UNIVERSITY OF
CALGARY

The author of this thesis has granted the University of Calgary a non-exclusive license to reproduce and distribute copies of this thesis to users of the University of Calgary Archives.

Copyright remains with the author.

Theses and dissertations available in the University of Calgary Institutional Repository are solely for the purpose of private study and research. They may not be copied or reproduced, except as permitted by copyright laws, without written authority of the copyright owner. Any commercial use or publication is strictly prohibited.

The original Partial Copyright License attesting to these terms and signed by the author of this thesis may be found in the original print version of the thesis, held by the University of Calgary Archives.

The thesis approval page signed by the examining committee may also be found in the original print version of the thesis held in the University of Calgary Archives.

Please contact the University of Calgary Archives for further information,

E-mail: uarc@ucalgary.ca

Telephone: (403) 220-7271

Website: <http://www.ucalgary.ca/archives/>

THE UNIVERSITY OF CALGARY

A Comparison of WISC-III Short Forms
for the Screening of Gifted Elementary Students in Canada

by

Betty Anne Reiter

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DIVISION OF APPLIED PSYCHOLOGY

CALGARY, ALBERTA

OCTOBER, 2001

© Betty Anne Reiter 2001

ABSTRACT

This investigation explored the accuracy of six short forms of the WISC-III in estimating the Full Scale IQ of potentially gifted elementary students in Canada. Data from the WISC-III Canadian standardization study served as the analysis sample ($n = 192$), while WISC-III data obtained from the psychology files of a large urban school division in Western Canada was used for the cross validation sample ($n = 1,058$).

The results of the predictive discriminant analyses indicated that all six short forms separated gifted students from non-gifted students with a high degree of accuracy. Gender and chronological age of the subject did not affect the discriminant accuracy of the short forms investigated. Short form IQ scores were then calculated for each of the six short forms via two different methods: linear equating (Tellegen & Briggs, 1967) and standard multiple regression. An examination of the mean differences between short form IQ scores and Full Scale IQ scores suggested that both methods were equally effective with the gifted students in this study.

When psychometric soundness and clinical utility were considered together with discriminating power, the Dumont-Faro short form of the WISC-III (Information, Vocabulary, Picture Completion, Coding, and Block Design) emerged as providing the best compromise between all factors. This short form, which accurately classified gifted and non-gifted students at a rate of 97.7%, was found to have excellent reliability (.921) and validity (.906). Pearson product-moment coefficients revealed that it yielded short form IQ scores that correlated highly (.967) with Full Scale IQ scores. Paired t-tests indicated that short form IQ scores rendered from the Dumont-Faro were closer to the actual scores than the other short forms investigated. Moreover, there were relatively few

cases where the short form IQ scores either underestimated or overestimated the Full Scale IQ scores. The Dumont-Faro short form was estimated to reduce the typical WISC-III administration time by at least half, while at the same time providing a strong measure of three broad areas of intelligence (e.g., crystallized intelligence, visual processing ability, and processing speed).

It was concluded that the Dumont-Faro short form may be used as an alternative to full battery assessment for screening potentially gifted elementary students in Canada. Recommendations for the practical application of this short form to school psychology practice are provided.

ACKNOWLEDGEMENTS

I would like to formally acknowledge The Psychological Corporation for compiling the WISC-III Canadian standardization data that was used in this investigation.

I would also like to acknowledge the Calgary Board of Education for allowing me to access the archival WISC-III data that was used for the cross validation sample.

I wish to extend my sincere appreciation to my supervisor, Dr. Marilyn Samuels, for guiding, challenging, and encouraging me throughout my doctoral program. Dr. Samuels has been a true mentor to me, both academically and professionally. I would like to express my gratitude to Dr. Michael Pyryt for willingly sharing his time and his statistical expertise throughout this study. Dr. Pyryt's enthusiasm for conducting research, coupled with his ability to teach others about the statistical methods of inquiry, has been an inspiration to me. I extend my appreciation and thanks to Dr. Judy Lupart for her insightful examination of this dissertation. I would also like to acknowledge Dr. Lupart for her guidance in helping me to develop a greater understanding of giftedness. I thank Dr. Tim Rogers for his interest in the study and for serving as a committee member. I would like to acknowledge my external examiner, Dr. Don Saklofske. Dr. Saklofske's enthusiasm for my research, coupled with his thought provoking questions, helped to make my dissertation oral examination an extremely rewarding experience.

I am greatly indebted to my brother, Robert Reiter, for his editorial assistance in the preparation of this dissertation. Last but not least, special acknowledgement and gratitude is extended to my husband, Patrick Hartigan. Thank you for believing in me.

DEDICATION

This dissertation is dedicated to my parents, Adel and Al Reiter. Without their encouragement, support, and unfailing confidence, I could not have achieved this goal.

TABLE OF CONTENTS

Approval Page.....	ii
Abstract.....	iii
Acknowledgements.....	v
Dedication.....	vi
Table of Contents.....	vii
List of Tables.....	x
List of Figures.....	xiii
 CHAPTER ONE: INTRODUCTION.....	 1
Identification and Assessment of Gifted Students.....	1
Individual Intelligence Tests Widely Used with Gifted Students.....	2
Statement of the Problem	3
Merits and Limitations of Short Forms	5
Purpose of the Current Investigation	6
Operational Definitions.....	7
Assumptions	8
Delimitations of the Current Investigation.....	8
Overview of the Current Investigation.....	8
 CHAPTER TWO: LITERATURE REVIEW	 9
Introduction to Short Forms of Individual Intelligence Tests.....	9
Types of Short Forms.....	11
Selected Subtest Short Forms	11
Selected Item Short Forms	12
Selected Subtest versus Selected Item Short Forms	13
Time Saved by Employing a Short Form	15
Evaluating the Validity of Short Forms	17
Reliability of Short Forms	20
Methodological Issues in Short Form Research	21
Development of Short Forms Based on Standardization Data	21
Deriving Short Form Intelligence Quotients	21
Rescoring Protocols to Obtain Short Form Data.....	24
Short Form Subtest Scatter	26
Summary of Short Form Research Issues	27
Short Form Research with the Wechsler-Bellevue Intelligence Scales	28
Short Form Research with the Wechsler Intelligence Scale for Adults (WAIS) and the Wechsler Intelligence Scale for Adults-Revised (WAIS-R)	29
Short Form Research with the Wechsler Intelligence Scale for Children (WISC).....	33
Short Form Research with the Wechsler Intelligence Scale for Children- Revised (WISC-R).....	34

WISC-R Short Form Research with Non-Gifted Populations	34
WISC-R Short Form Research with Gifted Populations	36
Short Form Research with the Wechsler Intelligence Scale for Children-	
Third Edition (WISC-III)	41
WISC-III Short Form Research with Non-Gifted Populations	41
WISC-III Short Form Research with Gifted Populations	49
Summary.....	52
Hypothesis for the Current Investigation	53
Research Questions.....	53
Rationale for the Selection of Short Forms Used in the Current	
Investigation	53
Short Forms Used in the Current Investigation	54
 CHAPTER THREE: METHODOLOGY	 63
Instrumentation	63
Standardization of the WISC-III	63
Reliability and Validity of the WISC-III	64
Importance of Canadian Normative Data	65
Data Collection Procedure	65
Data Analysis	67
 CHAPTER FOUR: RESULTS	 69
Introduction.....	69
Descriptive Statistics	69
Discriminant Analysis	70
Classification Results for the Total Analysis Sample.....	84
Classification Results for the Analysis Sample by Gender.....	85
Classification Results for the Analysis Sample by Age.....	85
Linear Equating Technique	89
Multiple Regression Analysis	89
Pearson Product-Moment Correlation Coefficients.....	90
Paired t-Tests	92
Psychometric Properties of the Short Forms	99
Frequency of Underestimated and Overestimated Scores	101
Cross Validation	103
Discriminant Analysis Classification Results.....	106
Pearson Product-Moment Correlation Coefficients.....	106
Paired t-Tests.....	108
Frequency of Underestimated and Overestimated Scores.....	113
Summary.....	116
 CHAPTER FIVE: DISCUSSION	 118
Summary	118

Discriminant Analysis	118
Validity of the Short Forms	121
Reliability of the Short Forms	123
Clinical Value of the Short Forms	124
Best Short Form for the Population in the Current Investigation	129
Linear Equating vs. Multiple Regression	132
Subtest Scatter	133
Practical Application of the Findings from the Current Investigation	133
Implications for School Psychology Practice	137
Limitations of the Current Investigation	139
Directions for Future Research	139
Conclusion	140
REFERENCES	143
APPENDIX A: WISC-III Subtest Administration Times for Gifted Elementary Students.....	165
APPENDIX B: Linear Equating Formulae based on the Analysis Sample Data.....	166
APPENDIX C: Regression Equations based on the Analysis Sample Data.....	168
APPENDIX D: Formulae Used in the Current Investigation.....	169

List of Tables

Table 1	Summary of the Short Forms Used in the Current Investigation.....	56
Table 2	Size of the Analysis Sample and the Cross Validation Sample by Chronological Age and Gender	71
Table 3	Means and Standard Deviations for WISC-III Subtests and the IQ Scales by Classification for the Analysis Sample	72
Table 4	Means and Standard Deviations for WISC-III Subtests and IQ Scales by Classification for the Cross Validation Sample.....	73
Table 5	Summary of Discriminant Analyses Between Gifted and Non-Gifted Students	75
Table 6	Summary of Discriminant Analyses Between Gifted and Non-Gifted Males	77
Table 7	Summary of Discriminant Analyses Between Gifted and Non-Gifted Females	78
Table 8	Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages 6 – 7	79
Table 9	Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages 8 – 9	80
Table 10	Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages 10 – 12	81
Table 11	Summary of Standardized Discriminant Function and Structure Coefficients for the Analysis Sample	83
Table 12	Classification Results for the Analysis Sample	86
Table 13	Classification Results for the Analysis Sample by Gender	87
Table 14	Classification Results for the Analysis Sample by Chronological Age.....	88
Table 15	Standard Multiple Regression Summary for the Analysis Sample.....	91

Table 16	Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Analysis Sample	93
Table 17	Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Analysis Sample	94
Table 18	Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Analysis Sample	96
Table 19	Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Analysis Sample.....	97
Table 20	Effect Size for Group Differences Based on the Analysis Sample...	98
Table 21	Psychometric Properties of the Six Short Forms Derived from the Standardization Data Reported in the WISC-III Canadian Manual...	100
Table 22	Short Form Standard Error of Estimate Values and Confidence Intervals By Technique Based on the Analysis Sample	102
Table 23	Frequency of Linear Equating Derived SFIQ Scores Falling Outside the 95% Confidence Interval for the Analysis Sample	104
Table 24	Frequency of Multiple Regression Derived SFIQ Scores Falling Outside the 95% Confidence Interval for the Analysis Sample	105
Table 25	Classification Results for the Cross Validation Sample	107
Table 26	Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Cross Validation Sample	109
Table 27	Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Cross Validation Sample	110
Table 28	Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Cross Validation Sample....	111
Table 29	Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Cross Validation Sample	112

Table 30	Effect Size for Group Differences Based on the Cross Validation Sample	114
Table 31	Frequency of Linear Equating Derived SFIQ Scores Falling Outside the 95% Confidence Interval for the Cross Validation Sample	115
Table 32	Frequency of Multiple Regression Derived SFIQ Scores Falling Outside the 95% Confidence Interval for the Cross Validation Sample	117

List of Figures

- Figure 1 Flowchart depicting the practical application of the Dumont-Faro short form for elementary students who are referred for giftedness.. 138

CHAPTER ONE: INTRODUCTION

Identification and Assessment of Gifted Students

Potentially gifted students are typically referred to school psychologists to determine whether they are eligible to participate in programming for the gifted (Gallagher, 1996; Piirto, 1999). The purpose of identifying gifted students is to find those individuals who require differentiated educational programming in order to fully realize their potential (Assouline, 1997; Feldhusen, 1998; Feldhusen & Jarwan, 1993; Gallagher, 1996; Kaufman, 1992; Pyryt, 1996; Sparrow & Gurland, 1998). Ideally, the identification process should highlight students whose needs, abilities, and characteristics fit the services that are offered (Feldhusen, 1998; Feldhusen & Jarwan, 1993; Louis, Subotnik, Breland, & Lewis, 2000; Shore, Cornell, Robinson, & Ward, 1991). For example, the Calgary Board of Education offers a special program designed for students who are “intellectually gifted” (Calgary Board of Education, 1999). Within this framework of giftedness, it is appropriate for school psychologists to use assessment methods that focus on cognitive and academic abilities (Assouline, 1997).

Individually administered intelligence tests are the most reliable and valid tools available for identifying potentially gifted students (Assouline, 1997; Feldhusen, 1998; Kaufman & Harrison, 1986; Klausmeier, Mishra, & Maker, 1987; Pyryt, 1996; Sparrow & Gurland, 1998). These instruments represent an objective and well-standardized means of assigning a quantitative value to a student’s present abilities and future academic achievement (Anastasi & Urbina, 1997; Harrison, 1997; Reschly, 1997). While no test is able to measure every attribute of intelligent behaviour, an intelligence test is the single best indicator of a student’s range of cognitive knowledge at a given point in

time (Assouline, 1997; Feldhusen, 1998; Kaufman & Harrison, 1986; Saccuzzo & Johnson, 1995; Sattler, 2001).

Although individually administered intelligence tests play a prominent role in the identification of potentially gifted students, for educational decision making it is important that these instruments are used in conjunction with other identification techniques (Gallagher, 1996; Kaufman, 1992; Runco, 1997; Sabatino, Spangler, & Vance, 1995; Silverman, 1998; Spangler & Sabatino, 1995). Additional methods of identifying gifted students might include any combination of the following: teacher and parent nominations (Boatman, Davis, & Benbow, 1995; Feldhusen, 1998; Gross, 1999; Jin & Feldhusen, 2000); gifted rating scales and checklists (Howley, Howley, & Pendarvis, 1995; Piirto, 1999); superior grades (Hadaway & Marek-Schroer, 1997; Sternberg, 1985); portfolios (Borland & Wright, 1994; Coleman, 1994; Hadaway & Marek-Schroer, 1997); standardized achievement tests (Feldhusen, 1998; Feldhusen & Jarwan, 1993; Glascoe, 1996; Mantzicopoulos, 2000); above-level testing (Assouline & Lupkowski-Shoplik, 1997; Silverman, 1998; Stanley, 1997); group intelligence tests (Assouline, 1997; Feldhusen, 1998; Feldhusen & Jarwan, 1993); and dynamic assessment (Kirschenbaum, 1998; Maker, 1996; Sarouphim, 1999; Zorman, 1997).

Individual Intelligence Tests Widely Used with Gifted Students

Two of the most common individual intelligence tests used to assess potentially gifted students include the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) and the Stanford-Binet Intelligence Scale-Fourth Edition (SB-IV). Due to differences in test construction (e.g., reliability, validity, normative data, ceiling items), however, these

instruments vary with respect to their effectiveness in assessing gifted students (Kluever & Green, 1990; Tyler-Wood & Carri, 1991).

Historically, the third version of the Stanford-Binet (SB-LM) was the preferred measure when assessing potentially gifted students (Gross, 1992). The dominance of the Stanford-Binet Scales tumbled following the introduction of the SB-IV in 1986 as this revision has consistently been shown to yield significantly lower results for gifted students than the SB-LM (Kluever & Green, 1990; N. M. Robinson & Robinson, 1992; Silverman & Kearney, 1992). Furthermore, due to changes in test construction, the SB-IV has also been found to be less effective in distinguishing highly gifted students from moderately gifted students (Silverman, 1998).

