cyr and Brooker (1984) provides a closer mean estimate of Full-Scale IQ, but it also has the smallest number of correct qualitative classifications, and the lowest correlation with the full-form Full-Scale IQ. The Silverstein (1982) 4-subtest short-form is now evidencing the highest correlation with the full-form Full-Scale IQ. Based on these results, there is support for Hypothesis 3a that the 4-subtest short forms are superior to the 2-subtest short forms.

Comparison of Linear-Equated Short-Form Full-Scale IQ With WAIS-R-Based-Conversion Full-Scale IQ -- Hypotheses 4 and 5.

The fourth and fifth hypotheses proposed that the IQ estimates provided by the linear-equated estimation procedures would be more accurate than those provided by those based on the WAIS-R conversion tables. To assess the utility of the older WAIS-R conversions against the linear equating procedure, several empirical and statistical methods were utilized. An empirical investigation of the mean estimates showed that, while both estimation procedures tended to overestimate the actual Full-Scale IQ, the linear equating procedure yielded mean estimates which were closer to the actual Full-Scale IQ than the WAIS-R conversion tables. An empirical investigation of correct qualitative intelligence classification rates also supported the superiority of the linear-equating estimation procedure.

The fourth hypothesis postulated that the correlations between the full-form IQ and the linear-equated estimates would be both higher and significantly different from the correlations of the full-form IQ with the WAIS-R-based estimates. Consistent with this

hypothesis, the correlations for each of the estimates, WAIS-R conversion and linear equating, were tested for significant differences according to the Hotelling-Williams test (Darlington & Carlson, 1987). The results showed that the correlations between FSIQ and the estimated FSIQ using the two conversion procedures were significantly different for the Silverstein (1982) 4-subtest short form ( $Z = 8.93$ ) and the Silverstein (1982) 2-subtest conversion tables ( $Z = 5.90$ ), at an alpha level of .05. For these short forms, the linear-equated correlation with FSIQ was higher than the WAIS-R conversion correlation with FSIQ. For the remaining short forms, the correlations were not statistically different from each other.

The fifth hypothesis stated that the difference between the linear-equated and WAIS-R-based estimates of IQ would favour the linear-equated estimates, by showing them to be significantly lower than the WAIS-R-based estimates. The t-tests comparing the mean estimates of the linear-equating procedure to the mean estimates of the WAIS-R-conversion procedures supported this hypothesis, as all of the t-values were negative and all were significant at the .05 alpha level. The largest mean difference was that between the Silverstein 2-subtest estimates, with a mean difference of 13.58 points. The mean difference of the Full-Scale IQ estimates between the two conversion procedures for all of the short forms except the Ward et al. (1987) short form are given in Table 9, along with the associated t-values, degrees of freedom, and probability levels.

**Table 9.**  
**T-Tests of the Difference Between the Linear-Equated Full-Scale IQ Estimates and the WAIS-R Conversion Full-Scale IQ Estimates.**

Short Form	Mean	Std Dev	df	t-value
Silverstein (1982) 2-subtest				
Brooker & Cyr (1982)	- 0.94	1.18	99	- 8.00*
Silverstein (1982)	-13.58	0.98	99	-160.30*
Silverstein (1982) 4-subtest	-3.55	6.91	99	- 5.97*
Reynolds et al. (1983) 4-subtest	-0.29	0.59	99	- 5.71*

\*  $p < .0001$

Together with the empirical observations from the classification rates, and the added statistical evidence from the comparison of correlations with the full-form Full-Scale IQ and the t-tests favouring the linear equating procedure, the linear equating procedure can be said to show less over-estimation than the WAIS-R conversion procedure, and to be more accurate in their estimation.

As the linear equating procedure appeared to provide more accurate estimations of Full Scale IQ, it is of some interest to determine the accuracy of the estimation procedure. To examine this, the absolute difference between the full-form FSIQ and the estimated IQ for each of the linear-equated short forms was tabulated for the entire sample, and are presented in Table 10.

**Table 10.**  
**Frequency Distribution of Absolute Differences of Linear-Equated Full-Scale IQ Estimates from the Full-Form Full-Scale IQ.**

Difference	Silverstein (1982) 2-subtest IQ	Silverstein (1982) 4-subtest IQ	Ward et al. (1987) 4-subtest IQ	Reynolds et al. (1983) 4-subtest IQ
0	17	22	14	19
1	5	21	18	21
2	12	20	15	15
3	14	14	9	9
4	13	9	12	5
5	8	3	12	12
6	9	1	3	6
7	3	4	6	1
8	6	1	2	5
9 or more	13	5	9	7

As can be seen from the table, the 4-subtest short-form advocated by Silverstein (1982) provides a greater number of estimates which are within the age-corrected 3-point standard error of measurement that is associated with the Full-Scale IQ. The standard error of measurement indicates how much an individual's IQ score would vary from their true score (Kaplan & Saccuzzo, 1993), and thus can provide a measure of the degree of accuracy of the estimation.

General Cognitive Trends in the WAIS-III

Age Trends in Cognition -- Hypothesis 6.

The sixth hypothesis postulated that the linear-equated IQ estimates from the

Wechsler short forms composed of four subtests would show a decreasing linear trend. The entire sample was split into five age groups: 65 - 69, 70 - 74, 75 - 79, 80 - 84, and 85 and above. For each age group the mean Full-Scale IQ, Verbal IQ, Performance IQ, linear-equated Full-Scale IQ estimates, and SILS-Conversion Full-Scale IQ estimates were computed. These values are presented in Table 11.

**Table 11.**  
**Sample Sizes, Means, and Standard Deviations of Full-Scale IQ, Verbal IQ, Performance IQ, Linear-Equated Full-Scale IQ Estimates, and SILS-Conversion Full-Scale IQ Estimates for Each Age Group.**

	FSIQ	VIQ	PIQ	SF1	SF2	SF3	SF4	SF5	SF6
65 - 69 (n=9)									
Mean	90.10	90.10	92.00	94.74	91.37	88.55	91.96	98.24	89.96
Std Dev	17.50	15.28	18.97	19.71	18.01	18.15	18.19	13.07	14.45
70 - 74 (n=14)									
Mean	81.07	85.78	78.50	84.44	83.70	81.65	83.36	93.54	88.22
Std Dev	8.33	7.52	9.50	8.71	9.58	9.78	11.12	10.90	12.27
75 - 79 (n=23)									
Mean	92.30	93.39	92.48	91.63	93.25	91.18	94.59	95.01	93.86
Std Dev	14.66	12.92	16.06	15.28	15.07	15.38	14.84	15.16	19.35
80 - 84 (n=18)									
Mean	93.06	98.83	88.22	92.98	94.54	90.73	94.31	95.09	99.06
Std Dev	15.24	14.36	17.12	16.19	16.18	15.28	15.88	17.78	24.74
85 + (n=36)									
Mean									
Std Dev	87.06	91.47	83.44	86.41	87.81	85.09	88.72	87.57	97.28
	9.99	10.52	10.13	10.61	10.79	11.90	10.93	11.12	19.20

**Note:** FSIQ = Full-Scale IQ, VIQ = Verbal IQ, PIQ = Performance IQ, SF1 = Silverstein (1982) 2-subtest short-form, SF2 = Silverstein (1982) 4-subtest short-form, SF3 = Ward et al. (1987) 4-subtest short-form, SF4 = Reynolds et al. (1983) short-form, SF5 = Paulson and Lin (1970b) SILS estimate, and SF6 = Zachary et al. (1985) SILS estimate.

The plots for the means of the five age groups for the Full-Form IQ, the Verbal IQ, and the Performance IQ are given in Figures 1, 2, and 3, respectively. All three of the figures show a distinctly non-linear trend, which is most noticeable for the plot of the Performance IQ.

**Figure 1. Plot of Full-Form Wechsler IQ versus age.**

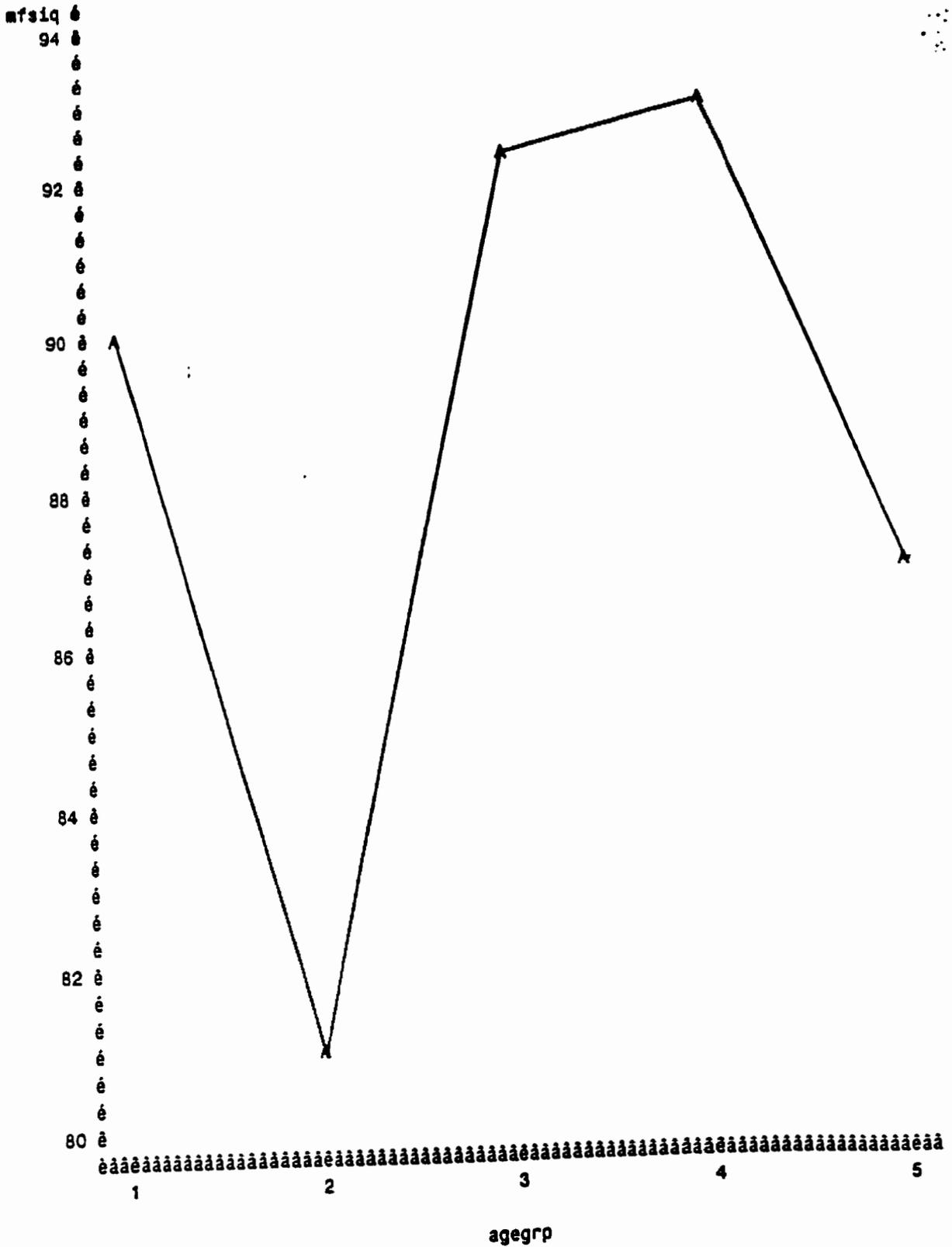


Figure 2. Plot of Verbal Wechsler IQ versus age.