At this time, the WISC-III is the most widely used intelligence test employed by school psychologists in North America (Prifitera, Weiss, & Saklofske, 1998; Reschly, 1997). Moreover, it is currently the test of choice for the assessment of potentially gifted students (Sevier, Bain, & Hildman, 1994; Sparrow & Gurland, 1998). The popularity of the WISC-III is not surprising given that its standardization procedures have been shown to be excellent and its psychometric properties set a high standard for other tests to follow (Braden, 1995; Edwards & Edwards, 1993; Kaufman, 1994; Sandoval, 1995). Sternberg (1993), one of the staunchest critics of formal assessment, admitted that “when all is said and done, the WISC-III is about as good as one gets with conventional intelligence tests” (p. 164).

Statement of the Problem

As the full battery of the WISC-III typically requires sixty to ninety minutes for a school psychologist to administer (Wechsler, 1991), individual intellectual assessment is

a labour intensive, time-consuming, and costly process. Abbreviated or “short form” versions of intelligence tests have been used as a means of obtaining an estimated Intelligence Quotient (IQ) score while involving less testing time (Sattler, 2001; Silverstein, 1990a; Watkins, 1986). Interest in short forms of a well-established instrument such as the WISC-III is due to the fact that this test is generally considered to be a highly reliable and valid measure for assessing intelligence in school-aged children. Brief tests of intelligence are available (e.g., Kaufman Brief Intelligence Test-KBIT, 1990; Wechsler Abbreviated Scale of Intelligence-WASI, 1999), but these instruments do not have the benefit of years of research backing to support their clinical value (Beal, Dumont, Branche, & Cruse, 1996; Reynolds, Willson, & Clark, 1983).

Although the initial intellectual assessment of students who have learning disabilities or cognitive delays involves in-depth evaluation of a broad range of skills and achievement, the assessment of gifted students typically relies on an overall IQ score (Mark, Beal, & Dumont, 1998). For example, one of the five criteria used to determine eligibility for gifted programming within the Calgary Board of Education is “Very Superior” (i.e., Full Scale IQ of 130 and above) ability. As a good short form of an intelligence test provides a reasonably reliable and valid estimate of overall cognitive functioning in less than half the time necessary to administer the full test, it is viable as a gifted screening instrument (Linville, Rust, & Kim, 1999; Mark et al., 1998).

While a substantial amount of research regarding the use of short forms with gifted populations has been conducted with the WISC-R (DiPasquale, Rosehart, Kempa, Rooney, & Steinman, 1992; Dirks, Wessels, Quarfoth, & Quenon, 1980; Elman, Blitz, & Sawicki, 1981; Fell & Schmidt-Fell, 1982; Fineman & Curran, 1986; Karnes & Brown,

1981; Killan & Hughes, 1978; Kramer, Shanks, Markley, & Ryabik, 1983; Linn & Lopatin, 1990; Lustberg, Motta, & Naccari, 1990; Ortiz & Gonzalez, 1989), there has been relatively little investigation involving the WISC-III (Linville et al., 1999). Furthermore, it appears that only one short form research study has employed Canadian WISC-III normative data (Mark et al., 1998). Given that Canadian students have been observed to perform differently on the WISC-III as compared to American students (Beal, Dumont, Cruse, & Branche, 1996; Wechsler, 1996), it is questionable as to whether short forms developed using the American standardization data are valid in Canada. According to the Standards for Educational and Psychological Testing (1999), prior to adopting any short form estimate of intelligence, the psychometric properties of the abbreviated measure must be examined to ensure that they are as similar as possible to the original test. Owing to the paucity of Canadian short form research, however, it is impossible to determine which abbreviations are valid for use in Canada.

Merits and Limitations of Short Forms

The use of short forms of intelligence tests may improve the efficiency of school psychology service on a number of different levels. First, the use of a short form could save time in terms of administering and scoring the instrument (Sattler, 2001; Silverstein, 1990a). Time saved from the administration of a short form may be used to conduct a more in-depth evaluation of student functioning (e.g., creativity, self-esteem) or to assess a greater number of students. From a practical point of view, this could result in a saving of school psychology resources. Second, short forms may be employed to screen students or to determine whether the administration of the full test battery is warranted (Linn & Lopatin, 1990; Silverstein, 1990a). Third, short forms may also be used when

re-evaluation of a student's cognitive abilities is necessary (Silverstein, 1990a; Zimet & Adler, 1990). Fourth, school psychologists are very familiar with the administration procedure involved with traditional full length intelligence tests, consequently using short forms of such instruments requires no additional training (Klein, 2001). Finally, administration of a short form could place less demand on the individual student being assessed (Linn & Lopatin, 1990).

With respect to the limitations of short forms, it must be acknowledged that these tools have decreased validity and reliability as compared to the full test battery (Sattler, 2001; Silverstein, 1990a). Moreover, they tend to yield high levels of misclassifications (Goh, 1979; Resnick & Entin, 1971). Consequently, they must never be used alone to make placement decisions. The use of a short form also limits profile analysis and may lead school psychologists to overlook cognitive strengths and weaknesses as certain subtests are omitted completely (Sattler, 2001; Silverstein, 1990a). Finally, employing a short form reduces the opportunity to observe the student and to gather clinically relevant information (Sattler, 2001; Silverstein, 1990a).

Purpose of the Current Investigation

The purpose of this study is to determine the accuracy of six WISC-III short forms in estimating the Full Scale IQ of potentially gifted elementary students in Canada. At this time, it appears that only one short form research study has been conducted using Canadian normative data. The effectiveness of six previously proposed short forms of the WISC-III will be compared utilizing the Canadian WISC-III standardization data. The WISC-III protocols from a sample of elementary students in a large urban school district in Canada will be used for cross validation purposes.

The availability of a psychometrically and clinically sound intelligence test short form for gifted students could assist school psychologists in meeting increased service demands while providing effective intellectual assessment. Furthermore, an investigation of this nature has a great deal of merit for Canadian school boards that consider IQ scores when determining which students are eligible for gifted programming. As assessing students is a time-consuming and costly process, a short form of the WISC-III could provide a reliable and valid estimate of ability without the need to administer the entire test battery. Moreover, it may also function as a screening device for determining when the remaining WISC-III subtests should be administered (i.e., for those students who fall within the area just below the gifted classification range) and when testing should be discontinued (i.e., for those students who are most likely not performing in the gifted range).

Operational Definitions

The following terms, derived from the literature, are offered in the interest of clarity.

Full Scale Intelligence Quotient (FSIQ): The overall WISC-III score based on the administration of 10 standard subtests which include: Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, and Comprehension.

Short Form Intelligence Quotient (SFIQ): A short form IQ score obtained by combining fewer than the standard 10 subtests of the WISC-III or by administering only certain items from the 10 standard subtests.

Assumptions

Although chartered psychologists with appropriate assessment credentials examined the students in the cross validation sample, the actual administrations of the WISC-III were not observed directly. Consequently, it was assumed that standardized administration procedures were followed when the WISC-III was given. The test protocols were not scrutinized for accuracy; however, it was assumed that protocols were scored correctly.

Delimitations of the Current Investigation

In this study, junior high and high school students were excluded. Moreover, ethnic/racial differences were not examined in this investigation. As the WISC-III Canadian standardization data did not separate the elementary students into specific grade levels, the variable of grade was not included in the present study.

Overview of the Current Investigation

In Chapter One, an introduction to the identification and assessment of gifted students has been provided. This chapter also outlined the background and purpose of the research study. The literature review offered in Chapter Two discusses the history of short form use with respect to types of short forms, evaluating the validity and reliability of short forms, time saved using short forms, and methodological issues related to short form research. An overview of short form investigations regarding the Wechsler Scales is also provided. Chapter Three contains a description of the methodology, while the results of the research findings are outlined in Chapter Four. Finally, the results of the current investigation are discussed in Chapter Five.

CHAPTER TWO: LITERATURE REVIEW

Introduction to Short Forms of Individual Intelligence Tests

There has been a long history of interest in the development and application of abbreviated or “short form” versions of individual intelligence tests. The first short form was proposed by Doll (1917) immediately following the 1916 publication of the Binet-Simon Scale in North America. According to Doll, the complete administration of an intelligence test was “too laborious, too complex, and too time-consuming” for the everyday practice of applied psychology (p. 197). Although almost a century has passed since the development of the first short form, time constraints related to the increased demand for psychological services continues to be cited as the primary reason for shortening intelligence test batteries (Linn & Lopatin, 1990; Lustberg et al., 1990). Despite the recent development of brief intelligence tests (e.g., Kaufman Brief Intelligence Test-KBIT; Wechsler Abbreviated Scale of Intelligence-WASI), the advantage of creating abbreviations of well-established individual intelligence tests lies in their impressive psychometric properties (Beal, Dumont, Branche, & Cruse, 1996; Reynolds et al., 1983).

From a pragmatic perspective, short form versions of intelligence tests make a great deal of sense for a number of reasons. They may be used as screening instruments to provide a quick method of estimating an individual’s global intellectual status (Anastasi & Urbina, 1997; Bersoff, 1971; Kaufman, 1990; Kaufman, Ishikuma, & Kaufman-Packer, 1991; Sattler, 2001; Silverstein, 1990a; Watkins, 1986; Zimet & Adler, 1990). They have also been used for re-evaluating the ability of special needs students (Silverstein, 1990a; Zimet & Adler, 1990) or for research purposes to determine the

overall intelligence levels for sample subjects (King & King, 1982; Sattler, 2001; Silverstein, 1990a). Short forms can also be administered to provide a snapshot view of an individual's intellectual status when cognition is not central to the referral question (Sattler, 2001; Silverstein, 1990a). Furthermore, these abbreviated tools place less demand on the individual being assessed (Linn & Lopatin, 1990).

When evaluating a short form, the following factors should be considered: amount of time saved (Levy, 1968; Sattler, 2001); level of validity (Kaufman et al., 1991; Kaufman, Kaufman, Balgopal, & McLean, 1996; Sattler, 2001; Silverstein, 1985a); level of reliability (Cyr & Brooker, 1984; Kline, 2001; Sattler, 2001; Silverstein, 1990b); provision of useful clinical information (Kaufman, 1972; Kaufman et al., 1991; Kaufman et al., 1996; Sattler, 2001; Silverstein, 1982); ability to answer the referral question (Sattler, 2001); and characteristics of the student (Sattler, 2001). Despite their apparent utility, short forms are limited in that they have decreased validity and reliability as compared to their full-length counterparts (Sattler, 2001; Silverstein, 1990a). Short forms also tend to yield high levels of misclassifications that can become a problem for decision-making purposes (Goh, 1979; Resnick & Entin, 1971). Moreover, the use of an abbreviated form reduces the opportunity to observe the examinee's approach to problem solving, in addition to providing less information about his or her cognitive strengths and weaknesses (Sattler, 2001; Silverstein, 1990a).

Educational decisions rely on precise measures, consequently, the sole use of short forms for placement, educational, or clinical decision-making purposes is strongly discouraged (Kaufman, 1994; Sattler, 2001; Silverstein, 1990a). Furthermore, Kaufman (1976) reported that "the complete battery, with its richer and more varied content, is the

instrument that should [be] administered if time is permitted” (p. 194). Nevertheless, short forms are viable for screening purposes in order to provide an overall estimate of a student’s level of intellectual functioning. Accordingly, the use of abbreviated forms with students who have been referred for gifted eligibility purposes appears to have merit. The utilization of a short form could enable school psychologists to screen a greater number of gifted referrals, in addition to economizing on time so that other types of assessments may be conducted for certain students.

Types of Short Forms

There are essentially two types of short forms: selected subtest and selected item. In order to understand the development of these two approaches, coupled with their relative merits and limitations, a brief explanation of each is warranted.

Selected Subtest Short Forms

The selected subtest approach involves administering a reduced number of subtests in their entirety. This is accomplished by using different combinations of two or more subtests in order to discover which set requires the shortest amount of time to administer, while at the same time maintaining as much statistical accuracy of the full test battery as possible. The selected subtest technique is based on the research that Rabin (1943) conducted with short forms of the Wechsler-Bellevue Intelligence Scale. Rabin used a rationally based method for deriving a short form whereby clinical experience and observation determined which subtests were the most indicative of the general mental level of the examinees. The scores from these subtests were then prorated to obtain the FSIQ score. Doppelt (1956) later proposed a statistically driven method for choosing which subtests to include in the short form. Within this model, subtest scores were used

in a regression equation (where differential weights are assigned to certain subtests) to estimate FSIQ.

Overall, selected subtest short forms provide a relatively quick means of yielding scaled scores from certain subtests, in addition to an overall SFIQ score. This type of abbreviation also provides the option of administering the remaining subtests should further assessment be warranted (Kaufman, 1977; Sattler, 2001; Silverstein, 1990a). In light of the fact that many subtests are omitted completely, this approach does not provide a full profile of ability, therefore, profile analysis and Verbal/Performance IQ estimates are not possible. Nevertheless, Zimet and Adler (1990) concluded that there are a number of valid and reliable selected subtest short forms that have a great deal of utility, provided the limitations of these instruments are taken into account.

Selected Item Short Form

The selected item approach involves the administration of certain items from most subtests (i.e., the exception on the Wechsler Intelligence Scale for Children is Coding and Digit Span which are given in their entirety), in order to derive a partial raw score for each subtest. Depending on the number of items used, the partial raw scores are then multiplied by a constant (e.g., 2 or 3) to yield full raw scores. Raw scores are then converted to scaled scores and used to derive SFIQ scores. Wolfson and Bachelis (1960) were the first investigators to use this technique. They developed a selected item abbreviation of the WAIS Verbal scale that sampled all the functions contained in the original scale. Satz and Mogel (1962) extended this research to create a selected item model that was representative of both the Verbal and Performance scales of the WAIS. The Satz-Mogel paradigm has served as the foundation for a variety of other selected

item short forms of the Wechsler scales (Cella, 1984; Finch, Ollendick, & Ginn, 1973; Goh, 1979; Hobby, 1982; Satz, Van de Riet, & Mogel, 1967; Yudin, 1966).

The primary advantage of utilizing a selected item short form is that it provides a sampling of a large range of intellectual functioning in a reduced amount of time. This type of abbreviation makes it possible to obtain estimated Verbal, Performance, and Full Scale IQ scores. While the selected item approach appears to have considerable merit as a short form, this procedure has been associated with the following shortcomings: loss of validity and reliability; questionable profile analysis of subtest patterns; and differences between the SFIQ estimates and the complete battery scores (Dean, 1977; Erikson, 1967; Gayton, Wilson, & Bernstein, 1970; Rasbury, Falgout, & Perry, 1978; Satz, Van de Riet, & Mogel, 1967; Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). An additional problem with the selected item short form is that it may only be used as a replacement for the complete test battery, as the nature of this method precludes the option of administering the remaining items should full test administration become necessary (Kaufman, 1977; Sattler, 2001; Silverstein, 1990a).

Selected Subtest versus Selected Item Short Forms

A review of the short form research literature suggests that there are a greater number of selected subtest studies than selected item studies. Finch, Childress, Wilkins, and Kendall (1974) were among the few researchers to directly compare selected subtest and selected item short forms. They discovered that the Yudin (1966) selected item short form and approximately half of the 30 WISC selected subtest short forms they investigated yielded significantly different mean scores as compared to the full form. It

was concluded that while certain selected subtest short forms had merit for screening purposes, this was not the case for the selected item approach.

Wechsler (1958) speculated that the use of the selected subtest abbreviated scale minimizes the effectiveness of individual assessment and “[t]he loss of qualitative observations and discrepancies on individual subtest performance is particularly significant” (p.113). Nevertheless, Silverstein (1990a) discovered that the best combinations of four or five selected subtests yielded validity coefficients comparable to those achieved from the selected item approach. Moreover, short form reliability appears to be superior in certain selected subtest abbreviations as compared to selected item short forms (Ryan, 1981; Silverstein, 1982; 1989a; 1990a, 1990b). Selected subtest short forms are also more useful as screening devices, in that the remaining subtests may be administered if warranted (Kaufman, 1977; Sattler, 2001; Silverstein, 1990a). Sattler (2001) noted that due to half of the test items being omitted in the selected item approach, the difficulty of items increases while the opportunity for practice decreases. He added that the significant departure from the standard administration procedure used in the selected item short form also makes it less desirable as compared to the selected subtest abbreviation.

With respect to the application of short forms with gifted populations, there is evidence to suggest that the higher the intellectual level, the lower the magnitude of the correlation between selected item SFIQ scores and FSIQ scores (Satz, Van de Riet, & Mogel, 1967). Consequently, the selected item short form approach may not be suitable for High Average to Superior ability students.

Time Saved by Employing a Short Form

The main purpose in constructing a short form of an intelligence test is to save time (Kaufman, 1990; Kaufman et al., 1991; Sattler, 2001; Silverstein, 1985a, 1990a). Regarding the psychoeducational evaluation process as a whole (e.g., observation, scoring, report writing, conferences, etc.), test administration is the most time-consuming component (Lichtenstein, 1998). In order to be viable, a short form should reduce testing time by approximately 50% (Levy, 1968). In terms of selected subtest short forms, there can be little doubt that some subtests require more time to set-up, administer, and score as compared to others. Given that the majority of past short form investigations routinely extracted short form subtest data from fully administered tests, determining the amount of time saved has typically been based on estimates.

Research has been conducted which provides the individual administration times of each subtest on the WAIS-R with clinical samples (Ryan & Rosenberg, 1984; Ward, Selby, & Clark, 1987). This information has typically been applied in subsequent investigations in order to provide an objective means for estimating the amount of time saved by using a particular combination of subtests. With respect to children, Terminie (1998) recently documented the average administration time of each WISC-III subtest based on research conducted with 27 potentially gifted students referred for assessment. Terminie reported the time required for each subtest beginning with the initial set-up (if required) and ending after a ceiling had been reached. The time in between subtests was not included, nor was the time required to score each subtest (refer to Appendix A for the administration time means and standard deviations for each WISC-III subtest). It was found that while the Information and Coding subtests were the quickest to administer,

subtests that require set-up (e.g., Block Design and Object Assembly) demanded significantly more time.

It has only been in the last several years that researchers have begun to administer short forms as separate tests. Thompson et al. (1986) investigated precisely how much time was required to administer a WAIS-R selected subtest short form with a clinical sample. They discovered that while a two subtest short form (e.g., Vocabulary and Block Design) required approximately 19 minutes to administer, a four subtest short form (e.g., Vocabulary, Arithmetic, Block Design, and Picture Arrangement) took approximately 35 minutes. Given that the administration of the full test battery required 77 minutes in this study, it was confirmed that the use of a short form could reduce testing time by approximately one half.

It appears that a direct comparison of the amount of time needed to administer a selected item short form as opposed to a selected subtest short form has yet to be conducted. Nevertheless, Watkins (1986) speculated that the selected subtest short form would result in a greater saving of time. He based this conjecture on the premise that the selected item approach is more labour intensive for the following reasons: instructions must be read for every subtest; more overall test items must be administered; and all raw scores must be multiplied by a particular number to yield SFIQ scores. It was concluded that while the selected item short form theoretically cuts the test in half, there would be little savings in terms of examiner time and effort. In order for the issue of time to be adequately addressed, however, research is needed to compare the administration and scoring times of selected subtest short forms with that of selected item short forms.

Finally, there is some evidence to suggest that the administration time of the Wechsler scales may be influenced by the intelligence level of the subjects (Kaufman, 1990; Thompson et al. 1986; Satz, Van de Riet, & Mogel, 1967; Ward et al, 1987). For example, in a study of WAIS-R subtest times that was based on a clinical sample that had a mean FSIQ of 77 (Ward et al., 1987), shorter administration times were reported as compared to a similar study where the sample had a mean FSIQ of 90 (Ryan & Rosenberg, 1984).

While the amount of time saved from the use of a short form of an intelligence test is an important consideration, one cannot disagree with Doppelt (1956) when he stated that “a compromise must be made between economy of time and effort and accuracy of prediction (p.63). Clearly, it is the school psychologist’s responsibility to ensure that a short form is an appropriate measure to employ given the particular referral question.

Evaluating the Validity of Short Forms

Validity research related to short forms is necessary in order to determine the extent to which the estimated SFIQ and the FSIQ are similar (Sattler, 2001). McNemar (1950) provided a formula for determining the validity of all possible short forms that required only the subtest intercorrelations reported in the various test manuals. It was later discovered that the McNemar formula yielded artificially high validity coefficients when the short form information was obtained from the same administration of a test as the full form data (Levy, 1968; Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). According to Levy (1967), this was due to the inclusion of the error variance of the part score in both the SFIQ and the FSIQ. In essence, the short form was not only correlated

with unused items of the long form, but also with itself. Formulas were subsequently introduced for correcting the correlations that were contaminated due to the presence of the abbreviated form within the full test (Bashaw & Anderson, 1967; Levy, 1967; Tellegen & Briggs, 1967).

Although there is a significant amount of correlational data which suggests that short forms perform a reasonable job in estimating the FSIQ score (Silverstein, 1989b), there is some question as to whether high correlations between the short form and the full form are sufficient in and of themselves. Resnick and Entin (1971) proposed three validation criteria which have been used by many investigators to evaluate short form intelligence measures. First among the Resnick-Entin criteria is that the correlation between the SFIQ and the FSIQ must be highly significant and account for a substantial percentage of the variance shared by the two measures. Acceptable short form correlations are defined as those that exceed .80 (Kline, 2001). The second criterion holds that the difference between the mean SFIQ and the mean FSIQ must be small and statistically non-significant. The last criterion requires that there must be a high correspondence between the SFIQ classification and the FSIQ classification.

A great deal of debate exists regarding the utility of the Resnick-Entin criteria (Bersoff, 1971; Sattler, 2001; Silverstein, 1985; Thompson, 1987; Thompson et al., 1987; Watkins, 1986). It is almost certain that there will be high correlations between SFIQ scores and FSIQ scores. Therefore, the first criterion is generally always met. With respect to the second criterion, a significant difference between the full battery mean and the short form mean could occur with a sufficiently large sample that serves to emphasize trivial differences. Lastly, due to the very nature of IQ classifications the third criterion

is almost never met. For example, Kaufman (1990) noted that a difference of one IQ point may change a classification of Average to Low Average (e.g., 90 vs. 89), while a significantly larger difference in points may result in no change in classification (e.g., 90 vs. 109). Regardless of how high the correlation is between a short form and the full form, the percentage of classification disagreement will likely be high. As the misclassification criterion cannot account for this type of discrepancy, its use as a validity indicator is questionable (Bersoff, 1971; Kaufman, 1990; Silverstein, 1990a).

Silverstein (1985a) concluded that rather than relying solely on the Resnick-Entin validation criteria, a more profitable approach would be for examiners to base the appropriateness of a short form on a variety of other factors (e.g., reliability and validity of the short form, purpose of the test, examinee characteristics, time constraints, etc.). Thompson (1987) added that a more useful validation criterion would be for SFIQ scores to fall within the confidence interval of the FSIQ for a substantial proportion of the subjects. With respect to the WISC-III, Kramer (1993) devised a confidence interval scale for each of the individual subtests. By referring to this scale, the school psychologist may compare the SFIQ estimate to the confidence interval for the FSIQ to determine the student's classification. In actual practice, the use of the confidence interval for short forms of the WISC-III has been found to significantly reduce misclassifications of students (Beal, Dumont, Cruse, & Branche, 1996; Dumont & Faro, (1993). Cravens (1999) evaluated the validity of a short form by determining the frequency with which it underestimated the FSIQ score. By summing the number of cases where the FSIQ score was underestimated by more than six points by the SFIQ

score, more meaningful information was rendered with respect to the accuracy of the short form.

Reliability of Short Forms

In addition to validity, it is also important to consider reliability when choosing a short form (Cyr & Brooker, 1984; Kline, 2001). Reliability tends to be higher in longer tests as opposed to shorter ones (Enburg, Rowley, & Stone, 1961; Maxwell, 1957; Sattler, 2001; Silverstein, 1967a). Nevertheless, Piotrowski (1976) found that reducing one to five subtests on the WISC-R seemed to have a negligible effect on reliability. Moreover, Silverstein (1990b) discovered that selected subtest short forms of the Wechsler scales had greater reliability as compared to selected item short forms.

Although Tellegen and Briggs (1967) developed a formula for determining the reliability of a short form, relatively few of the early investigators included the concept of reliability in their research (Kaufman, 1972, 1976; Silverstein, 1982, 1985b). To have confidence in the reliability of a short form, Kline (2001) recommended that the reliability coefficients of these instruments should be greater than .90. In addition to evaluating the worth of the short form, determining the reliability is also necessary in order to calculate the standard error of measurement so that confidence intervals may be constructed around the derived SFIQ score. As the reliability of a test decreases, the confidence interval around the observed score increases, thus reflecting the increased error. As a rule of thumb, Kline (2001) recommended that a confidence interval of at least ± 10 points should be used with two subtest short forms, while ± 8 points should be used with four subtest short forms.

Methodological Issues in Short Form Research

With respect to short form research, there are a number of methodological issues that must be considered. These issues include the following: development of short forms based on standardization data, deriving short form IQ estimates, short form data obtained from rescoring protocols, and short form subtest scatter.

Development of Short Forms Based on Standardization Data

McNemar (1950) maintained that in order for short forms to be valid, they should be based on samples that are neither too homogeneous nor too heterogeneous. In the past, however, considerable effort has been directed towards creating the best short form for a specific clinical population (e.g., psychiatric inpatients, cognitively delayed, etc.). The problem with this practice is that short forms developed for one population (i.e., with a certain clinical label, spanning a certain age range, in a certain geographic location) do not tend to generalize well to other populations (Kaufman, 1990; Silverstein, 1990a). To ameliorate this problem, current research has endeavored to utilize short forms that have been derived from the standardization data of the Wechsler scales (Kaufman, 1976; Kaufman et al., 1996; Silverstein, 1982; Sattler, 2001). As these short forms are based on a “normal” population they may yield scores that vary for atypical populations (e.g., cognitively delayed, gifted). Hence, cross validation research must be conducted in order to determine the predictive power of short form instruments with specific populations (Beck, Ray, Seidenberg, Young, & Gamache, 1983; Vollmerhausen, Elder, & Clark, 1986).

Deriving Short Form Intelligence Quotients

Following the administration of a short form, it is necessary to convert the scaled scores into a SFIQ, which is an estimate of FSIQ. In order to obtain IQ estimates, early

short form research relied on prorating scores and then referring to the test manual tables to arrive at a FSIQ (Rabin, 1943). Doppelt (1956) proposed a statistically driven method whereby subtest scores were used in a regression equation to estimate FSIQ. Variations in this regression procedure have been applied in other studies with different Wechsler scales with good results (Clarizio & Veres, 1984; Enburg et al., 1961; Finley & Thompson, 1958; Reynolds et al., 1983; Schwartz & Levitt, 1960; Kennedy & Elder, 1982; Vollmerhausen et al., 1986; Zimet, Farley, & Dahlem, 1985).

According to Tellegen and Briggs (1967), proration is based on the false assumption that the child would have obtained an average score on the subtests that were not administered equal to their average score on those that were administered. As a consequence, proration tends to inflate the standard deviation. Tellegen and Briggs added that regression equations might yield either an underestimate or an overestimate of FSIQ, in addition to shrinking the standard deviation. As an alternative, Tellegen and Briggs devised a linear equating technique whereby a short form score may be transformed into a short form deviation quotient (SFIQ). A SFIQ, like an IQ on the Wechsler scales, has a mean of 100 and a standard deviation of 15. While an IQ is derived from summing all the subtests in the original scale, a SFIQ is obtained by summing the scaled scores on selected subtests. To reduce the need for computation on the part of the examiner, SFIQ tables have been published for various Wechsler scales (Brooker & Cyr, 1986; Kaufman, 1972, 1976; Silverstein, 1982; 1990a; Tellegen & Briggs, 1967).

Despite the apparent utility of the linear equating technique, it also appears to have some flaws. For instance, it is not possible to calculate the standard error of

estimate with the conventional formula when a SFIQ is calculated using the linear equating method (Kaufman, 1977). To resolve this problem, Silverstein (1984, 1985b) developed a procedure for determining the standard error of estimate for SFIQ scores. Linear equating has also been observed to result in an increase in the standard error of estimate of SFIQ scores, however, this increase appears to be only marginal (Silverstein, 1984).

Although several studies have endeavored to examine the relative value of obtaining SFIQ scores via linear equating compared to regression analysis, the results remain unclear. Dumont and Faro (1993) performed both procedures with a sample of children with learning disabilities. It was concluded that with FSIQ scores below 90 regression equations yielded SFIQ scores that were closer to the actual score, whereas with FSIQ scores above 110 the linear equating technique resulted in SFIQ scores that were closer. Nevertheless, the researchers noted that in both cases the differences were within a single point. It was concluded that either the linear equating or the regression technique might be used with equal accuracy for populations with learning disabilities. Based on a sample of special education students, Vollmerhausen and her colleagues (1983) tentatively concluded that multiple regression was marginally better in estimating FSIQ for students with below average ability, while linear equating seemed to be slightly better for students with higher ability. Albeit not statistically significant, Beck et al. (1983) and Donders (1992) each found regression equations to provide a slightly more accurate estimation of FSIQ based on a sample of children referred to a community mental health facility and a sample of children with traumatic brain injury, respectively.

Clearly, there does not appear to be a definitive answer with respect to which method is superior for deriving short form estimates of FSIQ.

Rescoring Protocols to Obtain Short Form Data

The common practice of taking fully administered test protocols and rescoring them to extract the short form information, poses a number of fundamental problems (Doppelt, 1956; Mogel & Satz, 1963; Silverstein, 1968; Yudin, 1966; Zytowski & Hudson, 1965). Although early investigators used the McNemar (1950) formula to determine the validity of a short form which was derived from the administration of the full form, it was later discovered that this practice resulted in artificially high validity coefficients (Zytowski & Hudson, 1965). While correction formulas were devised (Bashaw & Anderson, 1967; Levy, 1967; Tellegen & Briggs, 1967) to account for this type of rescoring contamination, a controversy arose between Silverstein (1970, 1971, 1975) and McNemar (1974, 1975) regarding whether the original formula needed to be corrected. In an attempt to resolve this dispute, Kaufman (1977) concluded that depending upon the circumstance, both the original and the corrected formulae may be sound. For example, the McNemar formula appears most useful in screening situations where the full battery might later be administered. On the other hand, the corrected formula is most useful when the short form is meant to replace the full battery.

Administering both the short form and the full form to the same sample over different occasions could circumvent the correlated error variance problem, however, this could also lead to complications related to retesting effects (Wechsler, 1991). King and King (1982) reasoned that the practical value of such an exercise would be slight,

especially if the short form is being used for screening purposes and the remaining subtests could potentially be administered later.

The practice of rescoring protocols presents another problem in that it fails to take into account the effect that variables such as practice and fatigue might have on a student's test performance (Bersoff, 1971; Sattler, 2001; Satz & Mogel, 1962; Thompson, 1987; Thompson et al., 1986). It is also possible that due to alterations in the continuity of item difficulty, coupled with the change in the test length, subjects might perform differently on a separately administered short form as compared to the full form administration (Bersoff, 1971; Sattler, 2001; Silverstein, 1990a).