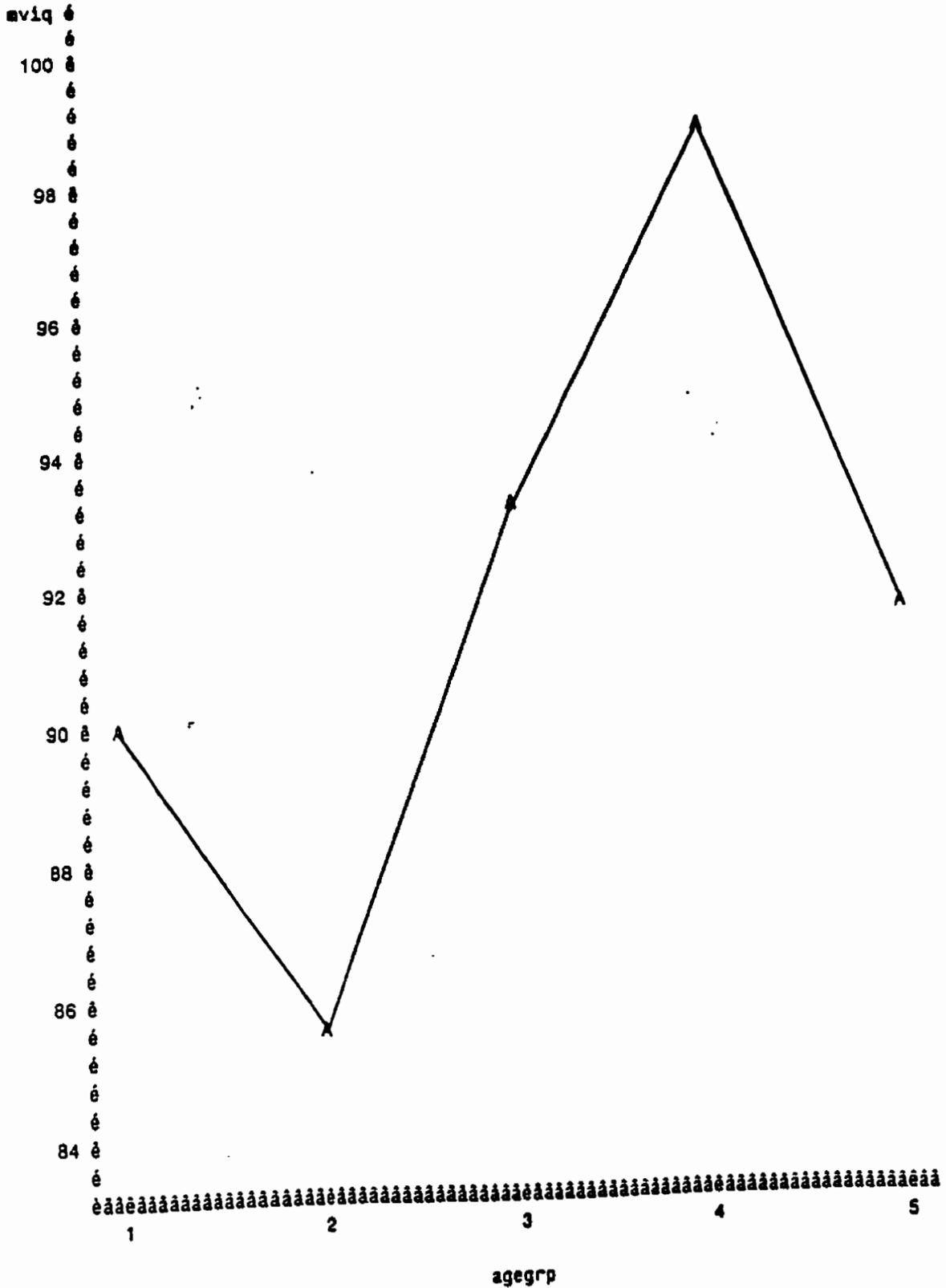
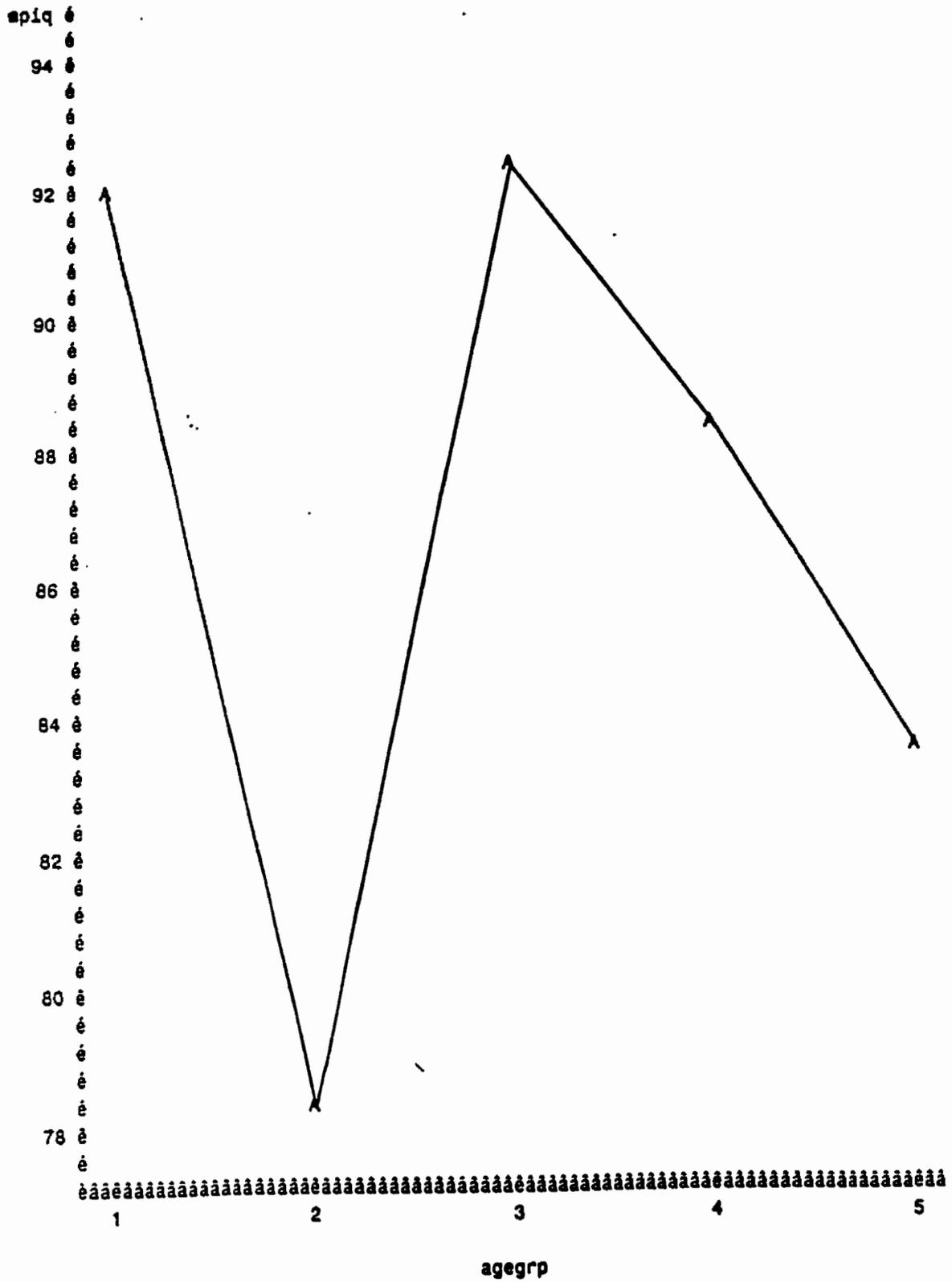
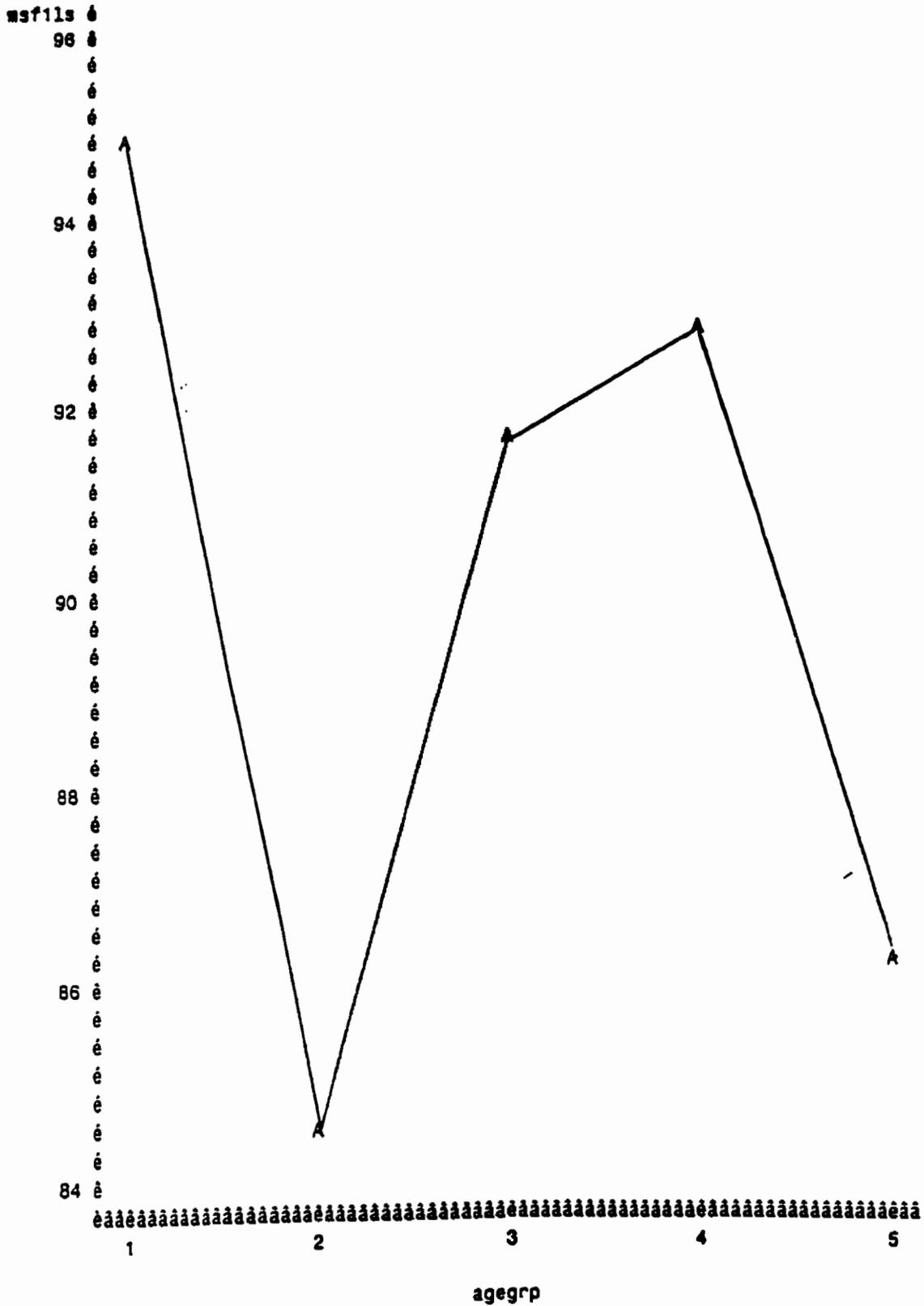


Figure 3. Plot of Performance Wechsler IQ versus age.



With the exception of the 70-74 age group scores, the plots for FSIQ and VIQ show stability or increases up until the 85 and over age group. For the PIQ, there is a slight increase followed by a decrease in scores as age increases. Plots for the linear-equated full-form IQ estimates for the short forms and the SILS conversion estimates are given in Figures 4 through 9. Each of the linear-equated short form estimates appear to show the same trends as the full-form IQ scores across the age groups, with a similar dip in scores at the 70-74 age group. The SILS conversion estimates do not follow the full-form IQ pattern. The Paulson and Lin (1970b) conversion produces a steadily declining pattern in the full-form IQ estimates, and the Zachary et al. (1985) conversion procedure produces a generally increasing trend in full-form IQ estimates.

**Figure 4.** Plot of Linear Equated IQ Estimates for Silverstein's (1982) 2-Subtest Short Form.



**Figure 5.** Plot of Linear Equated IQ Estimates for Silverstein's (1982) 4-Subtest Short Form.

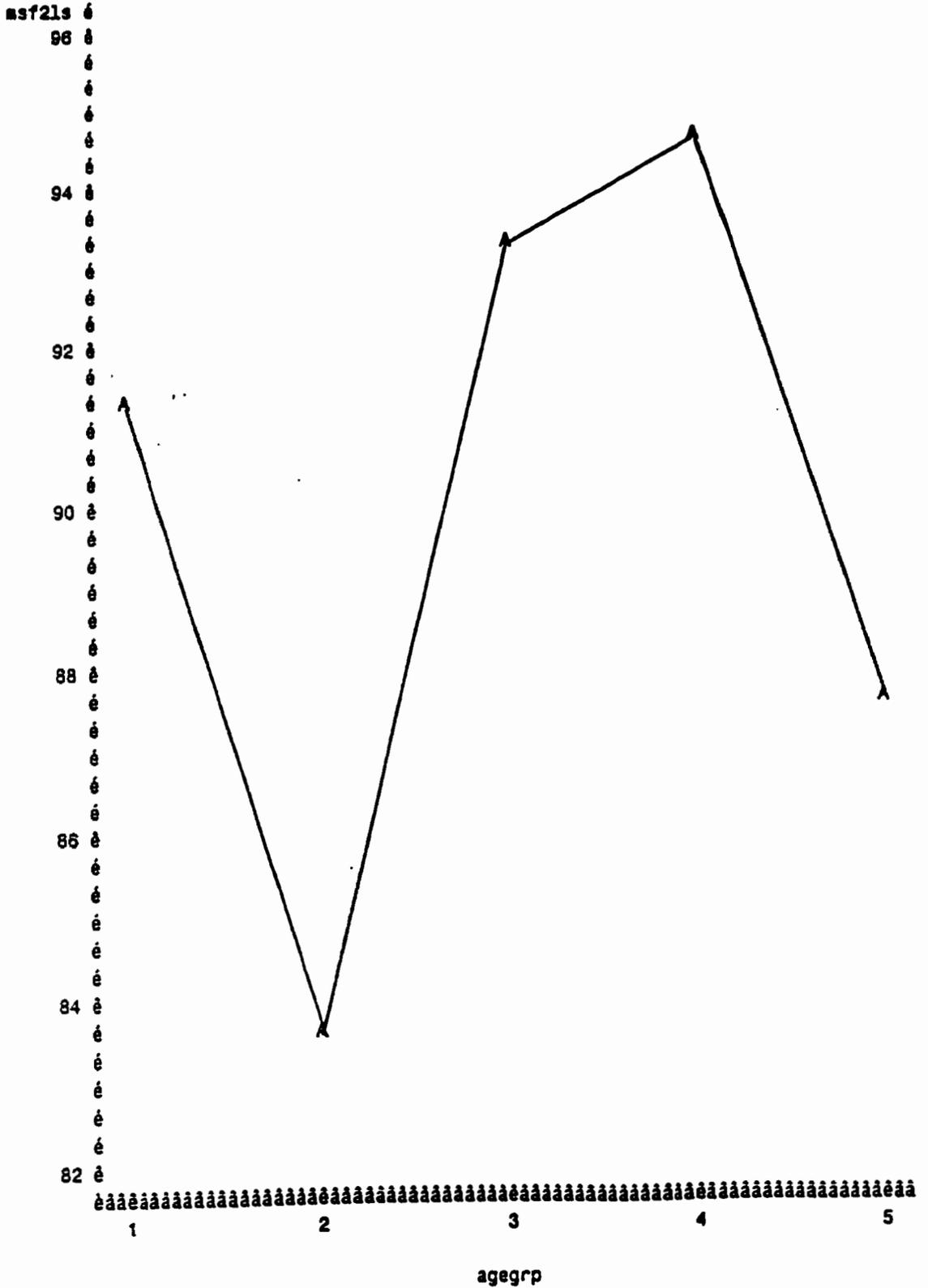
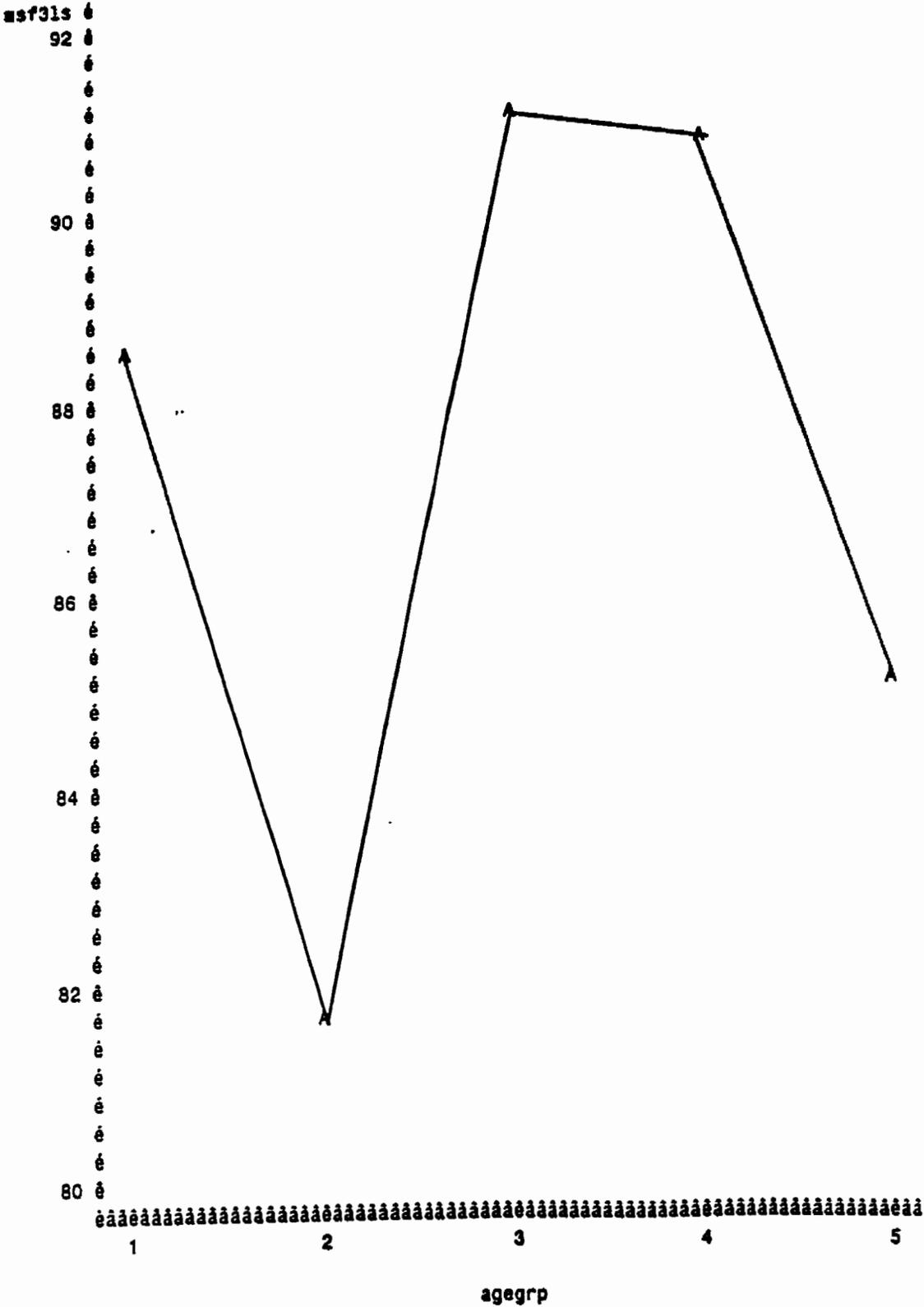
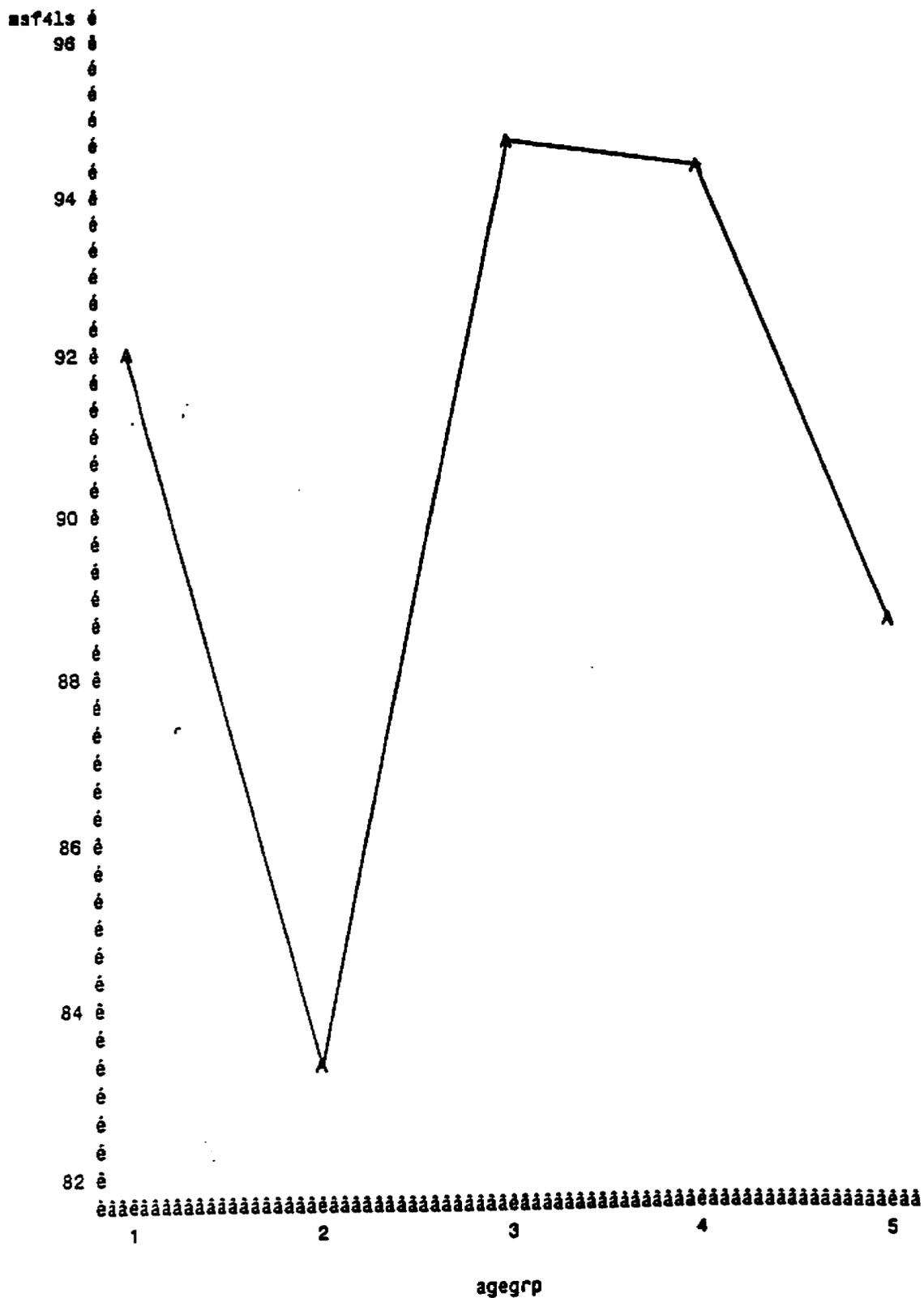