In order to explore whether short form assessment resulted in reduced levels of effort, attention, and motivation as compared to full battery assessment, Mogel and Satz (1963) separately administered a WAIS selected item short form to a group of subjects. It was concluded that there was no significant difference between the results derived by rescoring and those obtained by separately administering the short form. Finch, Childress, and Ollendick (1973) found similar results when investigating a selected item short form of the WISC. Thompson and his colleagues (1986) evaluated a variety of administrative conditions in order to assess the validity of a WAIS-R selected subtest short form. They compared the following assessment situations with a clinical sample: (a) administering a two subtest short form (Vocabulary-Block Design) followed by the remaining subtests in standard order, (b) administering a four subtest short form (Arithmetic, Vocabulary, Block Design, Picture Arrangement) followed by the remaining subtests in standard order, and (c) administering the full test in standard order. Thompson and his colleagues discovered that when the two subtest short form was

administered first, it produced a SFIQ estimate that significantly overestimated the actual FSIQ. This finding indicated that subjects might be more motivated to perform better at the outset of testing as they are less tired and more attentive. On the other hand, the four SFIQ scores were very similar to the FSIQ scores. It was concluded that whether a four subtest short form was given first or embedded in the total battery in the standard order, the magnitude of the estimated IQ was not affected. Consequently, selected subtest short forms that are comprised of four or more subtests do not appear to be negatively impacted by a change to the standardization order of the test.

With respect to gifted populations, Linn and Lopatin (1990) used a four subtest short form to explore whether there was a difference between the results obtained from rescored test protocols with those from the separately administered short form. It was concluded that altering the WISC-R subtest order did not prejudice the outcomes when assessing gifted students. Similarly, Kramer et al. (1983) found no evidence for the administration format effecting the estimated SFIQ scores of a WISC-R five subtest short form with potentially gifted students.

Short Form Subtest Scatter

Subtest scatter refers to the difference between the examinee's highest and lowest subtest scaled scores. The presence of subtest scatter on a short form may be indicative of cognitive strengths or deficits that warrant further assessment in order to obtain a more complete evaluation of the individual's functioning. The issue of subtest scatter is particularly important when conducting short form research with gifted populations. Examination of subtest scatter on full administrations of the WISC-R (Patchett & Stansfield, 1992) and the WISC-III (Fishkin, Kampsnider, & Pack, 1996; Silverman,

1998; Sparrow & Gurland, 1998) suggested that gifted students might demonstrate more subtest scatter than typical students. Hence, reliance on a single aggregate score (i.e., an average of a variety of abilities) may result in an underestimate of a gifted student's potential.

In order to have confidence that an estimated SFIQ is representative of the student's overall ability, the subtest scores should be examined for subtest scatter immediately following the administration of the short form. Kaufman (1977) recommended that "when the child's short form subtest profile exhibits scatter or is otherwise clinically interesting" (p. 1160), the remaining subtests in the intelligence test battery should be administered in order to obtain a clearer representation of the student's ability. Based on Kaufman's recommendation, LoBello (1991) suggested that subtest scatter of more than 6 points (i.e., greater than two standard deviations on the subtest scale) should be viewed as a "flag" by the school psychologist for determining when the complete test should be administered. More conservative short form research has advised the administration of the remaining subtests when the range of scatter consists of 4 to 5 scaled points (Donders & Warschausky, 1996). It is important to acknowledge, however, that the absence of subtest scatter does not preclude such scatter if the remaining subtests were administered (LoBello, 1991; Thompson, 1992).

Summary of Short Form Research Issues

The primary purpose of short form research is to develop an intellectual assessment tool that is quick to administer. Based on an extensive review of the literature, Silverstein (1990a) concluded that the best short form research contains the following elements: uses a selected subtest short form; investigates a short form that has

been derived from the standardization sample; applies a linear equating technique to derive SFIQ scores; and reports as much information as possible about the psychometric properties and the clinical features of the short form. Moreover, it was suggested that future short form research should focus on validating previously developed short forms with new populations, as little can be gained from adding to the plethora of abbreviations already available (Bersoff, 1971; Silverstein, 1990a). From a methodological perspective, short forms comprised of four or more subtests do not appear to be negatively influenced by changes in the standard administration format. Finally, it is important to be aware that the presence of subtest scatter within the short form may result in a SFIQ score that fails to accurately reflect the student's overall ability.

Short Form Research with the Wechsler-Bellevue Intelligence Scales

Following the 1939 publication of David Wechsler's first intelligence test, the Wechsler-Bellevue Intelligence Scales, Rabin (1943) investigated the utility of a selected subtest short form in order to save administration time. Based on clinical impression of how well individual subtests correlated with the FSIQ, coupled with ease of administration, a short form consisting of Comprehension, Arithmetic, and Similarities was chosen. Although the research data was derived from rescoring existing test protocols from a sample of student nurses and a sample of psychiatric patients, it was estimated that this short form saved approximately 70% of the time required for test administration. SFIQ scores were then estimated based on the prorating method. In order to evaluate the short form, correlations were compared with the original scale and with the Army Alpha test. It was found that the difference between the mean IQ on the

short form and that on the original scale was not statistically significant. Furthermore, intellectual classifications remained the same in the majority of cases.

Although Rabin deemed the short form of the Wechsler-Bellevue to be useful in terms of speed of administration and accuracy of results, the following limitations were acknowledged: the short form subtests were only representative of the Verbal scale; there was less opportunity to observe the examinee in a wide variety of cognitive situations; and the results were not verified with other populations of subjects. Nevertheless, this initial investigation of a selected subtest short form was the catalyst for a variety of other Wechsler-Bellevue research studies (Cummings, MacPhee, & Wright, 1946; Geil, 1945; Hilden, Taylor, & Dubois, 1952; Kriegman & Hansen, 1947). Moreover, it established the foundation for short form research with future versions of Wechsler intelligence tests.

Short Form Research with the Wechsler Intelligence Scale for Adults (WAIS) and the Wechsler Intelligence Scale for Adults-Revised (WAIS-R)

Following the revision of the Wechsler-Bellevue Intelligence Scale in 1955, the test instrument was renamed the Wechsler Intelligence Scale for Adults (WAIS). Although there has been a considerable amount of short form research involving both the WAIS and the WAIS-R, these studies have typically utilized populations consisting of psychiatric inpatients or individuals with cognitive delays. There do not appear to be any investigations that examine the use of short forms with adults who have superior intellect.

In order to save administration time, Doppelt (1956) developed a selected subtest short form of the WAIS. This short form included an equal number of Verbal and Performance subtests that had the highest correlations with their respective scales in the standardization sample. The resultant short form was comprised of Arithmetic,

Vocabulary, Block Design, and Picture Arrangement. Regression equations were utilized in order to obtain SFIQ scores. Doppelt reported that SFIQ scores correlated highly (.95) with FSIQ scores.

Based on the WAIS standardization data, Silverstein (1967) discovered that a short form consisting of Vocabulary and Block Design correlated highly (.92) with the FSIQ. This dyad was composed of one Verbal and one Performance subtest, each of which were reported to be the most reliable and best measure of “g” in their respective scales. After the release of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) in 1981, Silverstein (1982) determined that the Vocabulary-Block Design dyad continued to correlate the highest (.91) with the FSIQ. The addition of one Verbal (Arithmetic) and one Performance (Picture Arrangement) subtest to the dyad created an equally balanced Verbal and Performance tetrad. This short form, which was the same abbreviation Doppelt (1956) used in his WAIS research, was found to be more reliable (.94) and more valid (.95) than the Vocabulary-Block Design dyad. Although other tetrads were observed to correlate more highly with the FSIQ, this short form was selected for its clinical, practical, and empirical value. Silverstein (1982) estimated that while the dyad could be administered in approximately 19 minutes, the tetrad required approximately 35 minutes. Cross validation of this dyad and tetrad with clinical populations (e.g., low functioning adults, dementia patients) demonstrated that both short forms correlate highly (.86 and above) with the FSIQ (Haynes, 1985; Margolis, Taylor, & Greenlief, 1986). Although the tetrad was found to be more valid than the dyad, it was concluded that both abbreviations could be used for screening purposes.

Reynolds and his colleagues (1983) created a tetrad based on the WAIS-R standardization data which consisted of Information, Arithmetic, Picture Completion, and Block Design. These subtests were selected for the following reasons: they correlated highly with their respective scales, they measured a variety of mental processes, they were quick to administer and score, and they formed a clinically interesting unit. This short form, which reported an overall correlation coefficient of .87, was estimated to require between 20-30 minutes to administer. Reynolds et al. reasoned that while the Verbal dyad consisting of Vocabulary-Arithmetic was superior to Information-Arithmetic, it also required more time to administer and score. Moreover, while the Performance dyad of Picture Completion-Block Design was equally as strong as Picture Arrangement-Block Design, the Picture Completion subtest required less time to administer than Picture Arrangement. It was also noted that Picture Completion was not reliant upon motor skill, thus making it a good complement to Block Design. Overall, this short form appeared to be an effective and practical screening device that was just as reliable and valid as the tetrad utilized by Silverstein (1982), especially when time was a consideration.

Kaufman (1990) reported that there were certain subtests on the Wechsler scales that took longer to administer and score (e.g., Vocabulary, Comprehension, Block Design, Picture Arrangement, Object Assembly), while other subtests required less time (e.g., Information, Digit Span, Similarities, Arithmetic, Picture Completion, Digit Symbol). With the primary object of creating a short form that was quick to administer and score, Kaufman and his colleagues (Kaufman & Ishikuma, 1989; Kaufman et al., 1991) set out to develop the most efficient and accurate dyad, tetrad, and pentad based on

the WAIS-R standardization sample. The dyad, consisting of Information and Picture Completion, was chosen because it included the first two subtests given in the WAIS-R. As this dyad maintained the proper order of administration, it eliminated the possibility that motivation and attention factors might have impacted the short form results. The Information-Picture Completion short form was found to have a correlation coefficient of .88, which was almost as good as Silverstein's (1982) Vocabulary-Block Design (.94) dyad. Moreover, this dyad required less time to give (12 minutes), thus saving approximately 85% of administration time.

Kaufman and his colleagues selected a short form triad that included the first three subtests of the WAIS-R in their designated order (Information, Picture Completion, Digit Span). The advantage of this triad was that it represented all three WAIS-R factors (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility). It was found to have a correlation coefficient of .92 and required 16 minutes to administer, which constituted a saving of 80% of administration time. Owing to the finding that subtest order did not appear to be a factor in tetrad short forms (Thompson, 1987; Thompson et al., 1986), an abbreviation was proposed which included Similarities, Arithmetic, Picture Completion, and Digit Symbol. It was estimated that this short form would require approximately 19 minutes to administer, thus resulting in a saving of 75% of administration time. This short form was found to correlate highly (.95) with the FSIQ. The researchers reported that while this short form required one-third less time to administer than the Reynolds et al. (1983) tetrad and one-half less time than the tetrad utilized by Silverstein (1982), it still compared favorably to these measures in terms of psychometric properties.

An overview of short form research involving the WAIS and the WAIS-R adds considerably to our understanding of the development and application of abbreviated assessment instruments. Furthermore, given that many short forms of intelligence tests used with children are derived from adult abbreviations, it is beneficial to be familiar with all short form research.

Short Form Research with the Wechsler Intelligence Scale for Children (WISC)

The original version of the Wechsler Intelligence Scale for Children (WISC) was published in 1949. A review of the literature revealed that a significant amount of short form research was conducted with the WISC. These studies typically explored the use of abbreviations with students who were cognitively delayed (Carleton & Stacey, 1954; Cole, Williams, Nix & Litaker, 1967; Finch, Ollendick, & Ginn, 1973; Finley & Thompson, 1958; Reid, Moore, & Alexander, 1968; Schwartz & Levitt, 1960) or students who were emotionally troubled (Enburg et al., 1961; Finch et al., 1974; Simpson & Bridges, 1959; Yalowitz & Armstrong, 1955). In these earlier studies, short forms appear to have been chosen based on empirical grounds (i.e., the degree to which the SFIQ correlated with the FSIQ), as opposed to clinical value or psychometric soundness.

Thompson and Finley (1963) were the only investigators to explore the utility of a short form of the WISC with a sample of gifted students. The test protocols of 400 gifted elementary students with FSIQ scores of 125 and above were used to develop a selected subtest short form derived by using regression analysis. An additional 151 gifted students served as the cross validation sample. The resultant short form, referred to as the California Abbreviated WISC for the Intellectually Gifted (CAW-IG), was comprised of the following subtests: Information, Similarities, Picture Completion, Picture

Arrangement, and Block Design. A correlation coefficient of .75 was found between the short form and the full form. The researchers reported that this short form estimated the FSIQ within 4 points in 74% of the cases and within 8 points in 97% of the cases. Nearly identical findings were obtained when the short form was used with the cross validation sample. It was concluded that while this short form had difficulty distinguishing the relative standing of gifted students among themselves, it seemed to be a good predictor of who was gifted and who was not gifted.

Short Form Research with the Wechsler Intelligence Scale for Children- Revised (WISC-R)

The Wechsler Intelligence Scale for Children-Revised (WISC-R) was published in 1974. Owing to the high degree of correlation between the WISC-R and the WISC-III (Sevier et al., 1994; Wechsler, 1991), studies evaluating abbreviated versions of the WISC-R may be useful in determining the utility of WISC-III short forms.

WISC-R Short Form Research with Non-Gifted Populations

Kennedy and Elder (1982) used multiple regression analysis to develop a selected subtest short form of the WISC-R. The WISC-R protocols of 400 students with academic delays (mean FSIQ of 80) were rescored according to a statistically derived regression equation. It was reported that the best pentad, comprised of Information, Block Design, Comprehension, Picture Arrangement, and Coding, correlated highly (.95) with the FSIQ. Kennedy and Elder concluded that this short form was useful for re-evaluation purposes.

Several investigations have been conducted to validate the Kennedy-Elder short form with other samples. Zimet and her colleagues (1985) rescored the WISC-R protocols of 100 psychiatric inpatients between the ages of 6 to 16. They reported the

following findings: the short form correlate highly (.96) with the full form, no significant differences were found between the means of the short form and the long form, and only one child (1%) was misclassified by the short form. Consequently, this study had the distinction of being one of the few to meet all three of the Resnick-Entin (1971) validation criteria. In an attempt to replicate these impressive findings, Phelps and Rosso (1986) extracted the Kennedy-Elder short form data from the WISC-R protocols of 120 adolescents (aged 13 through 16) with emotional and behavioural difficulties. Although the SFIQ was found to correlate highly (.94) with the FSIQ, the short form tended to underestimate the FSIQ for this particular sample and 25% of the cases were misclassified.

In order to evaluate whether the method used to construct the short form yielded any difference in accuracy for estimating FSIQ, some studies have compared WISC-R short forms derived from rational and empirical criteria (e.g., Kaufman, 1976) with short forms developed solely from statistical data (e.g., Kennedy & Elder, 1982). Beck et al. (1983) and Donders (1992) each found that although the two abbreviations were almost equal in their ability to estimate FSIQ scores, the Kennedy-Elder short form (Information, Block Design, Comprehension, Picture Arrangement, Coding) was slightly better than the Kaufman¹ short form (Arithmetic, Vocabulary, Picture Arrangement, Block Design).

In another comparison study, Vollmerhausen et al. (1986) rescored the WISC-R protocols of 192 psychiatric inpatients and 200 special education students. The investigators reported that the Kennedy-Elder short form appeared to be a better predictor

¹ The Kaufman (1976) WISC-R short form is comprised of the same subtests Doppelt (1956) used in his WAIS short form. Consequently, this tetrad is referred to as the "Doppelt short form" in other areas of this dissertation.

for students with below average ability, while the Kaufman short form seemed to be better for students with higher ability. Zimet and Adler (1990) reported similar findings and concluded that the choice of which abbreviation to use was dependent upon the general intelligence level of the population studied. With respect to the use of short forms for re-evaluations, it was recommended that the remaining subtests should be administered in any situation where the short form resulted in an IQ classification change or if a low score was obtained (Vollmerhausen et al., 1986; Zimet & Adler, 1990).

WISC-R Short Form Research with Gifted Populations

It appears that a number of researchers have attempted to develop WISC-R short forms for the purpose of identifying potentially gifted students. In order to evaluate the effectiveness of different short forms in estimating FSIQ, Dirks et al. (1980) rescored the WISC-R protocols of 47 potentially gifted Grade 4 students according to a variety of short form combinations. Selected subtest short forms were constructed based on how well each subtest correlated with the FSIQ. The best dyad for this gifted sample consisted of Similarities and Block Design (.82), while the best triad included Similarities, Vocabulary, and Block Design (.85). Nevertheless, it was suggested that the dyad comprised of Similarities and Object Assembly (.77) and the triad made up of Similarities, Vocabulary, and Object Assembly (.72) represented more clinically interesting short forms. The authors concluded that for this sample of potentially gifted students, all of the reported dyads or triads were effective in estimating FSIQ and could be useful as screening measures.

Other investigators have also studied the Dirks et al. (1980) short form consisting of Similarities, Vocabulary, and Block Design, for use with gifted students. Fineman and

Carran (1986) rescored the WISC-R protocols of 200 potentially gifted students in order to extract the relevant short form data. It was discovered that the Similarities-Vocabulary-Block Design triad was more effective than the Vocabulary-Block Design dyad in accurately screening potentially gifted students.

Although the Vocabulary-Block Design dyad has generally been reported to result in high misclassification rates with clinical populations, its use as a screening instrument to determine whether additional assessment is warranted has received a significant amount of attention. For example, Mumpower (1955), Simpson and Bridges (1959), and Silverstein (1967a) found this dyad to correlate highly (.95, .87, and .87, respectively) with the FSIQ on the WISC. Silverstein (1970) reported a .86 correlation between the SFIQ and the FSIQ on the WISC-R, while Sattler (1974) obtained a .91 correlation. Haynes (1982), Haynes and Bensch (1982), and Ryan (1981) reported high correlations (.82, .87, and .85, respectively) and endorsed this dyad as a screening instrument for students with behaviour problems. With respect to gifted populations, Killan and Hughes (1978) and Elman et al. (1981) reported that the Vocabulary-Block Design dyad correlated highly (.92 and .81, respectively) with the FSIQ on the WISC-R. Both studies concluded that this dyad could be used effectively as a screen for identifying gifted students.

Karnes and Brown (1981) also conducted an investigation of the Vocabulary-Block Design dyad on the WISC-R. Utilizing a significantly larger sample consisting of 946 gifted students between the ages of 6 to 16, protocols were rescored to obtain the short form data. Unlike the impressive results found in earlier studies, this dyad was observed to correlate only modestly (.62) with the FSIQ. Nevertheless, it was still

reported to be the best dyad for this sample. Karnes and Brown went on to devise a short form for gifted students based on the subtests that correlated most highly with the FSIQ in this study. They concluded that in terms of accuracy and time, the two best tetrads consisted of Similarities, Vocabulary, Block Design, and Object Assembly (.73) and Similarities, Vocabulary, Picture Arrangement, and Block Design (.72).

In an effort to accurately screen potentially gifted minority students, Ortiz and Gonzalez (1989) utilized a short form of the WISC-R that Karnes and Brown (1981) reported as being one of the most effective in their sample of gifted students (Similarities, Vocabulary, Picture Arrangement, Block Design). A sample of 96 potentially gifted Hispanic students, in Grades 4 through 8, were each administered the short form separately. After several months, these students were administered the complete WISC-R battery. It was found that the Verbal dyad correlated highly (.71) with Verbal IQ, while the Performance dyad correlated to a lesser degree (.55) with Performance IQ. The researchers reported that this short form of the WISC-R was a valid instrument for screening potentially gifted minority students.

In order to identify which WISC-R dyad was the most useful for screening potentially gifted students, Fell and Schmidt-Fell (1982) analyzed the full test batteries of 92 gifted students. By extracting out the relevant short form data, it was found that the Similarities-Vocabulary dyad and the Similarities-Object Assembly dyad seemed to yield the best estimates of FSIQ for this sample. It was reported that the use of either of these short forms resulted in 62% accuracy in estimating the FSIQ, as opposed to the more popular Vocabulary-Block Design dyad that was found to have a 53% rate of accuracy.

Based on the success of the CAW-IG abbreviation of the WISC (Thompson & Finley, 1963), Kramer and his colleagues (1983) investigated the utility of this short form with the WISC-R. From a total of 73 students referred for gifted assessment, 18 were administered the short form (Information, Similarities, Picture Completion, Picture Arrangement, Block Design) separately, while 55 were administered the WISC-R in the standard manner and later rescored to extract the relevant short form data. It was determined that despite the different administration procedures, the SFIQ estimates from both groups correlated highly (.91 and .84, respectively) with the FSIQ. The researchers concluded that this short form could predict program eligibility without significantly overestimating FSIQ.

Linn and Lopatin (1990) developed a simultaneous screening and assessment procedure for identifying gifted students for special programming. They proposed that during the screening phase, a short form of the WISC-R could be administered separately to all gifted referrals. Examinees reaching or exceeding a designated cutoff score would immediately be given the remaining subtests to assess program eligibility, while the assessment would be terminated for those students who failed to reach this score. The WISC-R short form utilized in the Linn and Lopatin study (1990) consisted of Arithmetic, Vocabulary, Picture Arrangement, and Block Design. Although Doppelt (1956) first proposed this abbreviation for the WAIS, it was later adapted for use on the WAIS-R (Silverstein, 1982) and the WISC-R (Beck et al., 1983; Donders, 1992; Kaufman, 1976). This particular tetrad has enjoyed considerable popularity due to the following reasons: it contains an equal number of Verbal and Performance subtests; it represents all three WISC-R factors; and it is psychometrically sound. Compared to other

short forms, this abbreviation has been reported to be especially promising with gifted students (Vollmerhausen et al., 1986; Zimet & Adler, 1990).

The subjects in the Linn and Lopatin study (1990) consisted of 203 students, in Grades 2 through 5, who were referred for gifted assessment. To obtain the short form data, completed WISC-R protocols were rescored. It was found that the Arithmetic, Vocabulary, Picture Arrangement, and Block Design short form correlated highly (.89) with full form. To test for a possible bias on WISC-R scores due to alterations in the test sequence, the short form was administered separately to another 22 students referred for gifted assessment. The remaining subtests were administered immediately following the short form in standard order. The results from this smaller group were reported to be consistent with the original group that was based on the standard subtest administration. Consequently, altering the subtest order did not seem to impact the estimated FSIQ scores in this sample of gifted students. The researchers concluded that using this short form of the WISC-R with gifted students was both efficient and accurate. Moreover, it was estimated that this abbreviation could save approximately one hour of a school psychologist's time per child referred for assessment.

The first short form study utilizing Canadian students (American normative data) sought to compare a WISC-R short form against a Stanford-Binet (SB-IV) short form (Dipasquale et al., 1992). A total of 51 Grade 3 students referred for giftedness were separately administered a WISC-R short form consisting of Information, Similarities, Vocabulary, Picture Arrangement, and Block Design. It was noted that the researchers failed to report their rationale in selecting this particular WISC-R abbreviation. Moreover, it was impossible to calculate how well the short form correlated with the full

form, as the complete test battery was not administered. Nevertheless, Sattler (1992) reported this short form to be one of the top ten tetrads and to correlate highly (.94) with the FSIQ of the standardization sample. It was concluded that the short form derived from the older WISC-R yielded significantly higher SFIQ scores than did the short form from the newly published SB-IV. The researchers pointed out that the inflation of scores on older tests, which is referred to as the “Flynn effect” (Flynn, 1984, 1987), seemed to be especially apparent in students functioning above the average range.

Short Form Research with the Wechsler Intelligence Scale for Children- Third Edition (WISC-III)

The Wechsler Intelligence Scale for Children-Third Edition (WISC-III) was introduced in 1991. The WISC-III was normed in Canada in 1996.

WISC-III Short Form Research with Non-Gifted Populations

Utilizing the standardization data, Sattler (2001) developed the 10 most valid two, three, four, and five subtest abbreviations of the WISC-III. The Vocabulary-Block Design dyad continues to be one of the most popular short forms because the subtests have excellent reliabilities, they correlate highly with the FSIQ, and they provide good measures of “g”. It is important to note that although this dyad has a correlation coefficient of .91, the reliability of a short form in estimating the FSIQ increases as the number of subtests used in the short form also increases (Enburg et al., 1961; Maxwell, 1957; Sattler, 2001; Silverstein, 1967a).

The effectiveness of the Vocabulary-Block Design dyad of the WISC-III was evaluated with a sample of students who demonstrated either academic or behaviour problems (Herrera-Graf, Dipert, & Hinton, 1996). The WISC-III protocols of 197

subjects were rescored in order to derive the short form information. This dyad was found to correlate highly (.88) with the FSIQ for the subjects in this sample. Nevertheless, significant differences were found between the short form and long form means. Furthermore, approximately half of the sample showed changes with respect to their intelligence classification level. It was uncertain as to whether the heterogeneity of the FSIQ scores in this sample (ranging between 40-135) may have been responsible for these problems. The investigators concluded that this abbreviation should not be used, even as a screening instrument, with special education students.

As the Vocabulary-Block Design dyad is limited in its ability to provide clinical and diagnostic information, Sattler (1992) suggested that the WISC-III tetrad comprised of Information, Vocabulary, Picture Completion, and Block Design would yield more useful data. Although this particular short form requires more time to administer than the dyad, it correlates more highly (.93) with the FSIQ. Browne (1995) found this abbreviation provided a good estimate of FSIQ in a clinical sample. Furthermore, it yielded a high correlation coefficient (.93) and a low misclassification rate.

Based on the WISC-III standardization data, Sattler (2001) reported that the Information, Vocabulary, Block Design, and Symbol Search short form appeared to be especially promising. He added that the Arithmetic, Vocabulary, Block Design, and Symbol Search short form was also very appealing. It was noted that both of these abbreviations demonstrated high reliabilities (.93 and .92, respectively) and validities (.91 and .90, respectively). At this time, however, neither of these short forms appear to have been validated with a clinical sample.

Utilizing the WISC-III standardization data, the reliability and validity of three short forms were compared (Kaufman et al., 1996). The first short form, comprised of Similarities, Vocabulary, Picture Arrangement, and Block Design, appeared to have the best clinical and psychometric properties. Although this abbreviation correlated highly (.93) with the FSIQ, coupled with the fact that this combination of subtests was exceptional in terms of psychometric and clinical utility, it was estimated to require a significant amount of time to administer (e.g., 37 minutes). The second short form, chosen on the basis of its brevity in administration, included Information, Arithmetic, Picture Completion, and Symbol Search. This short form was estimated to require only 21 minutes to administer and correlated .89 with the FSIQ. A third short form, comprised of Similarities, Arithmetic, Picture Completion, and Block Design, was selected as it represented a compromise between the characteristics of the other two abbreviations. This short form required only 27 minutes to administer and was reported to correlate highly (.93) with the FSIQ. Kaufman and his colleagues concluded that for screening purposes, this short form represented the best combination in terms of psychometric properties, clinical soundness, and practical utility.

Connery, Katz, Kaufman, and Kaufman (1996) utilized the Similarities, Arithmetic, Picture Completion, and Block Design short form identified in the Kaufman et al. (1996) research with a clinical sample. With a group of 30 psychiatric inpatients, aged 12 through 16, the abbreviation was administered separately in order to obtain SFIQ scores. Actual administration of the short form revealed that it required even less time to administer than previously estimated (i.e., 25 minutes as opposed to 27). This short form was reported to correlate highly (.93) with the FSIQ. The researchers concluded that the

Similarities, Arithmetic, Picture Completion, and Block Design tetrad constituted a valid measure for this clinical sample.

Short form research with students who have learning disabilities appears to be rare (Calrizio & Veres, 1984). Nevertheless, recent interest in this population has increased possibly due to the three year re-evaluation requirement that has been instituted in some school districts. Utilizing a sample of students with learning disabilities, Dumont and Faro (1993) devised a short form for use with both the WISC-R and WISC-III. The Dumont-Faro short form is comprised of Information, Vocabulary, Picture Completion, Block Design, and Coding. This combination of subtests was chosen for the following reasons: they correlate highly with their respective Verbal and Performance scales; they measure a variety of mental processes; and they yield clinically useful information. Moreover, this short form was estimated to require 30 minutes to administer and to involve minimal subjective scoring. Based on the WISC-R protocols of 213 students with learning disabilities, aged 6 through 16, regression analysis was conducted to yield a formula that could be used to obtain SFIQ scores. The short form was reported to correlate highly (.90) with the full form. The Dumont-Faro WISC-R short form was then administered separately to a sample of 35 students who had been identified three years earlier as being learning disabled. Although 33% of the students were observed to change intelligence classifications with the short form, 100% were classified correctly when confidence intervals were utilized.

Dumont and Faro (1993) used the same method of regression analysis to develop a WISC-III short form formula based on the subtest scores of a sample of 98 learning disabled students. The short form of the WISC-III was found to correlate highly (.89)

with the full form. With a sample of 105 students that had been identified three years earlier as being learning disabled, half were administered the complete WISC-III and half were administered the short form separately. The correlation coefficient for the WISC-III FSIQ and the WISC-R FSIQ was .83, while the correlation between the WISC-III SFIQ and the WISC-R FSIQ was .81. Although 41% of the subjects were misclassified using the short form, the researchers reported that only 6% were misclassified when confidence intervals were applied. It was cautioned that while the Dumont-Faro short form was a time-efficient testing tool for re-evaluating students with learning disabilities, it should never be used as the initial evaluation of cognitive ability.

The effectiveness of the Dumont-Faro short form as a re-evaluation instrument for students with learning disabilities was investigated using Canadian norms (Beal, Dumont, Cruse, & Branche, 1996). This study was a follow-up to previous research that validated the Dumont-Faro short form with a sample of 44 Canadian students utilizing American norms (Beal, Dumont, Branche, & Cruse, 1996). The SFIQ (derived by rescoring the complete test) and FSIQ scores of 63 Canadian students with learning disabilities were compared based on both the WISC-III American and Canadian norms. Consistent with the Canadian WISC-III validation study (Wechsler, 1996), it was concluded that the Canadian norms more accurately represented the cognitive functioning of Canadian students. With respect to the utility of the Dumont-Faro abbreviation when Canadian norms were utilized, the short form was reported to correlate highly (.90) with the full form. Although misclassifications were observed in approximately 3% of the cases when confidence intervals were utilized, the researchers concluded that this was an acceptable

compromise for the efficiency that the abbreviation could provide when reassessing students with learning disabilities.

Comninel and Bordieri (2001) also examined the accuracy of the Dumont-Faro short form in estimating WISC-III scores for special education students referred for re-evaluation. This study involved rescoring the WISC-III protocols of 45 subjects, aged 6 through 16, who were receiving special education services. Although the short form correlated highly (.96) with the full form, there was a tendency for the SFIQ scores to underestimate actual scores. Moreover, 44% of the students were found to be misclassified when SFIQ scores were compared directly with FSIQ scores. The percentage of misclassifications obtained in the Comninel and Bordieri study was higher than the 41% (Dumont & Faro, 1993), 27% (Beal, Dumont, Cruse, & Branche, 1996), and 11% (Beal, Dumont, Branche, & Cruse, 1996) misclassification rates noted in other Dumont-Faro short form studies. Comninel and Bordieri concluded that the use of this short form with special education students required further investigation. A possible explanation for the limited success of the Dumont-Faro short form in this study might be related to the level of ability of the subjects as the FSIQ scores reportedly ranged between 47 to 131. It has been reported elsewhere that caution should be exercised when using this short form with students who have FSIQ scores less than 79 (Dumont & Faro, 1993).