Figure 6. Plot of Linear Equated IQ Estimates for Ward et al. (1987) 4-Subtest Short Form.



**Figure 7. Plot of Linear Equated IQ Estimates for Reynolds et al. (1983) 4-Subtest Short Form.**



**Figure 8. Plot of Estimated IQ from Paulson and Lin (1970b) SILS Conversion Formula.**

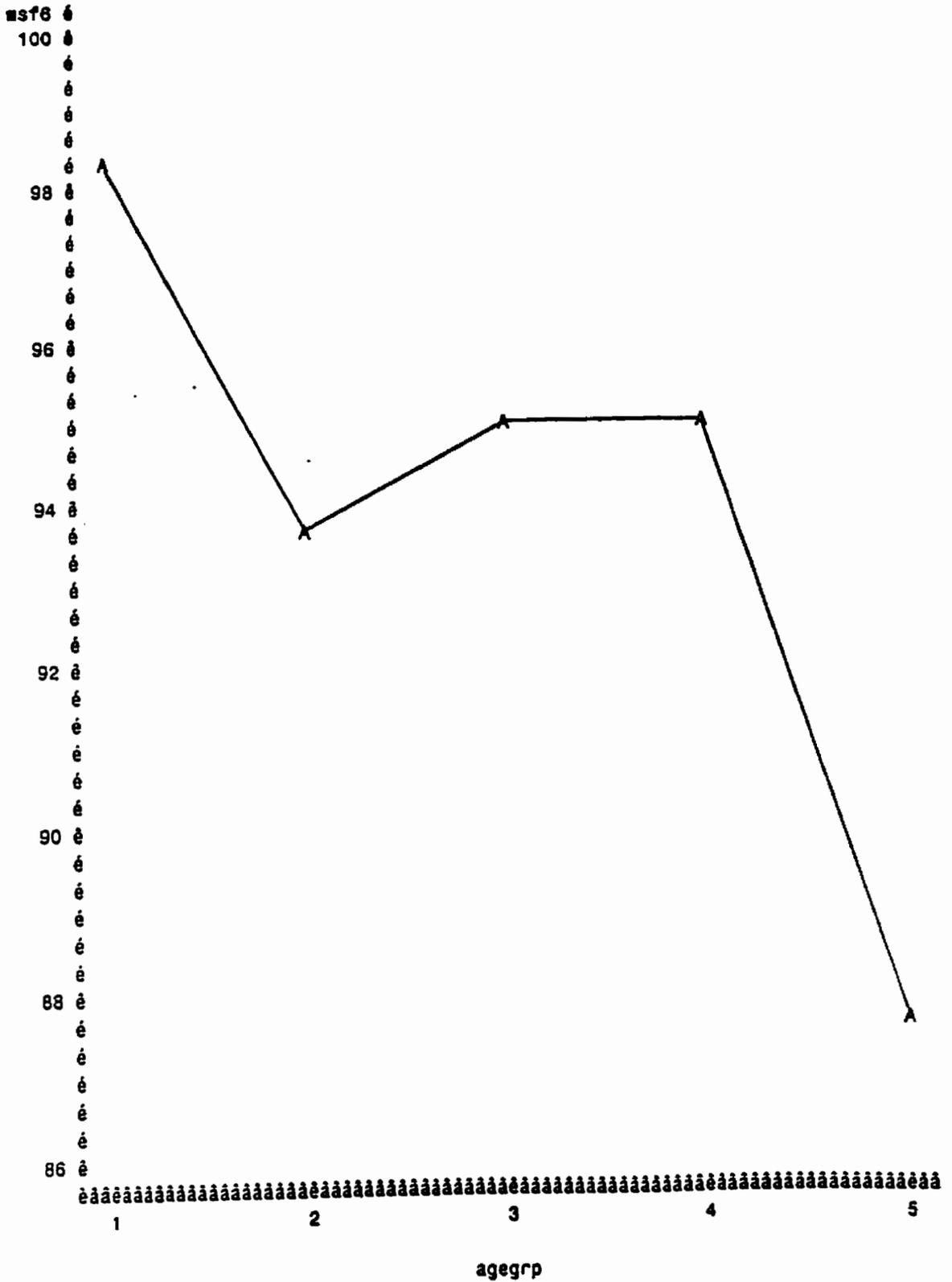
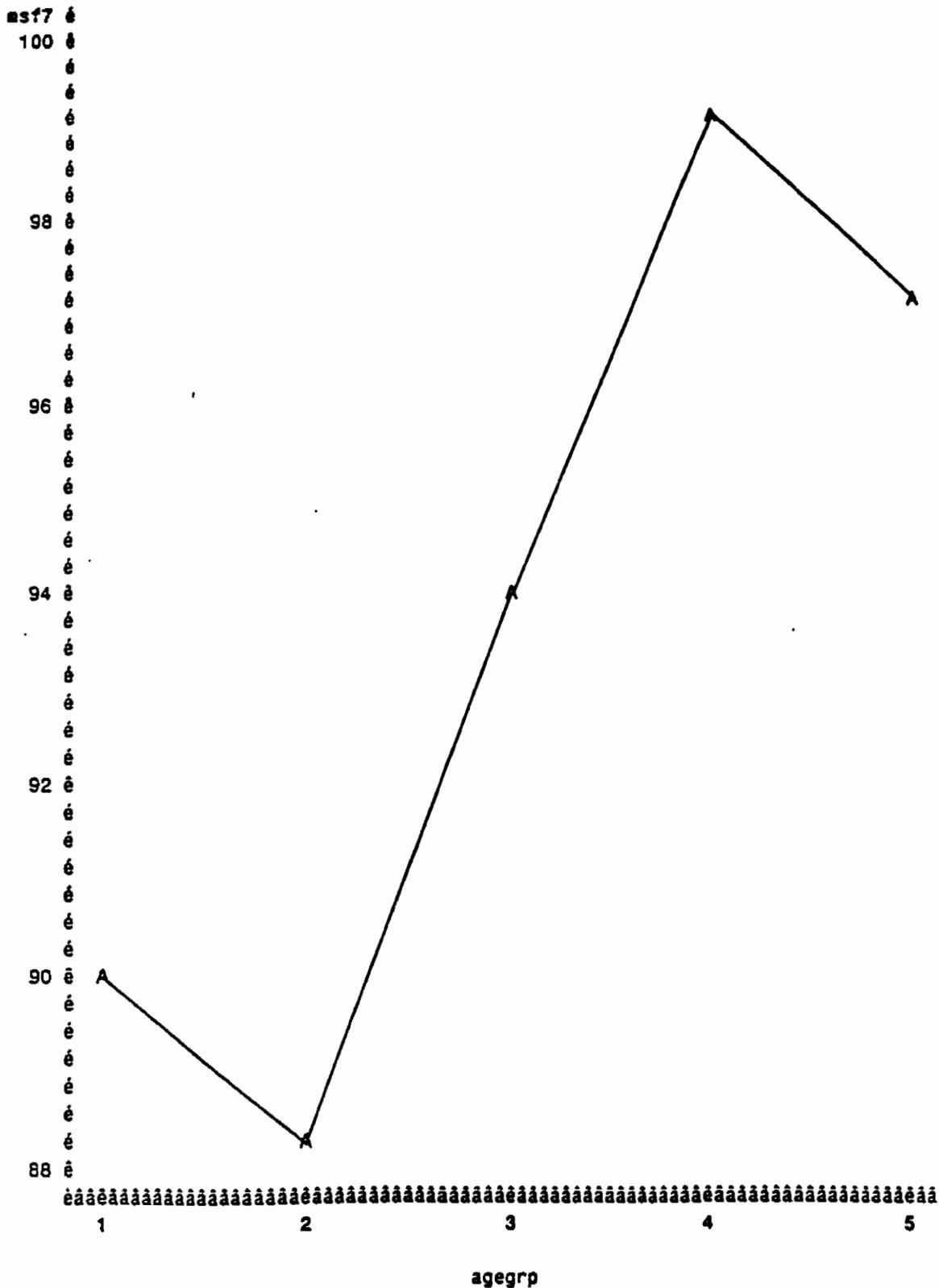


Figure 9. Plot of IQ Estimates from Zachary et al. (1985) SILS Conversion Formula.



A trend analysis (Glass & Hopkins, 1996; Keppel, 1991) was attempted on the data, using age grouping as a classification variable. As a preliminary step to determine if there were any significant differences between the age groups, an omnibus  $F$ -test was computed for each of the full form and short form Full-Scale IQ estimates. These omnibus tests were nonsignificant for all of the Full-Scale IQ scores, with  $F$ -values ranging from 1.6 to 2.4 (evaluated at 4, 95 df). As a result, any observed differences between the age groups is not due to the presence of a trend but is most likely the result of normal variation. Thus, Hypothesis H6 was not supported, and no further analyses of trend were carried out.

#### VIQ-PIQ Differences for Full-Form and 4-Subtest Short Forms -- Hypothesis 7.

The seventh hypothesis proposed that the number of significant absolute VIQ-PIQ differences evidenced by those Wechsler-based short forms composed of four subtests would be equivalent to the number of significant absolute VIQ-PIQ differences evidenced by the full-form. The absolute mean differences between VIQ and PIQ scores for the full-form Wechsler and the three short forms composed of 4 subtests are given in Table 12. The distributions of these differences for the full-form values and the short-form estimated values were consistent with a normal distribution via an examination of skewness and kurtosis values, as well as normal probability plots. For the full-form battery, the range of VIQ-PIQ differences was from 17 to 44 points, while the linear scaling procedure produced differences that ranged from 28 to 53 points.

**Table 12.**  
**Verbal and Performance IQ Differences with the Full-Form and Short-Form Wechsler Scales.**

Difference	Full-Form Mean (SD)	SF2 Mean (SD)	SF3 Mean (SD)	SF4 Mean (SD)
VIQ - PIQ	5.86 (10.92)	3.02 (13.09)	9.42 (13.31)	3.98 (13.03)
# over 8.76	51	46	55	47

Note: SF2 = Silverstein (1982) 4-subtest short-form, SF3 = Ward et al. (1987) 4-subtest short-form, and SF4 = Reynolds et al. (1983) 4-subtest short-form.

Table B.1 in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997) indicates that VIQ-PIQ differences greater than 8.76 are significantly different at the  $\alpha=.05$  level. For both the full-form and the short-form differences, the number of scores that fell outside of this range were tabulated, and are also presented in Table 12. From these values, there is support for the seventh hypothesis for all of the 4-subtest short forms that use a linear-equation procedure to obtain VIQ and PIQ estimates. In order to provide a clearer picture of the accuracy of the four-subtest short forms in simulating the characteristics of the full-form battery, frequency counts for absolute values of VIQ-PIQ differences were calculated, and are presented in Table 13. The short form by Ward et al. (1987) appears to provide a distribution that closely matches that of the full-form battery.