Donders (1997) proposed a six subtest short form of the WISC-III comprised of Vocabulary, Similarities, Picture Completion, Block Design, Arithmetic, and Coding. Based on the WISC-III standardization data, the following criteria were used to choose subtests for the abbreviation: level of reliability and validity; representation of the four-

factor structure of the complete scale; samples a range of tasks; and brevity for administration. This short form was found to correlate highly (.94) with the full form.

The Donders short form was cross validated on a sample of students with traumatic head injury (Donders & Warschausky, 1996). The WISC-III protocols of 171 individuals, aged 6 through 16, were rescored to derive the short form data. The short form was observed to correlate highly (.90) with the full form. The investigators concluded that the Donders short form was an accurate and valid alternative to the full administration of the WISC-III for the subjects in this sample. They added that while the SFIQ may be used to provide a global estimate of ability, for decision-making purposes the WISC-III should be administered in its entirety.

In Campbell's (1998) investigation of the validity of several short forms of the WISC-III, the Donders short form emerged as being one of the best abbreviations for use with a clinical population. The WISC-III protocols of 212 psychiatric inpatients, with a mean age of 12, were rescored according to a variety of different abbreviations. The Donders short form, which was estimated to require 30-40 minutes to administer, correlated highly (.91) with the full form. This abbreviation also provided a broader breadth of measurement as it sampled all four factors that comprise the WISC-III. It was reported that while the Vocabulary-Block Design dyad demanded considerably less time to administer, it did not correlate as highly (.82) with the FSIQ and it was found to yield larger standard errors. Depending on the needs of the examiner, it was concluded that the Donders sextet and the Vocabulary-Block Design dyad were valid and practical short forms with this inpatient population.

Cravens (1999) adapted several short forms developed for other Wechsler scales for use with the WISC-III. The WISC-III protocols of 137 African American psychiatric inpatients, aged 6 through 16, were rescored to extract the relevant short form data. It was found that a tetrad originally introduced by Doppelt (1956) for the WAIS (Arithmetic, Vocabulary, Picture Arrangement, Block Design) correlated reasonably well (.89) with the FSIQ and correctly classified 65% of the subjects. A tetrad proposed by Kaufman et al. (1991) for the WAIS-R consisting of Similarities, Arithmetic, Picture Completion, and Coding (instead of the Digit Symbol subtest that was used on the WAIS-R) was also investigated. Compared to the Doppelt abbreviation, this short form correlated to a similar degree with the full form (.88) and it classified a near equal percentage (64%) of subjects. Furthermore, it was estimated to require less time to administer. Cravens concluded that the Kaufman et al. tetrad seemed to be a better short form for the subjects in this clinical sample.

As composite factor scores yield distinctive and functionally meaningful ability profiles, they are more reliable than single subtest scores (Wechsler, 1991). Consequently, Prifitera, Weiss, and Saklofske (1998) recommended the use of an eight subtest short form in order to provide a complete measure of the Verbal Comprehension (Information, Similarities, Vocabulary, Comprehension) and the Perceptual Organization (Picture Completion, Picture Arrangement, Block Design, Object Assembly) factors on the WISC-III. The sum of the short form subtests were converted into a General Ability Index (GAI) score, which is similar to the FSIQ. Based on the WISC-III standardization data, GAI computational tables have been developed for use with American (Prifitera et al., 1998) and Canadian norms (Weiss, Saklofske, Prifitera, Chen, & Hildebrand, 1999).

The GAI short form was reported to correlate highly with the full form in both the American (.95) and the Canadian (.94) samples.

Although there does not appear to be any research available with respect to the GAI short form with gifted individuals, Prifitera et al. (1998) suggested that substituting this short form for the FSIQ may result in a greater number of potentially gifted students being identified. This assumption seems to be based on research reported in the Wechsler manual (1991) which suggested that gifted students demonstrated lower scores on the Processing Speed factor of the WISC-III as compared to the other three factors. Nevertheless, given that the GAI abbreviation requires the administration of eight of the 10 standard subtests in the WISC-III battery, employing this short form would result in very little time saved.

WISC-III Short Form Research with Gifted Populations

Linville, Rust, and Kim (1999) investigated the utility of the Information-Picture Completion dyad to determine whether it was an accurate screening device for gifted students. Sattler (1992) recommended this short form due to its speed of administration and ease of scoring. Given that this abbreviation preserves the standard order of subtest administration, the full test may be administered without switching test instruments and without discarding the information already obtained. Linville and his colleagues rescored the WISC-III protocols of 86 potentially gifted students using the Information-Picture Completion dyad. It was concluded that this short form, which reportedly correlated at .72 with the FSIQ, was a reasonable screen for gifted referrals. Of the 58 students who obtained FSIQ scores within the gifted range, the short form correctly identified 50. Had this short form been used initially, 28 complete assessments would not have been needed.

On the other hand, this short form also misclassified eight gifted students who would have been inaccurately screened out of the assessment process. To correct this problem, the researchers recommended that the cutoff score could be lowered to include all gifted students. It must be kept in mind, however, that while a lower cutoff score would have served to include all gifted students, it would also have added a higher percentage of non-gifted students. Despite its apparent merits, Campbell (1998) cautioned that the Information-Picture Completion short form has a tendency to overestimate the FSIQ and that its split/half reliabilities were among the lowest when compared to other dyads. Although this particular dyad requires less time to administer, it fails to yield as much clinical information as the Vocabulary-Block Design dyad (Sattler, 1992). For example, the Vocabulary subtest provides an indication of how a student deals with a less structured task, while the Block Design subtest offers information regarding problem-solving strategies.

Terminie (1998) rescored the WISC-III protocols of 400 gifted elementary students and concluded that the Vocabulary-Block Design short form was the best dyad in estimating FSIQ. With respect to administration time, the Information-Coding dyad was found to be the most efficient dyad. The fastest triad (Information, Coding, Similarities), tetrad (Information, Coding, Similarities, Vocabulary), and pentad (Information, Coding, Similarities, Vocabulary, Arithmetic, Comprehension) were also identified. Terminie reported that while short form dyads should not be used alone in the identification of gifted students, they seem to be appropriate instruments for screening purposes. What the researcher failed to acknowledge, however, is that the use of a short

form with less than four subtests might yield different results from the FSIQ due to issues surrounding motivation and attention (Thompson, 1987; Thompson et al., 1986).

At this time, there appears to be only one study examining the use of WISC-III short forms with gifted students using Canadian norms. Mark, Beal, and Dumont (1998) investigated the Dumont-Faro (1993) short form (Information, Vocabulary, Picture Completion, Coding, Block Design) of the WISC-III as a screening instrument for identifying gifted students in a Toronto school district. In this study, giftedness was defined as a FSIQ of 130 or above. When the WISC-III protocols of 158 potentially gifted Grade 3 students were rescored, 128 of the subjects (81%) were correctly identified as being either gifted or non-gifted. Twenty-five students fell within the area immediately below the gifted range and were identified as “possibly gifted” (i.e., 123-129 IQ). Finally, there were five students (3%) who were identified as being gifted based on the SFIQ, but were misclassified as their FSIQ scores were below 130.

Mark and his colleagues suggested a number of guidelines for employing the Dumont-Faro short form for gifted eligibility purposes. First, school psychologists should administer the abbreviation to all students referred for gifted assessment. Second, students should be accepted into gifted programming who receive a SFIQ of 130 or above, however, if the student demonstrates “one or two exceedingly high subtest scores” the full test battery should be administered to confirm overall ability. Third, students who receive a SFIQ estimate that is less than 123 should be excluded from further assessment. Finally, students who receive an SFIQ estimate between 123 to 129 should be administered the remaining subtests in order to confirm their true ability. The

researchers concluded that the Dumont-Faro short form is useful for determining gifted eligibility for Canadian students.

Summary

Short forms of well-established intelligence tests such as the Wechsler scales save time, effort, and resources. Consequently, these instruments are especially useful for school psychologists who endeavor to meet increasing service demands while at the same time providing effective intellectual assessments. Although short forms do not possess the same degree of validity and reliability as long forms, they are viable for screening, re-evaluation, and research purposes. A short form can provide a relatively accurate and efficient global estimate of ability, while enabling the school psychologist to spend more time in other components of the assessment process or to screen a greater number of students. The practical application of short forms may be demonstrated in the screening of gifted students who are eligible to participate in special programming. Although the SFIQ is only an estimate of the FSIQ, it offers a level of precision which is acceptable for screening purposes provided these results are combined with other selection criteria (e.g., nominations, portfolios, superior grades, achievement tests, etc.).

Despite the apparent utility of short forms, these instruments should never be used as the sole means in making placement or educational decisions. Depending on the circumstance, the use of short forms may fail to provide a broad enough representation of the student's true potential. Furthermore, the opportunity for making important qualitative observations is greatly diminished when a shortened instrument is used. Clearly, it is the responsibility of the school psychologist to determine when the short form is sufficient and when the full form of the intelligence test is warranted.

Hypothesis for the Current Investigation

There will be no statistically significant differences between estimated short form IQ scores and actual Full Scale IQ scores of the WISC-III.

Research Questions

1. Can short forms of the WISC-III effectively discriminate between gifted and non-gifted elementary students in Canada?
2. Does gender or chronological age effect the accuracy of short forms of the WISC-III in discriminating between gifted and non-gifted elementary students in Canada?
3. How accurate are short forms of the WISC-III for estimating the Full Scale IQ scores of potentially gifted elementary students in Canada?
4. Which short form of the WISC-III provides the best compromise among psychometric soundness, clinical utility, and discriminating power when identifying potentially gifted elementary students in Canada?
5. Is linear equating a more accurate method than multiple regression for estimating the Full Scale IQ scores of potentially gifted elementary students in Canada?

Rationale for the Selection of Short Forms Used in the Current Investigation

A number of criteria were considered in choosing the six WISC-III selected subtest short forms used in this investigation. First, the short form had to be quick to administer (Kaufman, 1990; Kaufman et al., 1991; Kaufman et al., 1996; Sattler, 2001; Silverstein, 1985a, 1990a). Second, the short form had to contain no less than four subtests (Thompson, 1987; Thompson et al., 1986). Third, the short form had to be psychometrically sound (Cyr & Brooker, 1984; Kaufman, 1990; Kaufman et al., 1991; Kaufman et al., 1996; Sattler, 2001; Silverstein, 1985a). Fourth, the short form had to

measure a variety of mental processes (Kaufman, 1990; Sattler, 2001). Fifth, the short form had to provide a clinically interesting unit (Kaufman, 1990; Kaufman et al., 1996; Sattler, 2001; Silverstein, 1982). Sixth, the short form had to include an equal or near equal number of Verbal and Performance subtests (Doppelt, 1956; Kaufman, 1990; Sattler, 2001). Finally, consideration was given to the similarity between the gifted population in the present investigation and the populations used in previous validation research (Beck et al., 1983; Vollmerhausen et al., 1986).

In order to determine the extent to which each short form measures a variety of mental processes, the Cattell-Horn-Carroll (CHC) theory of intelligence was used as a benchmark. The CHC theory is thought to be the most comprehensive and empirically supported model of cognitive ability currently available (Flanagan, McGrew, & Ortiz, 2000; McGrew, 1997; McGrew & Flanagan, 1998). This theory represents a hierarchical model of intelligence that holds “g”, or general intelligence, at the first level. Eight broad cognitive areas comprise the second level (i.e., fluid reasoning, crystallized intelligence, long-term storage and retrieval, visual processing, auditory processing, short-term memory, processing speed, and quantitative ability), while the third level consists of more than 70 narrower cognitive abilities. The WISC-III is capable of measuring five of the eight broad cognitive areas within the CHC theory (e.g., crystallized intelligence, visual processing, short-term memory, processing speed, and quantitative ability).

Short Forms Used in the Current Investigation

The six previously researched selected subtest short forms chosen for the present study are all comprised of four or more subtests. In order to address administration time,

psychometric utility, breadth and depth of abilities measured, an overview of each short form is offered. Table 1 provides a summary of all six short forms.

Short Form 1 (SF1): Similarities, Arithmetic, Vocabulary, Picture Completion, Coding, and Block Design (Donders, 1997). The Donders short form, which is based on an equal number of Verbal and Performance subtests, represents a clinically meaningful set of measures. The Similarities subtest provides the opportunity to assess language-based skills through a different format than the “what is” style assessed by Vocabulary. Moreover, the Similarities and Vocabulary subtests both have high factor loadings on the Verbal Comprehension Index. The Block Design and Picture Completion subtests both load high on the Perceptual Organization Index. Furthermore, the Picture Completion subtest is very brief to administer. The Arithmetic subtest, which assesses mental computation and concentration, provides a measure of the Freedom from Distractibility factor. The Coding subtest, an information processing task that involves discrimination and memory of visual information, taps into the Processing Speed factor.

Within the context of the CHC theory, the Donders short form is representative of four broad ability areas. Although Similarities and Vocabulary are both strong measures of crystallized intelligence, they assess two different areas within this broad ability. The Similarities subtest is a measure of language development, while Vocabulary is a measure of lexical knowledge. Block Design and Picture Completion are both measures of visual processing skill. More specifically, Block Design is a strong test of spatial relations, while Picture Completion is a moderate test of visualization. The Arithmetic subtest provides a strong measure of quantitative ability and the Coding subtest offers a strong indication of processing speed.

Table 1

Summary of the Short Forms Used in the Current Investigation

Short Form	Subtests*	Time	CHC Broad Cognitive Areas	Correlation with FSIQ	Research with gifted
SF1 (Donders, 1997)	Sim, Ari, Voc, PC, Cd, BD	28 min.	crystallized, visual processing, quantitative, processing speed	.90 - .98	no
SF2 (Doppelt, 1956)	Voc, Ari, PA, BD	24 min.	crystallized, visual processing, quantitative	.89 - .97	yes
SF3 (Dumont & Faro, 1993)	Inf, Voc, PC, Cd, BD	24 min.	crystallized, visual processing, processing speed	.89 - .96	yes
SF4 (Sattler, 1992)	Inf, Voc, PC, BD	20 min.	crystallized, visual processing	.93	no
SF5 (Kaufman et al., 1991)	Sim, Ari, PC, Cd	17 min.	crystallized, visual processing, quantitative, processing speed	.88 - .95	no
SF6 (Kaufman et al., 1996)	Sim, Ari, PC, BD	21 min.	crystallized, visual processing, quantitative	.93	no

* Note. Sim = Similarities, Ari = Arithmetic, Voc = Vocabulary, PC = Picture

Completion, Cd = Coding, BD = Block Design, PA = Picture Arrangement, Inf = Information.

Although slightly longer than other abbreviations (i.e., estimated to require 28 minutes according to Terminie, 1998), the Donders (1997) short form maintains the original four-factor structure of the WISC-III. Furthermore, previous investigations indicated that this abbreviation correlates highly (ranging between .90 to .98) with the FSIQ. At this time, there does not appear to be any research related to the use of the Donders short form with gifted students.

Short Form 2 (SF2): Vocabulary, Arithmetic, Picture Arrangement, and Block Design (Doppelt, 1956). Although this short form was originally developed by Doppelt (1956) for the WAIS, it has been applied to the WAIS-R (Silverstein, 1982), the WISC-R (Beck et al., 1983; Donders, 1992; Kaufman, 1976; Vollmerhausen et al., 1986) and the WISC-III (Cravens, 1999). Previous research with this abbreviation indicated that it correlates highly (ranging between .89 to .97) with the FSIQ. The Doppelt short form, which is comprised of an equal number of Verbal and Performance subtests, provides a clinically interesting unit of measures. The Vocabulary subtest, a good indicator of “g”, evaluates expressive language. The Arithmetic subtest, a test of mental computation, represents the Freedom from Distractibility factor. The Block Design subtest offers information regarding non-verbal problem solving that involves analysis and synthesis of information. Picture Arrangement, a nonverbal reasoning task that utilizes planning, organization, and sequencing skills, assesses an individual’s ability to comprehend and evaluate a situation.