**Table 13.**  
**Frequency Counts of Absolute VIQ-PIQ Differences for Full-Form and Four-Subtest Short Form Estimates.**

Difference	Full-Form VIQ - PIQ	Silverstein (1982) VIQ - PIQ	Ward et al. (1987) VIQ - PIQ	Reynolds et al. (1983) VIQ - PIQ
0	2	7	2	11
1	3	0	3	0
2	7	0	4	0
3	6	17	6	18
4	6	0	7	0
5	6	2	5	3
6	7	12	5	10
7	5	0	11	0
8	7	15	1	8
9 or more	51	47	56	50

An investigation of full-form and estimated VIQ and PIQ scores for each individual was undertaken in order to examine estimation errors that might arise from using these short forms to estimate the VIQ and PIQ values. The absolute errors in prediction for each of the three four-subtest short forms for the Verbal and Performance IQ scores were calculated, and are presented in Table 14. From these values it appears that the four-subtest short form by Silverstein (1982) provides more accurate estimations, with only 11 estimates deviating from the actual VIQ by more than 6 points. Accuracy on the Performance IQ estimation was not as good, with 26 estimates deviating from the actual PIQ by more than 6 points.

Table 14.

**Frequency Distribution of Absolute Differences of 4-Subtest Linear-Equated Verbal and Performance IQ Estimates from the Full-Form Verbal and Performance IQ.**

Difference	Silverstein (1982) 4-subtest		Ward et al. (1987) 4-subtest		Reynolds et al. (1983) 4-subtest	
	VIQ	PIQ	VIQ	PIQ	VIQ	PIQ
0	12	15	10	15	17	15
1	8	8	15	12	15	8
2	19	15	8	12	9	15
3	22	7	9	10	12	7
4	15	9	13	12	8	9
5	8	6	10	9	10	6
6	5	14	3	3	11	14
7	5	6	8	5	4	6
8	2	6	7	10	2	6
9 or more	4	14	17	12	12	14

## DISCUSSION

This study found varying degrees of support for the short forms that were studied. All of the short forms studied fit within the criteria for a valid test as proposed by Conrad (1931). All of the short forms provided IQ scores which are of equal units, were able to measure general as well as specific facets of intelligence (although the latter half of that statement is applicable to the 4-subtest short forms only), exhibited adequate reliability and validity, and were balanced with respect to assessing both verbal and non-verbal factors equally. One criterion that may not be applicable is that of the inclusion of an adequate number of tasks, but given that the purpose of a short form is to cut down on administration time, it appears that a trade-off between time and sampling of the cognitive domain is necessary when a short form is used, a conclusion that is in keeping with that made by Doppelt (1956). A second criterion that may not be applicable is that of a zero point for the scale, since even those individuals who score zero on the scales will receive a minimal IQ score.

With regards to ease of administration and scoring, all of the short forms were easy to administer and objective in their scoring. There were some problems observed with the Picture Arrangement subtest, which is used in the Silverstein (1982) 4-subtest short-form and the Reynolds et al. (1983) short-form. Many of the elderly found the small stimulus pictures for this subtest hard to see. Both the Vocabulary and the Information subtests require some subjective judgement to be made by the examiner with regards to the responses, but the examples given in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997) were more than enough to enable

quick and easy scoring of the responses. The Ward et al. (1987) short-form utilizes the Digit-Symbol Coding subtest, which did pose problems for some members of the sample. These problems were mainly centered around the elderly forgetting the instructions, and in some cases motor tremors made writing within the response boxes difficult. From a practical standpoint, the Ward et al. (1987) short form is the easiest short-form to be used, as three of its four subtests can be answered with verbal answers, whereas the other 4-subtest short forms require some degree of motor activity for their two Performance subtests. As far as ease of use for the examiner, the short form by Ward et al. (1987) was easier to use due to a minimum of materials outside of the Record Booklet and the Stimulus Book being required. The other short forms, which utilized the Block Design subtest and the Picture Arrangement subtests (with the exception of the 2-subtest short-form by Silverstein, 1982), required some extra manipulation by the experimenter by requiring more stimulus materials. There was a slight trade-off with regards to scoring, since the Silverstein (1982) and the Reynolds et al. (1983) short-forms can be scored using the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997), but the Ward et al. (1987) short-form requires the separate scoring template for the Digit-Symbol Coding subtest. As well, three of the subtests included in the Ward et al. (1987) short form are untimed tests, which makes progression through the subtests rather rapid. Thus, for ease of administration, the Ward et al. (1987) short-form is the most user-friendly.

Consistent with the findings of Paolo and Ryan (1993) for the WAIS-R, many of the elderly in this study found the WAIS-III enjoyable. The various subtests were seen as

“interesting”, and most of the individuals regarded the test as a challenge and a new experience. The division of the testing procedure into two one-hour sessions was very agreeable to many participants, and one hour appeared to be the limit for a period of sustained concentration for this sample. When testing in a clinical setting, especially when using the entire battery, such a procedure should be considered by the clinician so as to avoid detrimental fatigue and motivation effects. Of course, when using the short forms, they were all timed as taking 30 minutes or less, in which case these factors would not be entirely applicable to the testing situation.

The four Wechsler-based short-forms evaluated received mixed support for their utility in estimating full-form WAIS-III IQ scores when using the older WAIS-R conversion tables. The excellent reliabilities and adequate validities of each of these Wechsler-based short-forms attest to the ability of these short forms to assess the general construct of intelligence. All of the short forms exhibited reductions in testing time that were over 50%, a valuable criterion for the creation of a short form (Levy, 1968). The short form that had the shortest administration time was the 2-subtest short-form by Silverstein (1982), which required just under 17 minutes, and the longest was the 4-subtest short-form by Silverstein (1982), requiring just under 31 minutes. The administration time for the 4-subtest short-form of Ward et al. (1987) is in agreement with their original findings, as this short form showed an administration time that was comparable to the 2-subtest short-form advocated by Silverstein (1982).

Given the highly congruent reliabilities for all of the short forms investigated, the use of a 4-subtest short-form versus a 2-subtest short-form would hold more clinical

appeal due to an increased sampling from the cognitive domain. The short forms composed of four subtests exhibited higher reliabilities than the 2-subtest short-form, with all of the reliabilities ranging from .90 to .94. Both the Silverstein (1982) 4-subtest and the Ward et al. (1987) 4-subtest short-form exhibited the highest reliabilities, being matched at .94. With regards to validity, the Ward et al. (1987) 4-subtest short-form performed the worst, with a validity value of .73. The Silverstein (1982) 4-subtest short-form performed the best, with a validity of .94, and the Reynolds et al. (1983) 4-subtest short-form performed second-best, with a validity of .92. Based on these findings, the Silverstein (1982) 4-subtest short-form exhibits superior performance on some of the essential psychometric criteria for a test of intelligence.

For those short forms that used the WAIS-R conversion tables, all of them met the first hypothesis regarding high and significant correlations with the Full-Scale IQ from the full-form battery. These correlations are listed in Table 1, and all were significant at the .05 level. For the second hypothesis addressing the similarity in mean estimates of Full-Scale IQ, the only short forms to pass this hypothesis were the Cyr and Brooker (1984) 2-subtest short-form conversion, the Silverstein (1982) 4-subtest short-form conversion, and the Ward et al. (1987) 4-subtest short-form conversion. This can be seen in the contrasts for the repeated-measure ANCOVA in Table 4. None of the short forms showed support for the third hypothesis regarding the accuracy in qualitative classification. This can be seen in Table 7. Such findings are not surprising, given the failure of many evaluated short forms to meet all three of these evaluation criteria, most notably the third criterion regarding classification accuracy (Silverstein, 1985a, 1990a).

Of all the short forms using WAIS-R conversion tables, the 2-subtest short-form by Silverstein (1982) met the fewest of the three criteria proposed by Resnick and Entin (1971) by only showing a high and significant correlation with the full form IQ score. This short form exhibited IQ estimates that were substantially higher than the full-form IQ score, and also showed a high amount of misclassification according to the qualitative groupings used in the WAIS-III Administration and Scoring Manual (Psychological Corporation, 1997). Further support for the undesirability of this short form comes from the significant contrast with the repeated-measure ANCOVA, indicating that this particular short form provided estimates which differed from the full-form Full-Scale IQ, and which could not be attributed to normal variation. Similar findings regarding this short form have been found by Crawford et al. (1992). The remaining 4-subtest short-forms performed much better, generally meeting two of the three evaluation criteria -- a high correlation between the short-form estimates and the full-form IQ, and providing accurate estimates of the full-form IQ score. Two of the short forms, the Ward et al. (1987) and the Reynolds et al. (1983) 4-subtest short-forms, showed mean estimates that were within the 3-IQ point standard error of measurement for the full-form scale. All of the chosen short-form conversions performed very poorly with regards to classification of individuals, and exhibited classification rates that are comparable to those found in other studies (e.g., Robiner et al., 1988). The exception to this being the Ward et al. (1987) short-form for which a different scoring procedure was used to obtain the Full-Scale IQ estimates, and is thus not directly comparable to the other WAIS-R-based estimation procedures.

The use of linear equating procedures fared substantially better than their corresponding WAIS-R-based estimates. Similar to the WAIS-R-based conversions, all of the linear-equated estimates showed highly positive and significant correlations with the Full-Scale IQ scores, thus receiving support for the first hypothesis. Support for the second hypothesis was had for each of the linear-equated procedures, as can be seen from the lack of a significant main effect for type of short form in the repeated-measure ANCOVA given in Table 5. As with the WAIS-R conversion estimates, education level emerged as a significant covariate with the obtained IQ score. The differences in mean IQ scores from the full-scale IQ score were much smaller than those for the WAIS-R-based estimates, and all were within the 3-IQ-point standard error of measurement for the scale. With regards to the third hypothesis concerning accuracy of classification, the linear equating procedure yielded more correct classifications than did the WAIS-R-based short forms, with three of the short forms obtaining correct classification rates of over 70% -- the 4-subtest short-forms of Silverstein (1982), Ward et al. (1987), and Reynolds et al. (1983). The 4-subtest short-form by Silverstein (1982) provided the highest rate of correct classification, and also evidenced the highest correlation with the full-form Full-Scale IQ. The Ward et al. (1987) short-form provided the next highest correct classification rate, but was third-highest with regards to correlation with the full-form Full-Scale IQ. From the three evaluation criteria, and a consideration of the relative standing of the four Wechsler-based short-forms within each other, the two most useful short forms are the 4-subtest short-forms by Silverstein (1982) and by Ward et al. (1987). A deciding factor comes from a re-examination of the validities for the short forms, given

in Table 1. The Silverstein (1982) 4-subtest short-form provided the best validity, while the Ward et al. (1987) 4-subtest short-form provided the worst. Further support is found in an examination of the absolute differences in prediction for each of the short forms, given in Table 7. The Silverstein (1982) 4-subtest short-form had 77% of its estimated scores within 3 IQ points of the actual full-form Full-Scale IQ, whereas the Ward et al. (1987) short-form only had 56% of the estimates within this range. These findings place the Silverstein (1982) 4-subtest short-form as the most effective short form. A clinician using this short form, along with a linear-equated estimation procedure, will produce a Full-Scale IQ estimate that is within 3-IQ points of the actual Full-Scale IQ over 75% of the time. The only drawback for this short form is the amount of time required, which in this sample was nearly half of the administration time of the entire battery.

The overall findings regarding the Wechsler-based short forms do seem appealing, and would generally lead to the conclusion that practitioners who use these subtests from the WAIS-III could possibly use the conversion tables from the WAIS-R with the four-subtest short forms. Further support for this statement can be seen in some of the reviews for the three evaluation criteria (e.g., Silverstein, 1990), where the usefulness of any criterion beyond the finding of a high and significant correlation is questionable. Both types of estimation procedures (WAIS-R conversion and linear-equating) met this criterion for all of the short forms. What argues against continued use of the WAIS-R-based conversions is the high degree of misclassification evidenced by their use. Even though the mean estimates were accurate, the individual classifications provide a richer picture of the performance of the short forms on an individual level. The low

performance of the short forms in this regard argues against the use of the WAIS-R conversions, and argues in favour of using linear-equated estimates based on the WAIS-III. Further statistical support for the use of linear-equated estimates comes from the findings regarding the fourth and fifth hypothesis, which were concerned with the correlations between the short forms and the full-form with respect to Full-Scale IQ, and the degree of overestimation evidenced by each of the short form estimates. The fourth hypothesis found moderate support for two of the linear-equated estimation procedures, the Cyr and Brooker (1984) 2-subtest short-form estimation and the Reynolds et al. (1983) 4-subtest short-form, which had correlation values with the full-form Full-Scale IQ which were significantly different from their WAIS-R counterparts. The fifth hypothesis found support for all of the linear-equated conversions yielding mean estimates which were significantly lower than the WAIS-R-based conversion estimates, as evidenced by the uniform results of the t-tests reported in Table 9. The only short form excepted from these conclusions is the Ward et al. (1987) 4-subtest short-form, since only linear-equated estimates were used with this short form. Together, these findings support the production of new conversion tables for the WAIS-III for any of the short forms which are based on the WAIS-R, and for these new conversion formulae to be based upon the linear equating procedure outlined by Tellegen and Briggs (1967).

The Shipley Institute of Living Scale should probably not be used as an estimate of WAIS-III IQ in its present form with an elderly sample. With regards to the three evaluation criteria, the SILS estimates of Full-Scale IQ passed two of the evaluation criteria (finding support for Hypotheses 1 and 2), but failed to meet any of them to the

same degree as the Wechsler-based short forms. The SILS estimates supported the first hypothesis of a significant positive correlation, but the correlation values were not as high as those for the Wechsler-based short forms. There was also support for the second hypothesis, as the repeated-measures ANCOVA in Table 6 showed that there was no effect of type of conversion method. As with the Wechsler-based short-forms, the SILS estimation procedures failed to support the third hypothesis regarding qualitative classification. These two methods also produced many IQ estimates which were outside of the 3-IQ-point standard error of estimate (78% of the entire sample for both estimation procedures). The inferiority of the SILS in estimating Wechsler IQ scores has also been found by Watson et al. (1992). The distribution of values in Table 8 are not indicative of the possibility of a correction being made through the addition or subtraction of a constant to the WAIS-III Full-Scale IQ estimate, as had been indicated by Retzlaff et al. (1986) with regards to the WAIS-R Full-Scale IQ. Consistent with previous research on the SILS (e.g., Harnish et al., 1994; Morgan & Hatsukami, 1986), education emerged as a significant covariate affecting performance on this scale, and on the IQ estimates. Clinicians and practitioners who use this scale would be well advised to seek updated SILS conversion procedures for use with the WAIS-III.

Consistent with the findings of Morgan and Hatsukami (1986), many of the elderly individuals in this sample scored rather low on the Abstraction scale, averaging only 5 correct items. Many of the participants found this section of the SILS to be very difficult, as the instructions were vague and unclear. Willis (1996) has stated that elderly individuals tend to have a rather concrete method of thinking, and that in the face of

uncertainty will prevail upon a routine and see if it yields an answer, regardless of whether the answer is correct or not. This was the most likely cognitive process being engaged by the current sample, as many of the answers provided on the Abstraction test did follow a heuristic. It was simply the incorrect heuristic to employ.

The plots of the FSIQ, VIQ, and PIQ scores for the full-form Wechsler battery show a rough parallel to what has been seen in previous literature regarding cognition and the elderly (e.g., Hopp et al., 1997; Labouvie-Vief, 1985). Specifically, the FSIQ and VIQ scores show a slight increase as age progresses, and the PIQ scores show a plateau and then a sharp decline. The dip in scores at the 70-74 age group is somewhat discrepant with what is generally stated regarding cognition trends, and may be the result of a number of factors. The failure to find any statistical trend with the Full-Scale IQ indicates that any observed differences between the age groups is most likely due to normal variation within the groups, and thus the dip in scores may not be as serious as it initially appears.

To address the anomalous dip in scores in the 70-74 age group, Botwinick (1977) has stated that many noticeable deficits in cognition do not manifest until this age, and these individuals may be showing the decline prior to the development of possible adjustment mechanisms that can compensate for this decline. Hopp et al. (1997) have also shown that cognitive ability in the 70's and 80's remains stable, and have implied that individual's who maintain their level of functioning may form an "elite" group of elderly. A related explanation, and one that is quite morbid, is one of terminal drop, where individuals show a marked decline in performance prior to death. Given the findings by

Hopp et al. (1997), and some of the summary writings by Botwinick (1977), such an explanation is possible. This could only be borne out by a longitudinal study of cognitive ability in the elderly, which is beyond the scope of this project.

A cautionary note is worthy when interpreting these findings within the general theoretical framework of cognition in the elderly. Although the Wechsler index scores are analogous to the crystallized and fluid intelligence postulated by Horn and Cattell (1967, cited in Botwinick, 1977, p. 589), recently the interpretation of these index scores as facets of intelligence has been questioned (Caruso & Cliff, 1999). These authors, using the WAIS-III, found that a two-factor model did not divide the Verbal and Performance subtests according to their theoretical orientation, and they even found some substantial cross-loading of subscales between the two facets. They suggested that, in clinical practice, factor scores be used to replace the traditional VIQ and PIQ scores as they represent "a more realistic, and therefore more clinically useful, bifurcation of ability" (Caruso and Cliff, 1999, p. 204). What this means in light of the current study is unclear, although it does raise some psychometric issues involving the short forms. Traditional factor analysis requires a minimum of three indicators per factor (Comrey, 1973). Many of the short forms that are currently in the literature do not meet this criterion, and so would be unable to generate the two factors scores required by Caruso and Cliff (1999). The solution, it seems would be to have at least six subtests in a short form, which moves the short form into including over half of the original scale in the short form. This may lead to a compromise in testing time, something that the short-form is trying to avoid!

The four Wechsler-based short-forms and their respective linear-equated FSIQ estimates showed similar trends to those of the full-form IQ plot across the age groups, thus lending some initial confidence for the sixth hypothesis. Many of the mean estimates for each age group were within one standard deviation of the full-form IQ score for that age group, so it is evident that the linear equating procedure for each of the short forms was providing accurate scores across the age groups. This can also be seen in the estimates for each age group in Table 9. Further, the short forms closely mapped the pattern of cognition across the elderly life span in this sample, thus being able to capture the classical ageing pattern that is seen in cognition (Botwinick, 1977). As noted above, a test for significant trends among the estimates was not supported. The evaluation of these findings in light of the sixth hypothesis is unclear. Strictly speaking, there was no evidence of the linearly decreasing trend that is commonly seen in cross-sectional studies of elderly cognition, so in this sense the sixth hypothesis failed to be supported. Yet this hypothesis was made under the assumption that the actual full-form Full-Scale IQ scores would follow the trend that was predicted by theory. The results of the attempted trend analysis on the full-form Full-Scale IQ scores also failed to show a decreasing linear trend. As a result, since there was a lack of support for this hypothesis across all of the IQ scores, both actual and estimated, the actual validity of the sixth hypothesis can be put into question. It is possible that the amount of variance in the estimated IQ scores within each age group may have been responsible for the lack of a trend. A possible solution would be to restrict the range of scores for each age group, say at 1.5 standard deviations around the mean for that age group. This would encompass just under 90% of

the estimated IQ scores in the total sample, thus ensuring representativeness of the sample. Another method that may reduce the amount of variability is to trim a certain number of outliers from each age group. Although Keppel (1991) has stated that a restricted range can actually mask the type of trend that is evidenced by the data, given the high amount of variability typically seen with the elderly (e.g., Botwinick, 1977; Lindenberger & Reischies, 1999), such an explanation does not sound entirely plausible (although the small N in each age group does make such a conjecture possible).

It may be that the relationship between age and cognition is not strictly a linear one, even in a cross-sectional analysis, when one is looking at only the elderly. It may also be that the relationship is actually non-linear, although the lack of significance for quadratic trends in any of the attempted trend analyses argues against this consideration. A third consideration may be the presence of a higher-order trend (Glass & Hopkins, 1996; Keppel, 1991), one that is a blend of both linear and quadratic trends. The existence of a cubic trend may also be possible. Given the well-replicated research findings regarding the leveling and subsequent decline of intelligence with age (e.g., Botwinick, 1977), the possibility of a higher-order trend being present is much more plausible. Such possibilities should be explored in the future, with larger sample sizes which could be more balanced across the age groups than in the present study. This consideration can also have an impact on the estimation method used to generate the IQ scores from the short forms. There is an obvious non-linear trend to elderly cognition, and the use of a linear equating method could be partly responsible for the poor performance of the short forms, especially with respect to the qualitative IQ

classification. A way to correct this would be to incorporate a non-linear component into the construction of conversion tables for both current and future short-forms of the Wechsler battery.

Many clinicians use the difference between Verbal and Performance IQ scores to assess potential problems in cognitive functioning (O'Donnell & Leicht, 1990). The frequency counts presented in Tables 12 and 13 indicate that the short forms which utilize four subtests are comparable to the full Wechsler battery when detecting VIQ-PIQ differences. All of the short forms provided equivalent numbers of significant VIQ-PIQ differences as the full-form scale. The most proficient short form in regards to detecting VIQ-PIQ differences appears to be the Ward et al. (1987) short form, which possessed a distribution of scores that matched the actual distribution of VIQ-PIQ differences closely. Many of the differences that exceeded the minimum difference of 8.76 points were on the positive side of differences, indicating that the Verbal IQ was generally higher than the Performance IQ. Such a finding is consistent with previous literature regarding cognition in the elderly (e.g., Botwinick, 1977, Mitrushima & Satz, 1995).

#### The Flynn Effect

The action of the Flynn effect was observed with the use of the SILS conversion tables, evidenced by a consistent overestimation of FSIQ scores. An overestimation was also seen with the WAIS-R conversion tables associated with two of the 4-subtest short-forms, the Silverstein (1982) and the Reynolds et al. (1987) short-forms. This current study finds some support for the presence of the Flynn effect, but cannot make any clear statements about the mechanism that is at work. There is still some debate as to the

actual existence of the Flynn effect and with its exact mechanism (e.g., Neisser, 1998; Rodgers, 1998). Neisser (1997) put forward several possible mechanisms, such as an increasing degree of test-taking sophistication on the part of the examinees, increased nutrition, and increased years of schooling. Many of these potential mechanisms may not be applicable to the elderly, as they have most likely been out of school for several decades and have not been required to take a comprehensive test within that period. In the current sample, nutritional regime was not assessed. Neisser (1997) does intimate a possible cohort effect that is associated with the types of standardization sample used, an explanation that is consistent with the present data. The overestimation was seen primarily with the WAIS-R and SILS conversions, which utilized the WAIS-R standardization sample used to calculate the conversion procedures originally. The only exception to this would be the Paulson and Lin (1970b) SILS conversion which, being based upon the WAIS, should have shown more overestimation than the Zachary et al. (1985) procedure. This was not the case, as the Paulson and Lin (1970b) estimation produced a similar amount of overestimation as the Zachary et al. (1985) procedure. There is no ready-made explanation for why the two conversion procedures would not be affected by the Flynn effect, but it may be related to the SILS itself, which is not susceptible to age effects (Corotto, 1966), and one potential mechanism for the Flynn effect is a cohort effect (Neisser, 1997). This author agrees with Rodgers (1998) that the causes of the Flynn effect should continue to be investigated and clarified.

#### Education as a Covariate

It was also shown that education level had a significant effect on performance for

all of the short forms assessed in this study, a finding that has been seen in previous research (e.g., Satz, Hynd, D'Ella, & Daniel, 1990), and many practitioners would be advised to include this element in any cognitive assessment, both with the full-form and any short forms of the Wechsler batteries. Powell and Whitla (1994) have stated that education can play an important role in the assessment of an elderly individual. For instance, an individual with a low level of education but who is not cognitively impaired may register as impaired on the Wechsler battery. Alternatively, a person who is cognitively impaired but who is well-educated will most likely score in the normal IQ range. Some support for this has been found by Reitan and Wolfson (1996), where participants who were brain-damaged performed similarly to normal participants when matched on education (although see Shuttleworth, 1997, for an alternative interpretation). Further to this, if education is having a significant effect on performance, and crystallized intelligence is the reflection of this, then the Wechsler battery could be characterized as mainly a "crystallized" cognitive battery. Malec, Ivnik, Smith, and Tangalos (1992) have also presented regression equations that can be used with the WAIS-R to correct for education effects, as greater education is related to stronger performance on measures of memory (e.g., Vocabulary, Information). As these two subtests are part of three of the four Wechsler-based short-forms studied here, and a significant effect of education was found across all of the short forms studied, there is support for a similar correction to be applied to the WAIS-III. Clinicians should thus be made aware that a high score on these types of subtests could simply be caused by education effects, and may not be representative of the true cognitive level of the individual. Other factors that should be

investigated in future studies are the current biological state of the individual (e.g., are they suffering from a disease or sickness) and various sociological factors (e.g., socioeconomic status) for their influence on short-form performance.

#### Summation and Recommendations

In summary, for those clinicians who are seeking to reduce the administration time of the WAIS-III and still attempting to attain an adequate estimate of the current cognitive functioning of an elderly individual, the use of either the 4-subtest short-form by Ward et al. (1987) or the 4-subtest short-form by Silverstein (1982) are the most promising. The use of the linear equating method for the Silverstein (1982) short-form resulted in more accurate estimates and had a higher proportion of estimated scores within a 95% confidence interval of the full-form IQ score, making this short form an ideal choice for the assessment of general intelligence. The short form by Ward et al. (1987) provided similar accuracy in qualitative classification and required less administration time, but showed a much poorer validity than the Silverstein (1982) 4-subtest short form. The Ward et al. (1987) short form did provide more accuracy with regards to VIQ-PIQ differences. Both of these two short forms also showed superior performance in regards to mapping the cognitive ability of the sample across the elderly age groupings. If a clinician is mainly interested in using a short form to obtain a rough picture of VIQ-PIQ discrepancy, the short form by Ward et al. (1987) would be the most advantageous candidate. If a clinician is desiring an accurate estimate of global intellectual functioning, the Silverstein (1982) 4-subtest short form is the better candidate.

The present study has further evidenced that the short forms which were based on

the WAIS-R are still highly reliable, although the conversion procedures used with those subtests need to be re-evaluated. Clinicians who use the short forms with the older conversion tables will most likely make the mistake of coming to the wrong conclusion regarding the client's IQ level. This inaccuracy is not just with the actual level of cognitive functioning, but also extends to the more general assessment of cognitive classification. The use of the linear equating procedures ameliorated this effect somewhat, but there was still quite a high amount of misclassification evidenced. A possible explanation for this can be found in the findings that cognitive scores of the elderly become more variable than those of younger adults (Lindenberger & Reischies, 1999). Such an explanation is also consistent with the findings of the attempted trend analysis, in which the amount of variability within each age group resulted in no significant differences across the age groups. This could serve to make conversion tables for this age group a minimal advantage at best, as the tables could not hope to capture the amount of variability that this type of sample possesses. The inclusion of confidence bands could improve the diagnostic reliability of the short-form estimates somewhat.

The use of the Shipley Institute of Living Scale in the elderly as an estimator of WAIS-III IQ with the elderly is questionable. While both conversion formulas yielded accurate estimations of the mean full-form IQ, they still resulted in high misclassification rates along with a tendency to overestimate the full-form IQ. Accuracy rates for each of the Shipley conversion procedures were low, and a visual inspection of the plots of Full-Scale IQ by age group (Figures 8 and 9) show substantial deviations from the analogous plot for the full-form Full-Scale IQ (Figure 1). Continued use of this test would

necessitate the formulation of new conversion formulas, ones that take into account the education level of the individual, would yield more accurate qualitative classification rates, and an ability to follow the trend of scores evidenced by the full form. Such a task may be fruitful, as the SILS took just over 13 minutes to administer, a time that was the shortest of all the estimation methods studied.

Further research should investigate the performance of the individual subtests in both elderly and clinical populations. Ryan and Lopez (1999) have shown that the ordering of items on the Picture Arrangement subtest is not an increasingly difficult one in a sample of individuals diagnosed with alcohol dependence. This may lead to the individuals who fail early items to assume they have reached the limits of their ability, an occurrence that is also plausible with the elderly. One aspect of this study that was not controlled for was the current psychological state of the participants, since Wolff and Gregory (1991) have shown that a temporary dysphoric mood can affect performance on certain subtests of the WAIS-R. Dugbartey, Sanchez, Rosenbaum, Mahurin, Davis, & Townes (1999), using a mixed clinical and an immigrant sample, have also shown that the Matrix Reasoning subtest in the WAIS-III has a strong verbal component and implied that this test may not be a strictly nonverbal task. The previous statements are not meant to infer that the elderly sample in this current study could be considered as being a clinical sample, and even referring to it as a clinical-analogue sample would be inaccurate. The main point to be made here is that the performance of the individual subtests can be affected by a great many factors, and some of these factors may be present in elderly samples, be they clinical or not. Given that the participants in this study were

in personal care homes or attending day hospitals, it is possible that some were taking medications or could have been suffering from some sort of psychological distress, and this may have had an effect on their performance. Hendricks and Hendricks (1981) have stated that most individuals in personal care homes are typically suffering from health problems that have led to their admittance into the personal care home. Tobin and Lieberman (1976) have also shown that individual who are living in such institutions differ from elderly in the normal population, possessing slight cognitive disorganization, diminished self-esteem, and depression. These factors would almost certainly be present in clinical samples, and should be investigated in future research endeavours.

It should be noted that the development of a shorter test is not the only way to obtain an estimate of intelligence. Barona, Reynolds, and Chastain (1984, cited in Paolo, Ryan, Troster, and Hilmer, 1996) have developed a set of regression equations for the prediction of premorbid WAIS-R intelligence that are based on demographic information. Paolo et al., (1996) and Paolo, Ryan, and Troster (1997) have extended a set of such equations to the elderly WAIS-R sample, finding them able to detect intellectual deterioration in a brain damaged sample. These equations have not yet been extended to the WAIS-III.

Some limitations to this study are as follows. The relatively small sample size precludes the use of more complex and in-depth analytical procedures (e.g., structural equation modeling). There was no direct assessment of the presence (or absence) of psychopathology, as these could have had an impact on the performance of the elderly individuals in this study. Finally, the use of a comparison test would have been

advantageous to establish a measure of construct validity for the various short-forms studied.

## References

- Bartz, W. R., & Loy, D. L. (1970). The Shipley-Hartford as a brief IQ screening device. Journal of Clinical Psychology, 26, 74-75.
- Benedict, R. H. B., Schretlen, D., & Bobholz, J. H. (1992). Concurrent validity of three WAIS-R short forms in psychiatric inpatients. Psychological Assessment, 4, 322 - 328.
- Boone, D. E. (1990). Short forms of the WAIS-R with psychiatric inpatients: A comparison of techniques. Journal of Clinical Psychology, 46, 197-200.
- Boone, D. E. (1991). Use of the modified WAIS-R with psychiatric inpatients: A caution. Perceptual and Motor Skills, 73, 315 - 322.
- Boone, D. E. (1992). Evaluation of Kaufman's short forms of the WAIS-R with psychiatric inpatients. Journal of Clinical Psychology, 48, 239 - 245.
- Boone, D. E. (1995). A cross-sectional analysis of WAIS-R aging patterns with psychiatric inpatients: Support for Horn's hypothesis that fluid cognitive abilities decline. Perceptual and Motor Skills, 81, 371 - 379.
- Botwinick, J. (1977). Intellectual abilities. In Birren, J. E., & Schaie, K. W. (Eds.). Handbook of the psychology of aging. New York: Van Nostrand.
- Brooker, B. H., & Cyr, J. J. (1986). Tables for clinicians to use to convert WAIS-R short forms. Journal of Clinical Psychology, 42, 982-985.
- Cargnello, J. C., & Gurekas, R. (1987). The clinical use of a modified WAIS procedure in a geriatric population. Journal of Clinical Psychology, 43, 286 - 296.
- Cargnello, J. C., & Gurekas, R. (1988). The WAIS-SAM: A comprehensive

administrative model of modified WAIS procedures. Journal of Clinical Psychology, 44, 266 - 270.