According to the CHC theory of intelligence, the Doppelt short form represents three of the broad factors of ability. As the Vocabulary subtest is a strong measure of crystallized intelligence and the Arithmetic subtest represents a strong indication of

quantitative ability, these two Verbal subtests tap into a broad range of skills. Block Design is considered to be an excellent measure of visual processing ability, while Picture Arrangement is thought to be a moderate measure of the same ability. Nevertheless, as Block Design is a test of spatial relationships and Picture Arrangement is a test of visualization, these Performance subtests tap into two different types of visual processing.

The Doppelt short form of the WISC-R has been used with gifted students (Linn & Lopatin, 1990). Furthermore, other researchers have suggested that this abbreviation provides greater accuracy with cognitively advanced individuals (Vollmerhausen et al., 1986; Zimet & Adler, 1990). It is estimated that the Doppelt short form requires 24 minutes to administer (Terminie, 1998).

Short Form 3 (SF3): Information, Vocabulary, Picture Completion, Coding, and Block Design (Dumont & Faro, 1993). The Dumont-Faro short form represents a clinically useful set of measures. The Information subtest provides data related to long-term memory based on experience and school-based knowledge. The Vocabulary subtest, which is considered to be a strong measure of “g”, provides insight into expressive language and the retrieval of stored information. Block Design, another strong measure of “g”, provides information about visual-spatial processing. Picture Completion, a test of visual discrimination, is easy to score and offers information regarding word retrieval and word-finding difficulties. Although the addition of Coding resulted in an unequal number of Verbal and Performance subtests, it was included so that the Processing Speed factor could be represented.

The Dumont-Faro short form taps into three areas of intelligence according to the CHC model. The Information and Vocabulary subtests are both strong measures of crystallized intelligence. Nevertheless, Information specifically assesses general verbal information, while Vocabulary measures lexical knowledge. On the Performance Scale, Block Design and Picture Completion both tap into two different types of visual processing ability. Block Design is considered to be a strong measure of spatial relationships, while Picture Completion is a moderate measure of visualization. The Coding subtest provides a strong indication of processing speed ability.

Previous investigations have demonstrated that the Dumont-Faro short form correlates highly (ranging between .89 to .96) with the FSIQ. To date, this abbreviation of the WISC-III is the only short form to be used with Canadian normative data. Furthermore, this investigation was based on a sample of gifted students. The Dumont-Faro abbreviation is estimated to require approximately 24 minutes to administer (Terminie, 1998).

Short Form 4 (SF4): Information, Vocabulary, Picture Completion, and Block Design (Sattler, 1992). Sattler (1992) reported this short form to be among the top 10 tetrads in terms of validity and reliability. It is represented by an equal number of Verbal and Performance subtests, in addition to providing a clinically interesting combination of measures. Previous research has found the Sattler short form to correlate highly (.93) with FSIQ scores when applied in a clinical population.

The Sattler short form is somewhat limited in its range, as it represents only two areas of ability according to the CHC theory of intelligence. Although the Information and Vocabulary subtests are both strong measures of crystallized intelligence, they assess

two different skills within this broad area. Information is a test of general verbal information, while Vocabulary is a test of lexical knowledge. The Block Design and Picture Completion subtests are both measures of visual processing ability. More specifically, Block Design is a strong test of spatial relationships and Picture Completion is a moderate test of visualization.

While the Sattler abbreviation is very similar to the Dumont-Faro short form, it does not include the Coding subtest. There is some evidence to suggest that gifted students may demonstrate lower scores on the Processing Speed factor of the WISC-III (which includes the Coding subtest) as compared to the other three factors (Wechsler, 1991). Consequently, research involving the Sattler short form will be useful in determining the extent to which the Coding subtest adds to the Dumont-Faro short form. There does not appear to be any previous research involving this short form with gifted subjects. It was estimated that the Sattler short form requires approximately 20 minutes to administer (Terminie, 1998).

Short Form 5 (SF5): Similarities, Arithmetic, Picture Completion, and Coding (Kaufman, Ishikuma, & Kaufman-Packer, 1991). Kaufman and his colleagues originally developed this short form for the WAIS-R (using the Digit Symbol subtest instead of Coding). Previous research studies have found this abbreviation to correlate highly (.88 to .95) with FSIQ scores. Moreover, Cravens (1999) discovered that when this short form was used with the WISC-III it correctly classified a higher percentage of subjects compared to the Doppelt short form. The Kaufman et al. (1991) short form is estimated to require only 17 minutes to administer (Terminie, 1998), thus is one of the briefest

tetrads available. There appear to be no reported studies related to the use of this short form with gifted subjects.

This short form, which is comprised of an equal number of Verbal and Performance subtests, provides a clinically interesting unit. The Similarities subtest is a language-based task that measures verbal concept formation. The Arithmetic subtest taps into mental computation and concentration, in addition to being representative of the Freedom from Distractibility factor. Picture Completion is a test of visual discrimination that also lends insight into word retrieval. Coding is an information processing task that involves discrimination and memory of visual information. It provides representation of the Processing Speed factor.

The Kaufman et al. (1991) abbreviation represents four broad cognitive areas of CHC theory. The Similarities subtest is a strong measure of crystallized intelligence, while the Arithmetic subtest is a strong measure of quantitative ability. Although Picture Completion taps into visual processing, this subtest is thought to be only a moderate indicator of ability. The Coding subtest is considered to be a strong measure of processing speed.

Short Form 6 (SF6): Similarities, Arithmetic, Picture Completion, and Block Design (Kaufman, Kaufman, Balgopal, & McLean, 1996). Kaufman et al. (1996) proposed this short form as the best compromise between clinical and psychometric properties, coupled with brevity of administration. Research with this abbreviation showed that it correlates highly (.93) with the full form. It is estimated to require 21 minutes to administer (Terminie, 1998). Thus far, the Kaufman et al. (1996) short form has not been used with gifted populations.

This abbreviation is very similar to the Kaufman et al. (1991) short form, the only difference is that Block Design is used instead of Coding. By excluding the processing speed measure, however, this short form is only representative of three broad areas of CHC ability. Nevertheless, as there is some research that suggests gifted students score lower on the Processing Speed factor of the WISC-III as compared to the other three factors (Wechsler, 1991), it would be useful to include this short form as a means of comparison with the Kaufman et al. (1996) abbreviation. Regarding the remaining subtests, Similarities is a strong measure of crystallized intelligence and the Arithmetic subtest is a strong measure of quantitative reasoning. Although both Block Design and Picture Completion are measures of visual processing, they tap into two different areas of ability. More specifically, Block Design is a strong test of spatial relationships, while Picture Completion is a moderate measure of visualization.

CHAPTER THREE: METHODOLOGY

Instrumentation

The Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) is the most recent version of a well-established measure of general intelligence for children between the ages of 6 years, 0 months through 16 years, 11 months. This instrument consists of ten standard subtests and three supplemental subtests. The raw scores of each subtest are converted into standard scores, which have a mean of 10 and a standard deviation of 3. The standard scores of the five primary Verbal subtests are summed and converted into a Verbal IQ score. The same procedure is used with the five primary Performance subtests to yield a Performance IQ score. The Verbal and Performance IQ scores are then combined and converted into a Full Scale IQ. Each IQ score has a mean of 100 and a standard deviation of 15.

Standardization of the WISC-III

The American normative data for the WISC-III was obtained from a standardization sample of 2,200 children (Wechsler, 1991). The sample was divided into 11 age groups, with 100 males and 100 females comprising each group. The proportion of children selected from different demographic groups (i.e., age, race/ethnicity, geographic region, and parent education) was consistent with the normal U.S. population according to the 1988 Census survey.

A standardization sample of 1,100 Canadian children was used to develop the Canadian normative data (Wechsler, 1996). For each of the 11 age groups, 50 males and 50 females were represented. A stratified random sampling approach was used to ensure

that representative proportions of children from different demographic groups (based on the Statistics Canada 1986 Census) would be included in the sample.

Reliability and Validity of the WISC-III

Reliability, or the stability of test scores over time, is crucial for any psychometric instrument (Anastasi & Urbina, 1997). Reliability coefficients for the WISC-III subtests (excluding Coding and Symbol Search, as they are speeded tests) were calculated using the split-half method corrected by the Spearman-Brown formula. In the American normative sample, reliability coefficients for individual subtests across all ages ranged from .87 (Vocabulary and Block Design) to .69 (Object Assembly). Reliability coefficients for Verbal, Performance, and Full Scale IQ scores were .95, .91, and .96 respectively (Wechsler, 1991). Test-retest coefficients for the Verbal, Performance, and Full Scale IQ scores were reported as .94, .87, and .94, respectively.

Regarding the Canadian normative sample, reliability coefficients for individual subtests across all ages ranged from .87 (Vocabulary) to .61 (Object Assembly). The reliability of Verbal, Performance, and Full Scale IQ scores for the Canadian normative sample were .93, .89, and .95 respectively (Wechsler, 1996). Due to a lack of appropriate Canadian test-retest data, coefficients for Verbal, Performance, and Full Scale IQ scores were not reported. Nevertheless, it is unlikely that Canadian test-retest data would have differed significantly from the American sample.

Validity refers to whether a test is actually measuring what it purports to measure (Anastasi & Urbina, 1997). To establish the validity of an instrument, independent research must be carried out to determine whether the new test is highly correlated with other measures of intelligence. While considerable evidence of the instrument's validity

is presented in both the WISC-III American (Wechsler, 1991) and Canadian (Wechsler, 1996) manuals, there was limited documentation regarding gifted populations. In the American manual, only one study of gifted children was reported. This research was based on 38 children, aged 4 through 14 years, who were identified as gifted by independent evaluations. The mean WISC-III Verbal, Performance, and Full Scale IQ scores for these gifted subjects were 128.0, 124.6 and 128.7 respectively.

Importance of Canadian Normative Data

Canadian WISC-III validation research provided evidence that Canadian children performed differently on the WISC-III as compared to their American counterparts (Wechsler, 1996). The WISC-III Canadian Supplement Manual (1996) reported that Canadian children tend to receive significantly higher Verbal, Performance, and Full Scale IQ scores utilizing the American normative data. More specifically, Canadian children were found on average to receive Verbal IQ scores 1.40 points higher, Performance IQ scores 4.96 points higher, and Full Scale IQ scores 3.34 points higher than their American counterparts. Subsequent research with a group of Canadian students with learning disabilities corroborated the initial finding that Canadian and American children perform differently on the WISC-III (Beal, Dumont, Cruse, & Branche, 1996).

Data Collection Procedure

Data for this study was collected from two sources: the Canadian WISC-III standardization sample and archival Canadian school psychology files containing WISC-III protocols. The Psychological Corporation compiled the WISC-III standardization data that consisted of 192 total cases. An equal number of gifted ($n = 96$) and non-gifted

($n = 96$) subjects were represented, with a total of 102 males and 90 females. The standardization data served as the analysis sample for this research study. With respect to the archival data, documentation was gathered from a total of 1,058 files, with an equal number of gifted ($n = 529$) and non-gifted ($n = 529$) subjects. Of these subjects, 682 were males and 376 were females. The archival data was used to form the cross validation sample in order to examine the external validity of the results.

The archival data was obtained from a Western Canadian urban school division (the Calgary Board of Education) consisting of approximately 100,000 students. Criteria for selecting files for this study included the following: elementary aged student (i.e., Grades 1 through 6); WISC-III protocol (scored using Canadian norms) administered between September 1997 to May 2001; and FSIQ score of 120 or above. Although children with Full Scale IQ scores of 130 and above are considered eligible to apply to the gifted program within the school district studied, FSIQ scores of 120 and above were documented in order to take into account the standard error of measurement. The following information was collected from the selected WISC-III protocols: gender, chronological age, subtest standard scores, Verbal IQ score, Performance IQ score, and Full Scale IQ score.

Data was also collected for a group of “non-gifted” subjects (e.g., individuals with FSIQ scores between 85-115) who were matched with the gifted subjects in terms of gender and chronological age. Information from the non-gifted sample was gathered in order to provide a wider range of correlations for analysis. This was an important aspect of the study, as the homogeneity of the gifted sample alone would have limited the range of observed variance, thus affecting the obtained correlations.

The principal investigator gathered all the archival data for this study. In order to ensure that all data was recorded accurately from the test protocols, the subtest scores were summed and the Verbal, Performance, and Full Scale IQ scores were recalculated for each data set.

Data Analysis

Data analyses in this investigation included the following:

- (1) Descriptive statistics were calculated to summarize the age and gender of the subjects, in addition to providing subtest and scale score means and standard deviations;
- (2) A predictive discriminant analysis was used for every short form in order to define the discriminant functions that maximally separated the two groups (i.e., gifted and non-gifted) for the total sample. Discriminant analyses was also conducted with the sample after it was divided into male and female groups, and once again after divided into chronological age groups (6 – 7, 8 – 9, 10 – 12);
- (3) SFIQ scores were calculated for every short form based on the linear scaling technique developed by Tellegen and Briggs (1967);
- (4) For every short form, a standard multiple regression analysis was performed to determine the weighted formula required to obtain SFIQ scores;
- (5) Pearson product-moment correlation coefficients were performed to determine the significance of the correlations between the SFIQ scores and actual FSIQ scores;
- (6) Paired t-tests, using the Bonferroni correction for simultaneous comparisons, were performed to determine if mean SFIQ scores were significantly different from mean

FSIQ scores. Effect size was taken into consideration to determine the magnitude of the differences between the mean scores;

- (7) To determine the psychometric properties of the six short forms, reliability and validity coefficients were calculated based on the standardization data reported in the WISC-III Canadian manual. The standard error of measurement and standard error of estimate were also calculated for each of the short forms;
- (8) To verify the outcomes yielded from the analysis sample, the results obtained from the initial investigation were applied to the cross validation sample.

CHAPTER FOUR: RESULTS

Introduction

All raw data was entered onto a spreadsheet and subjected to statistical analyses using the Statistical Package for the Social Science (SPSS) program version 10.0. Unless otherwise specified, the alpha level for each analysis was set at $p < .05$. This level of significance was chosen in order to achieve a balance between committing either a Type I (i.e., rejecting a true hypothesis) or Type II (i.e., failing to reject a false hypothesis) error. Data analyses in this investigation included the following: descriptive statistics, discriminant analysis, linear equating, multiple regression, Pearson product-moment correlations, statistics for determining the psychometric properties of the short forms, and paired t-tests.

Descriptive Statistics

The WISC-III standardization group ($n = 192$) was the analysis sample in this investigation, while the CBE group ($n = 1,058$) served as the cross validation sample. Both samples were comprised of an equal proportion of gifted and non-gifted subjects. The analysis sample included 102 males and 90 females, while the cross validation sample consisted of 682 males and 376 females. Although there appears to be a disproportionate number of males (63%) to females (37%) in the combined analysis and cross validation samples, it is consistent with the number of males (62%) to females (38%) identified as being gifted within the Calgary Board of Education (K. Roberts, personal communication, June 13, 2001). Neither the analysis nor the cross validation samples contained any missing data. Given that the greatest number of independent variables used in this study was six, coupled with the statistical guideline of twenty

participants per variable (Hair, Anderson, Tatham, & Black, 1998), both samples satisfied the minimum requirement for the study. Table 2 provides an overview of the two samples according to chronological age and gender.