Caruso, J. C., & Cliff, N. (1999). The properties of equally and differentially weighted WAIS-III factor scores. Psychological Assessment, 11, 198 - 206.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378 - 399.

Cella, D. G. (1984). The modified WAIS-R: An extension and revision. Journal of Clinical Psychology, 40, 801 - 804.

Compton, D. M., Bachman, L. D., & Logan, J. A. (1997). Aging and intellectual ability in young, middle-aged, and older educated adults: Preliminary results from a sample of college faculty. Psychological Reports, 81, 79 - 90.

Comrey, A. L. (1973). A first course in factor analysis. New York: Academic Press.

Conrad, H. S. (1931). The measurement of adult intelligence, and the requisites of a general intelligence test. Journal of Social Psychology, 1, 72 - 86.

Coon, D. (1995). Introduction to psychology: Exploration and Application (7<sup>th</sup> Edition). St. Paul; West Publishing.

Corotto, L. V. (1966). Effects of age and sex on the Shipley-Institute of Living Scale. Journal of Consulting Psychology, 30, 179.

Cotzin, M., & Gallagher, J. D. (1950). The Southbury Scale: A valid abbreviated Wechsler-Bellevue for mental defectives. Journal of Consulting Psychology, 14, 358 - 364.

Crawford, J. R., Allan, K. M., & Jack, A. M. (1992). Short forms of the UK WAIS-R: Regression equations and their predictive validity in a general population sample. British Journal of Clinical Psychology, *31*, 191 - 202.

Cyr, J. J., & Brooker, H. (1984). Use of appropriate formulas for selecting WAIS-R short forms. Journal of Consulting and Clinical Psychology, *52*, 903 - 905.

Dalton, J. E., & Dubnicki, C. (1981). Sex, race, age, and education variables in Shipley-Hartford scores of alcoholic inpatients. Journal of Clinical Psychology, *37*, 885-888.

Dalton, J. E., Pederson, S. L., & McEntyre, W. L. (1987). A comparison of the Shipley vs. WAIS-R tests in predicting WAIS-R full scale IQ. Journal of Clinical Psychology, *43*, 278-280.

Darlington, R. B., & Carlson, P. M. (1987). Behavioral statistics: Logic and methods. New York: Macmillan Inc.

Dennis, D. M. (1973). Predicting full scale WAIS IQs with the Shipley-Hartford. Journal of Clinical Psychology, *29*, 266-268.

Doppelt, J. E. (1956). Estimating the full-scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. Journal of Consulting Psychology, *20*, 63 - 67.

Dugbartey, A. T., Sanchez, P. N., Rosenbaum, J. G., Mahurin, R. K., Davis, J. M., & Townes, B. D. (1999). WAIS-III Matrix Reasoning test performance in a mixed clinical sample. The Clinical Neuropsychologist, *13*, 396 - 404.

Edwards, A. J. (1993). Dementia. New York: Plenum Press.

Engelhart, C. I., Eisenstein, N., Johnson, V., & Losonczy, M. (1999).

Comparison of linear equating and prorated short forms for estimating WAIS-R FSIQ in a neuropsychological population. Clinical Neuropsychologist, *13*, 95 - 99.

Enns, R. A., & Reddon, J. R. (1998). The factor structure of the Wechsler Adult Intelligence Scale - Revised: One or two but not three factors. Journal of Clinical Psychology, *54*, 447 - 459.

Fowles, G. P., & Tunick, R. H. (1986). WAIS-R and Shipley estimated IQ correlations. Journal of Clinical Psychology, *42*, 647 - 649.

Frick, R. W. (1995). Accepting the null hypothesis. Memory and Cognition, *23*, 132 - 138.

Garret, H. E., & Woodworth, R. S. (1966). Statistics in psychology and education (Sixth edition). New York: David McKay Company.

Glass, G. V., & Hopkins, K. S. (1996). Statistical Methods in Education and Psychology (3<sup>rd</sup> Edition). Needham Heights: Allyn & Bacon.

Goebel, R. A., & Satz, P. (1975). Profile analysis and the abbreviated Wechsler Adult Intelligence Scale: A multivariate approach. Journal of Consulting and Clinical Psychology, *43*, 780 - 785.

Gregory, R. J. (1992). Psychological testing: History, principles, and applications. Needham Heights: Allyn & Bacon.

Groth-Marnat, G. (1990). Handbook of Psychological Assessment (2<sup>nd</sup> Edition). New York: Wiley & Sons.

Harnish, M. J., Beatty, W. W., Nixon, S. J., & Parsons, O. A. (1994).

Performance by normal subjects on the Shipley Institute of Living Scale. Journal of Clinical Psychology, 50, 881 - 883.

Haynes, J. P. (1985). Comparative validity of two WAIS-R short forms with clients of low IQ. Journal of Clinical Psychology, 41, 282 - 284.

Hayslip, B. (1983). Mental health and aging. In N. S. Ernst, & H. R. Glazer-Waldman (Eds.), The aged patient: A sourcebook for the allied health professional (pp. 158 - 181). London, England: Year Book Medical Publishing.

Heinemann, A. W., Harper, R. G., Friedman, L. C., & Whitney, J. (1985). The relative utility of the Shipley-Hartford scale: Prediction of WAIS-R IQ. Journal of Clinical Psychology, 41, 547-551.

Hendricks, J., & Hendricks, C. D. (1981). Aging in mass society: Myths and realities (2<sup>nd</sup> ed.). Cambridge, MA: Winthrop Publishers.

Himelstein, P. (1957). A comparison of two methods of estimating full scale IQ from an abbreviated WAIS. Journal of Consulting Psychology, 21, 246.

Hoffman, R. G., & Nelson, K. S. (1988). Cross-validation of six short forms of the WAIS-R in a healthy geriatric sample. Journal of Clinical Psychology, 44, 952-957.

Holmes, D. S., Armstrong Jr., H. E., Johnson, M. H., & Reis, H. A. (1965). Further evaluation of an abbreviated form of the WAIS. Psychological Reports, 16, 1163 - 1164.

Holmes, D. S., Armstrong Jr., H. E., Johnson, M. H., & Reis, H. A. (1966). Validity and clinical utility of the Satz and Mogel abbreviated form of the WAIS. Psychological Reports, 18, 992 - 994.

Hopp, G. A., Dixon, R. A., Grut, M., & Backman, L. (1997). Longitudinal and psychometric profiles of two cognitive status tests in very old adults. Journal of Clinical Psychology, 53, 673 - 686.

Jones, R. L. (1962). Analytically developed short forms of the WAIS. Journal of Consulting Psychology, 26, 289.

Iverson, G. L., Myers, B., Bengston, M. L., & Adams, R. L. (1996). Concurrent validity of a WAIS-R seven-subtest short form in patients with brain impairment. Psychological Assessment, 8, 319 - 323.

Kaufman, A. S., Ishikuma, T., & Kaufman-Packer, J. L. (1991). Amazingly short forms of the WAIS-R. Journal of Psychoeducational Assessment, 9, 4 - 15.

Kaplan, R. M., & Saccuzzo, D. P. (1993). Psychological testing: Principles, applications, and issues (3<sup>rd</sup> edition). Pacific Grove: Brooks/Cole.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3<sup>rd</sup> edition). New Jersey: Prentice Hall.

Kish, G. B., & Ball, M. E. (1969). Low education level as one factor producing a verbal-abstract disparity on the Shipley-Institute of Living Scale. Journal of Clinical Psychology, 25, 183-184.

Kramer, E., & Francis, P. S. (1965). Errors in intelligence estimation with short forms of the WAIS. Journal of Consulting Psychology, 29, 490.

Labouvie-Vief, G. (1990). Intelligence and cognition. In J. E. Birren, & K. W. Schaie, (Eds.) (1990) Handbook of the psychology of aging (3<sup>rd</sup> edition). New York: Van Nostrand.

Levy, P. (1968). Short-form tests: A methodological review. Psychological Bulletin, *69*, 410 - 416.

Lindenberger, U., & Reischies, F. M. (1999). Limits and potentials of intellectual functioning in old age. In P. B. Baltes, & K. U. Mayer, (Eds.) The Berlin Aging Study: Aging from 70 to 100. Cambridge: Cambridge University Press.

Malec, J. F., Ivnik, R. J., Smith, G. E., & Tangalos, E. G. (1992). Mayo's Older American Normative Studies: Utility of corrections for age and education for the WAIS-R. Clinical Neuropsychologist, *6*, 31 - 47.

Margolis, R. B., Taylor, J. M., & Greenlief, C. L. (1986). A cross-validation of two short forms of the WAIS-R in a geriatric sample suspected of dementia. Journal of Clinical Psychology, *42*, 145 - 146.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5<sup>th</sup> edition). Baltimore: Williams and Wilkins.

Maxwell, E. (1957). Validities of abbreviated WAIS scales. Journal of Consulting Psychology, *21*, 121 - 126.

McCornack, R. L. (1956). A criticism of studies comparing item-weighting methods. Journal of Applied Psychology, *40*, 343 - 344.

McNemar, Q. (1950). On abbreviated Wechsler-Bellevue scales. Journal of Consulting Psychology, *14*, 79 - 81.

Mitrushina, M. & Satz, P. (1995). Base rates of the WAIS-R intersubtest scatter and VIQ-PIQ discrepancy in normal elderly. Journal of Clinical Psychology, *51*, 70-78.

Mogel, S., & Satz, P. (1963). Abbreviation of the WAIS for clinical use: An

attempt at validation. Journal of Clinical Psychology, 19, 298 - 300.

Morgan, S. F., & Hatsukami, D. K. (1986). Use of the Shipley Institute of Living Scale for neuropsychological screening of the elderly: Is it an appropriate measure for this population? Journal of Clinical Psychology, 42, 796-798.

Mumpower, D. L. (1964). The fallacy of the short form. Journal of Clinical Psychology, 20, 111 - 113.

Neisser, U. (1997). Rising scores on intelligence tests. American Scientist, 85, 440 - 447.

Neisser, U. (1998). The rising curve: Long-term gains in IQ and related measures. Washington: American Psychological Association.

O'Donnell, J. P., & Leicht, D. J. (1990). WAIS VIQ-PIQ differences: Validation and theoretical implications for younger adults with learning disabilities. International Journal of Clinical Neuropsychology, 12, 140-146.

Oakes, W. F. (1975). On the alleged falsity of the null hypothesis. The Psychological Record, 25, 265 - 272.

Pangman, V. C. (1996). Comparison of the reliability of two mental status questionnaires (MMSE & SMMSE) for use in screening cognitive impairment in elderly clients in chronic care units: Implications for nursing practice. MA thesis, University of Manitoba.

Paolo, A. M., & Ryan, J. J. (1991). Application of WAIS-R short forms to persons 75 years of age and older. Journal of Psychoeducational Assessment, 9, 345 - 352.

Paolo, A. M., & Ryan, J. J. (1993). Is the WAIS-R an acceptable test for the elderly? Opinions of examinees 75 years and older. Journal of Clinical Psychology, 49, 720-723.

Paolo, A. M., Ryan, J. J., & Troster, A. I. (1997). Estimating premorbid WAIS-R intelligence in the elderly: An extension and cross validation of new regression equations. Journal of Clinical Psychology, 53, 647 - 656.

Paolo, A. M., Ryan, J. J., Troster, A. I., & Hilmer, C. D. (1996). Utility of the Barona demographic equations to estimate premorbid intelligence: Information from the WAIS-R standardization sample. Journal of Clinical Psychology, 52, 335 - 343.

Paulson, M. J., & Lin, T. (1970a). Age: The neglected variable in constructing an abbreviated WAIS. Journal of Clinical Psychology, 26, 336 - 343.

Paulson, M. J., & Lin, T. (1970b). Predicting WAIS IQ from Shipley-Hartford scores. Journal of Clinical Psychology, 26, 453 - 461.

Powell, D. H., & Whitla, D. K. (1994). Profiles in cognitive aging. Cambridge: Harvard University Press.

Prado, W. M., & Taub, D. V. (1966). Accurate prediction of individual intellectual functioning by the Shipley-Hartford. Journal of Clinical Psychology, 22, 294 - 296.

Psychological Corporation (1997). WAIS-III Administration and Scoring Manual. San Antonio: The Psychological Corporation.

Psychological Corporation (1997). WAIS-III - WMS-III Technical Manual. San Antonio: The Psychological Corporation.

Rabin, A. I. (1943). A short form of the Wechsler-Bellevue test. Journal of Applied Psychology, *27*, 320 - 324.

Reitan, R. M., & Wolfson, D. (1996). Relationships of age and education to Wechsler Adult Intelligence Scale IQ values in brain-damaged and non-brain-damaged groups. Clinical Neuropsychologist, *10*, 293 - 304.

Resnick, R. J., & Entin, A. D. (1971). Is an abbreviated form of the WISC valid for afro-american children? Journal of Consulting and Clinical Psychology, *36*, 97 - 99.

Retzlaff, P., Slicner, N., & Gibertini, M. (1986). Predicting WAIS-R scores from the Shipley Institute of Living Scale in a homogeneous sample. Journal of Clinical Psychology, *42*, 357-359.

Reynolds, J. R., Wilson, V. L., & Clark, P. L. (1983). A four-test short form of the WAIS-R for clinical screening. Clinical Neuropsychology, *5*, 111 - 116.

Robiner, W. N., Dossa, D., & O'Dowd, W. (1988). Abbreviated WAIS-R procedures: Use and limitations with head-injured patients. The Clinical Neuropsychologist, *2*, 365 - 374.

Rodgers, J. L. (1998). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? Intelligence, *26*, 337 - 356.

Roth, D. L., Hughes, C. W., Monkowski, P. G., & Crosson, B. (1984). Investigation of the validity of WAIS-R short forms for patients suspected to have brain impairment. Journal of Consulting and Clinical Psychology, *52*, 722 - 723.

Ryan, J. J. (1999). Two types of tables for use with the seven-subtest short forms of the WAIS-III. Journal of Psychoeducational Assessment, *17*, 145 - 151.

Ryan, J. J., Georgemiller, R. J., & McKinney, B. E. (1984). Application of the four-subtest WAIS-R short form with an older clinical sample. Journal of Clinical Psychology, 40, 1033 - 1036.

Ryan, J. J., Larsen, J., & Prifitera, A. (1983). Validity of two- and four-subtest short forms of the WAIS-R in a psychiatric sample. Journal of Consulting and Clinical Psychology, 51, 460.

Ryan, J. J., & Lopez, S. J. (1999). Order of item difficulty on picture arrangement: Extending the discussion to the WAIS-III. Perceptual and Motor Skills, 88, 1053 - 1056.

Ryan, J. J., Lopez, S. J., & Werth, T. R. (1998). Administration time estimates for WAIS-III subtests, scales, and short forms in a clinical sample. Journal of Psychoeducational Assessment, 16, 315 - 323.

Ryan, J. J., & Ward, L. C. (1999). Validity, reliability, and standard errors of measurement for two seven-subtest short forms of the Wechsler Adult Intelligence Scale-III. Psychological Assessment, 11, 207 - 211.

SAS Institute Inc. (1997). SAS [computer program]. Cary, North Carolina.

Satz, P., Hynd, G. W., D'Ella, L., & Daniel, M. H. (1990). A WAIS-R marker for accelerated aging and dementia, Alzheimer's type? Base rates of the Fuld formula in the WAIS-R standardization sample. Journal of Clinical and Experimental Neuropsychology, 12, 759 - 765.

Satz, P., & Mogel, S. (1962). An abbreviation of the WAIS for clinical use. Journal of Clinical Psychology, 18, 77 - 79.

Schaie, K. W., & Geiwitz, J. (1982). Adult development and aging. Toronto: Little, Brown, and Company.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. Journal of Psychology, *9*, 371-377.

Shuttleworth, J. A. B. (1997). Age and education effects on brain-damaged subjects: "Negative" findings revisited. Clinical Neuropsychologist, *11*, 205 - 209.

Silverstein, A. B. (1965). "Validity" of short form intelligence tests. Journal of Consulting Psychology, *29*, 392 - 393.

Silverstein, A. B. (1967a). Validity of WAIS short forms. Psychological Reports, *20*, 37 - 38.

Silverstein, A. B. (1967b). A short form of Wechsler's scales for screening purposes. Psychological Reports, *21*, 842.

Silverstein, A. B. (1968). Evaluation of a split-half short form of the WAIS. American Journal of Mental Deficiency, *72*, 839 - 840.

Silverstein, A. B. (1970). Reappraisal of the validity of a short short form of Wechsler's scales. Psychological Reports, *26*, 559 - 561.

Silverstein, A. B. (1982). Two- and four-subtest short forms of the Wechsler Adult Intelligence Scale - Revised. Journal of Consulting and Clinical Psychology, *50*, 415 - 418.

Silverstein, A. B. (1984). Estimating full scale IQs from short forms of Wechsler's scales: Linear scaling versus linear regression. Journal of Consulting and Clinical Psychology, *52*, 919.

Silverstein, A. B. (1985a). An appraisal of three criteria for evaluating the usefulness of WAIS-R short forms. Journal of Clinical Psychology, 41, 676-680.

Silverstein, A. B. (1985b). A formula for the standard error of estimate of deviation quotients on short forms of Wechsler's scales. Journal of Clinical Psychology, 41, 408 - 409.

Silverstein, A. B. (1985c). Two- and four-subtest short forms of the WAIS-R: A closer look at validity and reliability. Journal of Clinical Psychology, 41, 95-97.

Silverstein, A. B. (1990a). Notes on the reliability of Wechsler short forms. Journal of Clinical Psychology, 46, 194 - 196.

Silverstein, A. B. (1990b). Short forms of individual intelligence tests. Psychological Assessment, 2, 3-11.

Silverstein, A. B. (1990c). Critique of a Doppelt-type short form of the WAIS-R. Journal of Clinical Psychology, 46, 333 - 339.

Sines, L. K., & Simmons, H. (1959). The Shipley-Hartford scale and the Doppelt short form as estimators of WAIS IQ in a state hospital population. Journal of Clinical Psychology, 15, 452-454.

Spandel, H. Z. (1985). The psychoeducational use and interpretation of the Wechsler Adult Intelligence Scale - Revised. Springfield: Charles Thomas.

Swoyer, C., & Monson, T. C. (1975). Discussion: Theory confirmation in psychology. Philosophy of Science, 42, 487 - 502.

Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. Journal of Consulting Psychology, 5, 499 - 506.

Thompson, A. P. (1987). Methodological issues in the clinical evaluation of two- and four-subtest short forms of the WAIS-R. Journal of Clinical Psychology, 43, 142 - 144.

Tipton, R. M., & Stroud, L. H. (1973). Abbreviated forms of the WAIS. American Journal of Mental Deficiency, 78, 150 - 152.

Tobin, S., & Lieberman, M. (1976). The last home for the aged. San Francisco, CA: Jossey-Bass.

Troll, L. E. (1982). Continuations: Adult development and aging. Monterey: Brooks/Cole.

Vincent, K. R. (1979). The modified WAIS: An alternative to short forms. Journal of Clinical Psychology, 35, 624 - 625.

Ward, L. C. (1990). Prediction of verbal, performance, and full scale IQs from seven subtests of the WAIS-R. Journal of Clinical Psychology, 46, 436 - 440.

Ward, L. C., Selby, R. B., & Clark, B. L. (1987). Subtest administration times and short-forms of the Wechsler Adult Intelligence Scale-Revised. Journal of Clinical Psychology, 43, 276-278.

Watkins, C. E. (1986). Validity and usefulness of WAIS-R, WISC-R, and WPPSI short forms: A critical review. Professional Psychology: Research and Practice, 17, 36 - 43.

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

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale -

Revised. New York: Psychological Corporation.

Wechsler, D. (1997). Manual for the Wechsler Adult Intelligence Scale III. New York: Psychological Corporation.

Weiss, J. L., & Schell, R. E. (1991). Estimating WAIS-R IQ from the Shipley Institute of Living Scale: A replication. Journal of Clinical Psychology, *47*, 558 - 562.

Whitmyre, W., & Pishkin, V. (1958). The abbreviated Wechsler Adult Intelligence Scale in a psychiatric population. Journal of Clinical Psychology, *14*, 189-191.

Willis, S. L. (1996). Everyday problem solving. In J. E. Birren, & K. W. Schaie, (Eds.). Handbook of the psychology of aging (4<sup>th</sup> edition). New York: Van Nostrand.

Willis, S. L., & Baltes, P. B. (1982). Intelligence in adulthood and aging: Contemporary issues. In K. W. Schaie, & J. Geiwitz, (Eds. ) Readings in Adult Development and Aging. Toronto: Little, Brown and Company.

Wolff, K. C., & Gregory, R. J. (1991). The effects of a temporary dysphoric mood upon selected WAIS-R subtests. Journal of Psychoeducational Assessment, *9*, 340 - 344.

Zachary, R. A., Crumpton, E., & Spiegel, D. E. (1985). Estimating WAIS-R IQ from the Shipley Institute of Living Scale. Journal of Clinical Psychology, *41*, 532 - 540.

Zarantonella, M. M., Munley, P. H., & Milanovich, J. (1993). Predicting Wechsler Adult Intelligence Scale-Revised (WAIS-R) I scores from the Luria-Nebraska Neuropsychological Battery (Form 1). Journal of Clinical Psychology, *49*, 225-233.

Zimmerman, I. L., & Woo-Sam, J. M (1973). Clinical interpretation of the

Wechsler Adult Intelligence Scale. New York: Grune and Stratton.

## ***Appendix A***

Description of the subtests of the Wechsler-Bellevue (W-B), the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Adult Intelligence Scale - Revised (WAIS-R)

### Verbal Scales

**Information** — samples knowledge that average persons with average opportunities should be able to acquire.

**Digit Span** — the test taker must recall and repeat auditory information in the proper sequence.

**Vocabulary** — the test taker must define words that are presented both orally and visually. This subtest measures accumulated verbal learning, and represents an individual's ability to express a wide range of ideas with ease and flexibility.

**Arithmetic** — the test taker responds to orally presented math problems that are to be solved without the use of paper and pencil. This subtest measures numerical reasoning, speed of numerical manipulation, and concentration and attention.

**Comprehension** — the test taker must respond to a situation requiring an efficient way to deal with a specific problem. This subtest measures social judgement and the extent to which an examinee adheres to conventional standards

**Similarities** — the test taker is required to determine in which way two things are alike. This subtest requires the test taker to move from particular facts to a general rule or principle.

### Performance Scales

**Picture Completion** — the test taker must tell what is missing from each picture. This subtest measures visual concentration and involves paying attention to the environment in order to determine consistencies and inconsistencies.

**Picture Arrangement** — the test taker must rearrange pictures to form a proper sequence so as to tell a story. This subtest evaluates social judgement.

**Block Design** — the test taker must reproduce a design using red-and-white coloured blocks. This subtest measures nonverbal problem solving skills and one's ability to organize objects spatially.

**Object Assembly** — the test taker must assemble cutouts to form a flat picture of a familiar object. This subtest measures motor coordination and the ability to differentiate familiar configurations by fitting together pieces of a puzzle.

**Digit Symbol (Coding)** — the test taker must fill in as many symbols as possible under the numbers on the answer sheet. This subtest measures visual-motor speed, coordination, and the capacity for sustained concentration and effort.

***Appendix B***

Description of the new subtests of the Wechsler Adult Intelligence Scale — Version III.

**Matrix Reasoning** — a series of incomplete gridded patterns that the examinee completes by pointing to or saying of the correct response from five possible choices.

**Symbol Search** — a series of paired groups, each pair consisting of a target group and a search group. The examinee indicates, by marking the appropriate box, whether either target symbol appears in the search group.

**Letter-Number Sequencing** — a series of orally presented sequences of letters and numbers that the examinee simultaneously tracks and orally repeats, with the numbers in ascending order and the letters in alphabetical order.

Appendix C

Copy of the Shipley Institute of Living Scale

SHIPLEY INSTITUTE OF LIVING SCALE

Administration Form  
Walter C. Shipley, Ph.D.



Name: \_\_\_\_\_ Sex: M F Age \_\_\_\_\_  
Education: \_\_\_\_\_ Usual Occupation: \_\_\_\_\_ Today's Date: \_\_\_\_\_

Part I

Instructions: In the text below, the first word in each line is printed in capital letters. Opposite it are four other words. Circle the one word which means the same thing, or most nearly the same thing, as the first word. If you don't know, guess. Be sure to circle the one word in each line that means the same thing as the first word.

EXAMPLE:

LARGE	red	<u>big</u>	silent	wet
-------	-----	------------	--------	-----

- |                 |             |            |            |             |
|-----------------|-------------|------------|------------|-------------|
| (1) TALK        | draw        | eat        | sleep      | sleep       |
| (2) PERMIT      | allow       | sew        | cut        | drive       |
| (3) PARDON      | forgive     | pound      | divide     | tell        |
| (4) COUCH       | pin         | eraser     | sofa       | glass       |
| (5) REMEMBER    | swim        | recall     | number     | defy        |
| (6) TUMBLE      | drink       | dress      | fall       | think       |
| (7) HIDEOUS     | silvery     | tilted     | young      | dreadful    |
| (8) CORDIAL     | swift       | muddy      | leafy      | hearty      |
| (9) EVIDENT     | green       | obvious    | skeptical  | afraid      |
| (10) IMPOSTOR   | conductor   | officer    | book       | pretender   |
| (11) MERIT      | deserve     | distrust   | fight      | separate    |
| (12) FASCINATE  | welcome     | fix        | stir       | enchant     |
| (13) INDICATE   | defy        | excite     | signify    | bicker      |
| (14) IGNORANT   | red         | sharp      | uninformed | precise     |
| (15) FORTIFY    | submerge    | strengthen | vent       | deaden      |
| (16) RENOWN     | length      | head       | fame       | loyalty     |
| (17) NARRATE    | yield       | buy        | associate  | tell        |
| (18) MASSIVE    | bright      | large      | speedy     | low         |
| (19) HILARITY   | laughter    | speed      | grace      | malice      |
| (20) SMIRCHED   | stolen      | pointed    | remade     | soiled      |
| (21) SQUANDER   | tease       | belittle   | cut        | waste       |
| (22) CAPTION    | drum        | ballast    | heading    | ape         |
| (23) FACILITATE | help        | turn       | strip      | bewilder    |
| (24) JOCOSE     | humorous    | paltry     | fervid     | plain       |
| (25) APPRISE    | reduce      | strew      | inform     | delight     |
| (26) RUE        | eat         | lament     | dominate   | cure        |
| (27) DENIZEN    | senator     | inhabitant | fish       | atom        |
| (28) DIVEST     | dispossess  | intrude    | rally      | pledge      |
| (29) AMULET     | charm       | orphan     | dingo      | pond        |
| (30) INEXORABLE | untidy      | involatile | rigid      | sparse      |
| (31) SERRATED   | dried       | notched    | armed      | blunt       |
| (32) LISSOM     | moldy       | loose      | supple     | convex      |
| (33) MOLLIFY    | mitigate    | direct     | pertain    | abuse       |
| (34) PLAGIARIZE | appropriate | intend     | revoke     | maintain    |
| (35) ORIFICE    | brush       | hole       | building   | lute        |
| (36) QUERULOUS  | maniacal    | curious    | devout     | complaining |
| (37) PARIAH     | outcast     | priest     | lentil     | locker      |
| (38) ABET       | waken       | ensue      | incite     | placate     |
| (39) TEMERITY   | rashness    | timidity   | desire     | kindness    |
| (40) PRISTINE   | vain        | sound      | first      | level       |

Copyright © 1939 by The Institute of Living, The Neuro-Psychiatric Institute of the Hartford Retreat.  
Copyright © renewed 1967 by Barbara Shipley Boyle.  
Not to be reproduced in whole or in part without written permission of Western Psychological Services.  
All rights reserved. 23436739 Printed in U.S.A.



## ***Appendix D***

### **Introductory Letter to Centres and Hospitals**

Ian Clara  
University of Manitoba  
Winnipeg, Manitoba  
R3T 2N2  
204-261-4777  
email: [esquire@surf.pangea.ca](mailto:esquire@surf.pangea.ca)

To:

Re: Psychometric Assessments of the Shipley Institute of Living Scale with the WAIS-III and WAIS-III short forms for an Aging Sample of Clinical Patients

Dear \_\_\_\_\_

I am a student pursuing a Master's degree in Experimental Psychology at the University of Manitoba. The area of my particular study is in the psychometric properties of intelligence tests. This area interests me since I would like to evaluate intelligence tests in the elderly population, specifically those persons aged 65 years or older. Some of the most widely used intelligence tests have not been adequately evaluated with respect to the aging population, and I would like to fill this gap in the research area. It would be very beneficial to become aware of any limitations that need to be considered when using such tests with the elderly.

I will be focussing on the psychometric properties of three types of intelligence tests — the Wechsler Adult Intelligence Scale — Version III (WAIS-III), short forms of the WAIS-III, and the Shipley Institute of Living Scale (SILS). Reliability is the extent to which a test is consistent in what it measures. Using the WAIS-III as a “measuring stick”, I want to compare the short forms of the WAIS-III and the SILS in terms of how accurate they are in estimating a person's full-scale IQ (the IQ score they would have obtained if we had given them the entire WAIS-III). I intend to use the WAIS-III as a measuring stick since all of the short forms and the SILS are able to have their respective scores converted in to estimates of the full-scale WAIS-III IQ score. The validity of a test is the judgement of how a test measures what it is supposed to measure. Using the WAIS-III as a measuring stick for intelligence again, I want to determine if the short forms of the WAIS-III and the SILS are actually giving accurate estimates of intelligence scores, rather than the accuracy in estimation being due to some mathematical fluke.

An added incentive for my interest in this area is with the amount of time taken for the administration of the tests. The full-scale WAIS-III is a very lengthy test, which might have unfair consequences when administered to the aging population. part of the

reasoning for investigating the WAIS-III short forms and the SILS is that they take a substantially lower time to administer, and if it can be determined that these alternate testing forms also have good reliability and validity with the full-scale WAIS-III, then the use of the alternate forms as a way of obtaining an accurate estimate of the full-scale WAIS-III IQ score would be advantageous.

I have been talking about the WAIS-III and the SILS, and now I would like to take this chance to give a brief explanation of these tests. The WAIS-III is a recent update of the Wechsler Adult Intelligence Scale — Revised (WAIS-R), having just been released in the summer of last year. The WAIS-III contains 14 subtests, and these subtests can be combined to give three separate scores — a full-scale IQ score, a Performance IQ score, and a Verbal IQ score. The short forms of the WAIS-III that I wish to use are composed of combinations of two, three, or four subtests. The subtests involve either a written, verbal, or object-manipulation response on the part of the participant. Once I have administered the full WAIS-III, I will not need to re-administer the separate short forms, since their scores can be compiled from the individual subtest scores on the full-scale test. The SILS is a twenty-minute pencil-and-paper test, asking the participant to either circle an answer or fill in a series of blanks.

I would greatly appreciate the chance to work with you and your institution. The University of Manitoba Psychology Department has already given ethical approval for my study, but please let me know of any forms or requirements that need to be met at your institution. Participation from elderly individuals is on a volunteer basis only, and participants will be free to discontinue their participation at any time, and they will be informed of the aims and nature of the study beforehand. I will be utilizing the Folstein Mini Mental State Examination as a screening device to ensure that participants will be able to complete the tests. I will also require some background information from the participants (i.e., age, gender, educational level). For your perusal, a copy of the Introductory Letter that I would be using for the participants is also included in this package. Other than completing the two intelligence tests described above, no other requirements will be asked of the participants.

If you believe that you can help me with this project, please feel free to contact me at 261-4777, email me at [esquire@surf.pangea.ca](mailto:esquire@surf.pangea.ca), or reply by mail. I offer my thanks for considering my study, and I hope that you will accept my invitation to participate in this project and help me contribute something to the academic community.

Sincerely,

Ian Clara

*Appendix E*

Copy of the Mini-Mental State Examination

ID#  
 Examiner  
 Date

Mini-Mental State Examination

Maximum  
 Score      Score

ORIENTATION

- 5      (    )    What is the (year) (season) (date) (day) (month)?
- 5      (    )    Where are we: (state) (county) (town) (hospital) (floor).

REGISTRATION

- 3      (    )    Name 3 objects: 1 second to say each. Then ask the patient all 3 after you have said them. Give 1 point for each correct answer. Then repeat them until he learns all 3. Count trials and record.

ATTENTION AND CALCULATION

- 5      (    )    Serial 7's. 1 point for each correct. Stop after 5 answers. Alternatively spell "world" backwards.

RECALL

- 3 ( ) Ask for the 3 objects repeated above. Give 1 point for each correct.

**LANGUAGE**

- 9 ( ) Name a pencil, and watch (2 points)  
Repeat the following "No ifs, ands or buts." (1 point)

Follow a 3-stage command:

"Take a paper in your right hand, fold it in half, and put it on the floor" (3 points)

Read and obey the following:

CLOSE YOUR EYES (1 point)

Write a sentence (1 point)

Copy design (1 point)

\_\_\_\_\_ Total Score

ASSESS level of consciousness along a continuum

\_\_\_\_\_

	Alert	Drowsy	Stupor
--	-------	--------	--------

Coma

Source:

Folstein, M., Folstein, S. & McHugh, P. (1975). "Mini-Mental State" A practical method for grading the cognitive state for the clinician. Journal of Psychiatric Research, 12, 189-198.

*Appendix F*Instructions for Administration of the Mini-Mental State Examination**ORIENTATION**

- (1) Ask for the date. Then ask specifically for parts omitted, e.g., "Can you also tell me what season it is?" One point for each correct.
- (2) Ask in turn "Can you tell me the name of this hospital?" (town, county, etc.). One point for each correct.

**REGISTRATION**

Ask the patient if you may test his memory. Then say the names of 3 unrelated objects, clearly and slowly, about one second for each. After you have said all 3, ask him to repeat them. This first repetition determines his score (0-3) but keep saying them until he can repeat all 3, up to 6 trials. If he does not eventually learn all 3, recall cannot be meaningfully tested.

**ATTENTION AND CALCULATION**

Ask the patient to begin with 100 and count backwards by 7. Stop after 5 subtractions (93, 86, 72, 65). Score the total number of correct answers.

If the patient cannot or will not perform this task, ask him to spell the word "world" backwards. The score is the number of letters in correct order. E.G. dlrow = 5, dlrow = 3:

**RECALL**

Ask the patient if he can recall the 3 words you previously asked him to remember. Score 0-3.

**LANGUAGE**

**Naming:** Show the patient a wrist watch and ask him what it is. Repeat for pencil. Score 0-2.

**Repetition:** Ask the patient to repeat the sentence after you. Allow only one trial. Score 0 or 1.

**3-Stage command:** Give the patient a piece of plain blank paper and repeat the command. Score 1 point for each part correctly executed.

**Reading:** On a blank piece of paper print the sentence "close your eyes", in letters large enough for the patient to see clearly. Ask him to read it and do what it says. Score 1 point only if he actually closes his eyes.

**Writing:** Give the patient a blank piece of paper and ask him to write a sentence for you. Do not dictate a sentence, it is to be written spontaneously. It must contain a subject and verb and be sensible. Correct grammar and punctuation are not necessary.

**Copying:** On a clean piece of paper, draw intersecting pentagons, each side about 1 in., and ask him to copy it exactly as it is. All 10 angles must be present and 2 must intersect to score 1 point. Tremor and rotation are ignored.

Estimate the patient's level of sensorium along a continuum, from alert on the left to coma on the right.

**Source:**

Folstein, M., Folstein, S. & McHugh, P. (1975). "Mini-Mental State" A practical method for grading the cognitive state for the clinician. Journal of Psychiatric Research, 12, 189-198.

## *Appendix G*

### Introductory Letter to Participants

Ian Clara  
P411 Duff Roblin  
University of Manitoba  
R3T 2N2  
204-261-4777

ID #:

Greetings.

My name is Ian Clara, and I am a student at the University of Manitoba. I am writing my Masters thesis for the Department of Psychology, and I would like to ask for your help in becoming a participant for this research. I am very much interested in the use of intelligence testing in people over the age of 65 years. Many of the intelligence tests that are used today have not been adequately studied with people in this age group, and I would like to fill in this gap.

My survey is very simple. I would like to administer two intelligence tests — the Wechsler Adult Intelligence Scale — Version III (WAIS-III), and the Shipley Institute of Living Scale (SILS). Then, from the WAIS-III, I will compute what are called “short form scores”, which are simply scores from two, three, or four parts of the WAIS-III (it has a total of 14). Then, I will look at the short form scores and the SILS scores to see if they can accurately estimate the score that was achieved on the full-scale WAIS-III. The reason that I want to look at this is that the WAIS-III can be very lengthy, so if a way can be found to get an accurate estimate of a person’s IQ that takes less time, then that would be beneficial to everybody. Another part of this study involves making sure that these shorter measures actually do what I think they will do — namely, do they really give accurate estimates or is there some sort of mathematical trickery going on?

If you choose to participate, then all that would be required is roughly two hours of your time. Participation is completely voluntary. I will be using a short screening test prior to the actual testing, but this is only to ensure that you will be able to understand the questions that I will be asking you. You will be free to stop the study and take a break at any time. If at some point you wish to leave the study then you will be free to do that as well. For confidentiality purposes, I have given each letter a coded ID number. This will be used in place of your name, so that you will remain anonymous throughout the study. Your name and signature will be required on a consent form, but these will be kept in a locked file that can only be accessed by myself.

In order to make participation as easy as possible, I will travel to a location that is convenient for you to perform the testing. You will not have to come to the University of Manitoba. I have included my phone number both on the top of this letter and on the consent form. If you wish to participate, you can call me directly and arrange an appointment time.

I would greatly appreciate the chance to work with you. I believe that answering questions through research is an invaluable process, and I would genuinely enjoy working through this process with you. I thank you in advance.

Sincerely,  
Ian Clara

**Appendix H**

Ian Clara  
 P411 Duff Roblin  
 University of Manitoba  
 R3T 2N2  
 204-261-4777

**Consent Form** ID#:

By signing this form, I acknowledge the following:

- 1) that any personal information included in this study (my name and phone number and any other information) shall be used only within this study and not for any other purpose without my express consent.
- 2) that at any time during the interviews and testing I can declare that I do not wish to continue and I can withdraw from the study, or postpone the session until a later time.
- 3) that I will be given a full debriefing regarding the exact nature of the study and what the goals of the study are.
- 4) in order for the primary researcher to contact me in order to arrange for sessions, I will provide a phone number that is to be used for this purpose only
- 5) that the session shall be carried out (please circle one):
  - a) at the University of Manitoba
  - b) at my place of residence
  - c) at a place of my choosing

By signing this form I agree to the above conditions, and hold the researcher responsible for ensuring that these conditions are met.

Date: \_\_\_\_\_ Name: \_\_\_\_\_  
 (please print)

Signature: \_\_\_\_\_

---

This section will be detached and returned to the participant at the end of the testing session. It is an optional section, to be filled out ONLY if the participant wishes to leave a phone number to be contacted with.

Name: \_\_\_\_\_ Phone Number: \_\_\_\_\_  
 (optional)

## *Appendix I*

### **Debriefing Form**

I would first like to thank you for participating in this study. Without your help and cooperation it could not have been completed. I want to remind you that any personal information obtained is strictly confidential. The data gathered in this study are for analytical purposes only, and will not be used to reach any sort of decision that would affect you personally. It is not the nature of this study to evaluate individuals. I am now going to explain why I wanted to perform this study, and you can feel free to ask any questions that you may have.

This particular study has three goals. The first goal is to determine if a shorter form of the WAIS-III could be used instead of the full WAIS-III, since the administration time of the full WAIS-III can take up to 90 minutes. The WAIS-III is a standardized intelligence scale composed of 14 subtests. Each of these subtests can be grouped in certain ways to give what is called a Verbal and a Performance IQ score, as well as a full-scale IQ score. From the administration of the full WAIS-III it is hoped that certain subtests from both the Verbal and Performance areas can be used that will give a reasonable estimate of the full-scale WAIS-III score, and so help researchers and test takers by reducing the amount of time required to administer the test.

The second goal is to investigate the ability of the Shipley Institute of Living Scale to act as a measure of intelligence, when compared against the WAIS-III. Since the Shipley takes so little time to administer, and is very simple to administer, it is advantageous to know how well it can perform against such a standardized test as the WAIS-III. It was also part of this study to see if the conversion strategies that were used with earlier versions of the WAIS-III would still give accurate estimates of full-scale WAIS-III scores.

The third goal is to look at differences between the Verbal and Performance scales of the WAIS-III. It is our hope that we can present a method that will help clinicians and test evaluators better evaluate test takers, and also lead to more consistent decisions.

I would like to take this time now to let you ask any questions that you may have, or to give me any comments about the study.

If you have any further question, please feel free to contact me at 261-4777.

Thank you once again for your participation.

Sincerely,

Ian Clara

## *Appendix J*

### **Feedback Letter for the Participants**

Dear Participant,

My name is Ian Clara, and a short while ago you participated in a study that helped me to complete my Masters thesis in Psychology at the University of Manitoba. I would like to once again thank you for your participation. Now that the study has been completed, I can share some of the results with you.

There were two scales given in the study — the Wechsler Adult Intelligence Scale III (WAIS-III) and the Shipley Institute of Living Scale (SILS). My study focussed on the efficiency of shorter versions of the WAIS-III, as well as investigating the ability of the SILS to predict a person's score on the WAIS-III. Some of the group test results are given in the table below.

Age group	Mean WAIS-III score		Mean SILS score
	short	long	
65 - 69			
70 - 74			
75 - 79			
80 - 84			
85 - 89			

From the data that was collected, it was determined that a shorter version of the WAIS-III, comprising of the (*insert relevant subtests*) subscales of the WAIS-III, gives the best estimate of the full-scale WAIS-III score.

It was also determined that the SILS (*could / could not*) adequately estimate the full-scale WAIS-III score.

A third area of investigation was the comparison of what are called “comparable confidence intervals” to those of conventional confidence intervals. A confidence interval is an area around a number that, to some degree of certainty, we can be sure that the actual value of the mean or some other value really lies within that range. It was determined that comparable confidence intervals (*do / do not*) perform better than conventional confidence intervals. This has relevance to test administrators in that they will be able to have more precise measurements from the tests and scales that they use in the course of their profession.

If you have any questions about the study, or if you would like more information that what is presented here, please do not hesitate to contact me at 261-4777. I would be more than happy to answer your questions.

Sincerely,  
Ian Clara

## *Appendix K*

### Feedback From for Administrators and Institutions

Dear (Administrator Name),

My name is Ian Clara, and a short while ago your institution participated in a study that helped me to complete my Masters thesis in Psychology at the University of Manitoba. I would like to once again thank you for your participation. Now that the study has been completed, I can share some of the results with you.

There were two scales given in the study — the Wechsler Adult Intelligence Scale III (WAIS-III) and the Shipley Institute of Living Scale (SILS). In the investigation of the WAIS-III, I looked at the use of “short forms,” which are selected subtests of the WAIS-III that can hopefully be used to give accurate full-scale IQ scores without the need to administer the complete WAIS-III. There were four short forms evaluated, and I give the mean results in the table below.

Short Form Type	Mean Admin Time (min)	Mean Short Form Score	Mean Short Form Estimate of IQ	Mean Actual Full-Scale IQ	Reliability
Vocabulary and Block Design					
Vocabulary, Block Design, Information					
Vocabulary, Block Design, Arithmetic, Picture Arrangement					
Information, Arithmetic, Picture Arrangement, Block Design					

The second part of the study involved the SILS. The SILS has been used to give estimates of full-scale IQ on previous versions of the WAIS. We wanted to investigate if this relationship still held with the newest version of the WAIS, the WAIS-III., and also if any changes needed to be made to the conversion strategy used to compute the estimate. There were two different methods of estimating the full-scale WAIS-R IQ from the SILS. The mean results are given below.

Estimation Procedure	Mean SILS Score	Mean Full-Scale WAIS-III IQ Estimate	Mean Full-Scale WAIS-III IQ Full Score
----------------------	-----------------	--------------------------------------	--

(1)  $WAIS = 0.92 (SILS) + 63.10$   
(Paulson and Lin, 1970)

(2) procedure outlined by Zachary, Crumpton, and Spiegel (1985)

The third part of the study investigated what is called the VIQ-PIQ difference. The WAIS-III (and earlier versions of the WAIS) actually yields three scores — a full-scale IQ score, and a Verbal IQ score and a Performance IQ score. The VIQ and PIQ together make up the full-scale IQ score. Some researchers have looked at this VIQ-PIQ difference as a “rule of thumb” for indication of a deficit. It is of some interest to know if the short forms, being composed of a smaller number of subtests than the complete test battery, can produce a pattern of VIQ-PIQ differences that is similar to the full-form Wechsler. The current results found that the Ward et al. (1987) 4-subtest short form could produce estimates of the VIQ-PIQ difference that were quite similar to those found with the entire battery.

Once again, I would like to thank you and your institution for aiding me in the completion of my MA requirements. I can honestly say that I could not have completed this study without your help. If you have any questions, or if you would like more information about the procedures described above, please do not hesitate to contact me by phone (261-4777) or email ([esquire@surf.pangea.ca](mailto:esquire@surf.pangea.ca)).

Sincerely,

Ian Clara