Based on the full WISC-III administration for the total analysis sample, the mean Verbal IQ was 111.23 ($SD = 13.54$, range 79 to 138), the mean Performance IQ was 115.10 ($SD = 14.30$, range 82 to 146), and the mean Full Scale IQ was 114.18 ($SD = 13.48$, range 85 to 144). Table 3 provides the means and standard deviations for WISC-III subtests and IQ scales according to classification (i.e., non-gifted, gifted, and total). For the total cross validation sample, the mean Verbal IQ was 114.72 ($SD = 18.68$, range 86 to 154), the mean Performance IQ was 110.46 ($SD = 15.10$, range 85 to 152), and the mean Full Scale IQ was 113.59 ($SD = 17.60$, range 85 to 155). An overview of WISC-III means and standard deviations for the cross validation sample is provided in Table 4.

An examination of the means and standard deviations of the two samples suggested that while the overall subtest and scaled scores were very similar, the cross validation sample yielded higher gifted scores and lower non-gifted scores. The cross validation sample scores also tended to demonstrate greater variability, as evidenced by the larger standard deviations.

Discriminant Analysis

Discriminant analysis is a statistical procedure used for predicting group membership from a set of continuous predictor variables (Hair et al., 1998). Discriminant analysis accomplishes this task by finding the linear combinations of predictor variables that best capture the differences between the groups. According to Huberty (1994), discriminant analysis is an effective technique for classifying individuals to specific

Table 2

Size of the Analysis Sample and the Cross Validation Sample by Chronological Age and Gender

Age	<u>Analysis Sample (n = 192)</u>			<u>Cross Validation Sample (n = 1,058)</u>		
	Male	Female	Total	Male	Female	Total
6	24	18	42	10	10	20
7	6	6	12	68	20	88
8	10	16	26	214	130	344
9	22	16	38	138	62	200
10	10	6	16	114	72	186
11	16	14	30	122	72	194
12	14	14	28	16	10	26

Table 3

Means and Standard Deviations for WISC-III Subtests and IQ Scales by Classification
for the Analysis Sample (n = 192)

WISC-III Subtests and IQ Scales	<u>Non-gifted</u>		<u>Gifted</u>		<u>Total</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Information	10.11	2.37	13.34	2.08	11.73	2.75
Similarities	9.95	2.54	13.55	1.94	11.75	2.89
Arithmetic	9.96	2.37	13.38	2.42	11.67	2.94
Vocabulary	10.13	2.78	14.11	2.39	12.12	3.27
Comprehension	10.35	2.76	13.76	2.57	12.06	3.16
Digit Span	10.40	2.89	11.78	2.80	11.09	2.92
Picture Completion	10.69	2.66	13.62	2.30	12.16	2.88
Coding	10.45	2.79	13.47	2.96	11.96	3.25
Picture Arrangement	10.07	2.98	13.67	3.14	11.87	3.54
Block Design	10.56	2.70	14.48	2.76	12.52	3.36
Object Assembly	11.41	2.57	13.89	2.37	12.65	2.76
Symbol Search	10.59	2.78	13.59	2.99	12.09	3.25
Verbal IQ	100.84	9.90	121.61	7.25	111.23	13.54
Performance IQ	104.47	9.88	125.74	9.21	115.10	14.30
Full Scale IQ	102.57	8.19	125.78	5.11	114.18	13.48

Table 4

Means and Standard Deviations for WISC-III Subtests and IQ Scales by Classification
for the Cross Validation Sample (n = 1,058)

WISC-III Subtests and IQ Scales	<u>Non-gifted</u>		<u>Gifted</u>		<u>Total</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Information	9.56	2.32	15.02	2.46	12.29	3.63
Similarities	10.23	2.01	15.36	2.11	12.80	3.29
Arithmetic	9.01	2.22	14.75	2.99	11.88	3.90
Vocabulary	9.77	1.94	15.69	2.50	12.73	3.71
Comprehension	10.03	2.14	14.74	2.66	12.38	3.37
Digit Span	8.69	2.26	11.80	2.91	10.16	3.01
Picture Completion	9.77	2.01	12.88	2.37	11.33	2.69
Coding	8.92	2.51	12.17	2.99	10.55	3.20
Picture Arrangement	10.02	2.45	13.57	3.09	11.79	3.30
Block Design	10.30	2.29	15.03	2.80	12.67	3.48
Object Assembly	9.58	2.23	12.71	2.55	11.14	2.86
Symbol Search	9.88	2.67	13.23	2.81	11.49	3.21
Verbal IQ	98.15	7.84	131.29	9.34	114.72	18.68
Performance IQ	98.16	7.96	122.76	9.48	110.46	15.10
Full Scale IQ	97.54	7.44	129.64	6.98	113.59	17.60

populations. Pyryt and Heck (1991) recommend the use of multivariate techniques, particularly discriminant analysis, for investigations involving gifted populations.

Predictive discriminant analysis was carried out for each of the six WISC-III short forms to determine which set of subtests most clearly differentiated the gifted and non-gifted groups. Group membership (i.e., gifted or non-gifted) was the dependent variable, while the individual subtests in each short form were the independent variables. A different discriminant analysis was conducted for each short form to determine the optimum weighting of subtests in predicting group membership.

A single discriminant function was identified that maximized the separation of the gifted and non-gifted groups on the vector of scores for each of the six short forms. The resulting functions were subjected to a chi-square test in order to determine significance of the functions. The discriminant functions for the six short forms, which ranged from .822 to .778, each produced large canonical correlations that were found to be statistically significant ($p < .05$). The amount of variance in classification accounted for by function 1 was 68% in SF1, 64% in SF2, 64% in SF3, 61% in SF4, 61% in SF5, and 63% in SF6. The discriminant function analyses results are presented in Table 5.

In order to explore whether gender influenced the discrimination between gifted and non-gifted subjects, the analysis sample was divided into male and female groups. For each of the six short forms, a different discriminant analysis was conducted for males and females in the sample. Consistent with the total analysis sample, male and female groups both produced similar canonical correlations based on each of the six short forms. Furthermore, all results were statistically significant. It was noted, however, that males were observed to have slightly lower canonical correlations (ranging from .799 to .741)

Table 5

Summary of Discriminant Analyses Between Gifted and Non-Gifted Students (n = 192)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1 ^a	1	.822	.676	.324	210.500	6	.0001
2 ^b	1	.802	.643	.357	193.871	4	.0001
3 ^c	1	.802	.643	.356	193.579	5	.0001
4 ^d	1	.778	.605	.395	174.789	4	.0001
5 ^e	1	.783	.613	.387	178.670	4	.0001
6 ^f	1	.791	.626	.374	184.965	4	.0001

^a = Similarities, Arithmetic, Vocabulary, Picture Completion, Coding, Block Design.

^b = Vocabulary, Arithmetic, Picture Arrangement, Block Design.

^c = Information, Vocabulary, Picture Completion, Coding, Block Design.

^d = Information, Vocabulary, Picture Completion, Block Design.

^e = Similarities, Arithmetic, Picture Completion, Coding.

^f = Similarities, Arithmetic, Picture Completion, Block Design.

as compared to females (ranging from .863 to .812). Table 6 provides an overview of the results pertaining to males, while Table 7 offers the results for females.

Further investigation was conducted in an attempt to determine whether chronological age was a factor that might impact the ability of a short form in discriminating between gifted and non-gifted subjects. The analysis sample was divided into the following three age groups: 6 – 7, 8 – 9, and 10 – 12. A different discriminant analysis was conducted for each of the six short forms for the 6 – 7 group, the 8 – 9 group, and the 10 – 12 group. Based on the six short forms, all three groups produced similar canonical correlations as the total analysis sample. Moreover, all results were found to be statistically significant. While the 6 – 7 group tended to yield marginally lower correlations overall (ranging from .827 to .759), the 8 – 9 group was observed to obtain slightly higher correlations (ranging from .861 to .800). The correlations for the 10 – 12 group (ranging from .845 to .777) were very similar to the total analysis sample. Detailed results of these analyses are presented in Tables 8, 9, and 10.

The discriminant weights (standardized discriminant function coefficients) and the discriminant loadings (structure coefficients) between the independent variables and the discriminant function were computed for each of the six short forms for the total analysis sample. The standardized coefficients provide an index of the relative importance each variable has when discriminating between groups. Examination of these coefficients tends to be the most common method of interpreting the discriminant function (Huberty, 1994). Although variables with the highest absolute value typically contribute most to group separation, multicollinearity among the independent variables may result in smaller weights (Hair et al., 1998). The structure coefficients are

Table 6

Summary of Discriminant Analyses Between Gifted and Non-Gifted Males (n = 102)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1	1	.799	.638	.362	98.540	6	.0001
2	1	.773	.597	.402	89.203	4	.0001
3	1	.782	.611	.389	92.121	5	.0001
4	1	.741	.549	.450	78.199	4	.0001
5	1	.768	.589	.409	87.500	4	.0001
6	1	.744	.553	.446	79.115	4	.0001

Table 7

Summary of Discriminant Analyses Between Gifted and Non-Gifted Females (n = 90)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1	1	.863	.744	.255	116.081	6	.0001
2	1	.848	.719	.280	109.372	4	.0001
3	1	.843	.710	.290	105.919	5	.0001
4	1	.834	.695	.305	102.241	4	.0001
5	1	.812	.659	.340	92.708	4	.0001
6	1	.852	.725	.273	111.500	4	.0001

Table 8

Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages 6 - 7(n = 54)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1	1	.827	.683	.316	56.408	6	.0001
2	1	.809	.654	.345	53.249	4	.0001
3	1	.799	.638	.362	50.362	5	.0001
4	1	.763	.582	.418	43.655	4	.0001
5	1	.759	.576	.424	42.960	4	.0001
6	1	.783	.613	.387	47.511	4	.0001

Table 9

Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages 8 - 9(n = 64)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1	1	.854	.729	.271	76.950	6	.0001
2	1	.861	.741	.258	81.226	4	.0001
3	1	.846	.715	.285	74.670	5	.0001
4	1	.836	.698	.301	72.036	4	.0001
5	1	.800	.640	.360	61.369	4	.0001
6	1	.810	.656	.343	64.165	4	.0001

Table 10

Summary of Discriminant Analyses Between Gifted and Non-Gifted Students Ages10 – 12 (n = 74)

SF	Test of Function	Canonical Correlation	Squared Canonical Correlation	Wilks' Lambda	Chi Square	df	p
1	1	.845	.714	.286	86.456	6	.0001
2	1	.777	.603	.396	64.904	4	.0001
3	1	.842	.708	.291	85.690	5	.0001
4	1	.819	.670	.329	77.844	4	.0001
5	1	.834	.703	.304	83.263	4	.0001
6	1	.831	.690	.309	82.161	4	.0001

correlations between the predictor variables and the discriminant function. As structure coefficients reflect common variance among the independent variables, they can be more useful for interpretation purposes as compared to standardized coefficients (Hair et al, 1998). Variables with higher absolute values ($\geq .40$) for both weights and loadings most clearly contribute to group discrimination. When a variable has a high absolute value on only one of the indices, however, contribution to group differences is less certain (Lupart & Pyryt, 1996). Standardized discriminant function coefficients and structure coefficients for each of the six short forms in the current investigation are presented in Table 11.

The pattern of coefficients in SF1 suggests that Similarities and Block Design were the best predictors for distinguishing between gifted and non-gifted subjects as they both showed values $\geq .40$. Vocabulary (.535), Arithmetic (.497), and Picture Completion (.411) yielded good loadings, however, their weights (.317, .336, .181, respectively) were all below the established cutoff value. Coding, with a weight of .360 and loading of .366 appeared to contribute the least in terms of discriminating between the two groups.

Vocabulary, Block Design, and Picture Arrangement were found to be the best predictors for discriminating between gifted and non-gifted subjects in SF2 as they all demonstrated values $\geq .40$. With a loading of .534 and a weight of .385, Arithmetic did not appear to add as much as the other variables in this particular short form.

In SF3, Vocabulary and Block Design discriminated the most between subjects in the gifted and the non-gifted groups. Information, with a loading of .541 and a weight of .395, also appeared to be a reasonable discriminator. Picture Completion (loading of .252

Table 11

Summary of Standardized Discriminant Function and Structure Coefficients for the
Analysis Sample (n = 192)

Independent Variables	SF1		SF2		SF3	
	Stand.	Struct.	Stand.	Struct.	Stand.	Struct.
Information					.395	.541
Similarities	.427	.556				
Arithmetic	.336	.497	.385	.534		
Vocabulary	.317	.535	.578	.575	.444	.575
Picture Completion	.181	.411			.252	.442
Coding	.360	.366			.393	.392
Picture Arrangement			.407	.439		
Block Design	.440	.499	.528	.536	.495	.536

Independent Variables	SF4		SF5		SF6	
	Stand.	Struct.	Stand.	Struct.	Stand.	Struct.
Information	.398	.588				
Similarities			.638	.637	.600	.620
Arithmetic			.455	.570	.432	.554
Vocabulary	.465	.624				
Picture Completion	.305	.479	.321	.471	.250	.459
Coding			.438	.419		
Picture Arrangement						
Block Design	.567	.582			.491	.557

Note. Bold print indicates variables with a cutoff loading $\geq .40$ that were correlated with the discriminant function.

and weight of .442) and Coding (loading of .393 and weight of .392) seemed to furnish the least in terms of discriminating gifted from non-gifted subjects.

Similar findings were noted in SF4 as Vocabulary and Block Design were observed to be the best contributors in discriminating between subjects. Information, with a loading of .588 and a weight of .398, was also found to be a fair predictor of group membership. Picture Completion, with a loading of .479 and a weight of .305, appeared to provide the smallest amount of discriminating power.

Similarities, Arithmetic, and Coding were found to be the best predictors for distinguishing between gifted and non-gifted subjects on SF5. Picture Completion, with a loading of .471 and a weight of .321, seemed to offer less than the other variables in terms of discriminating between the two groups.

Finally, Similarities, Block Design, and Arithmetic contributed the most to overall discrimination in SF6. Picture Completion, with a loading of .459 and a weight of .250, appeared to provide the least amount of discriminating power.

Classification Results for the Total Analysis Sample

A classification analysis was performed in order to examine the effectiveness of the discriminant function in predicting group membership. Based on the discriminant analyses and the resulting models, it was concluded that all six short forms performed exceptionally well. Albeit not statistically significant, SF2 rendered the highest overall classification rates in that 94.8% of the cases were correctly classified, with 96.9% correctly predicted as gifted and 92.7% as non-gifted. SF1 and SF3 were both found to yield the second highest classification rates overall, with 93.2% of the subjects being correctly classified as either gifted or non-gifted. More specifically, SF1 resulted in 99%

of the subjects being correctly predicted as gifted and 87.5% as non-gifted, while SF3 correctly classified 97.9% of the subjects as gifted and 88.5% as non-gifted.

With respect to the cases that were misclassified by the resulting discriminant functions, SF6 rendered the largest number of cases (7.3%) where gifted students were falsely classified as being non-gifted. On the other hand, SF5 was observed to misclassify the greatest number of non-gifted subjects (14.6%) as being gifted. The classification results for the total analysis sample are contained in Table 12.

Classification Results for the Analysis Sample by Gender

The classification analysis for male and female groups indicated that all six short forms performed exceptionally well in terms of forecasting group membership. Although not statistically significant, the female group was consistently found to obtain higher classification rates. Furthermore, it was noted that SF2 yielded the best classification rates for both female (98.9%) and male (91.2%) groups. Classification results for each gender group are presented in Table 13.

Classification Rates for the Analysis Sample by Age

All three age groups appeared to perform extremely well with respect to correctly classifying subjects as either gifted or non-gifted. While the results were not statistically significant, SF1 (96.3%) and SF3 (96.3%) yielded a higher classification rate for ages 6 – 7 than did SF2 (94.4%). The reverse was true for ages 8 – 9 with SF2 (98.4%) surpassing SF1 (95.3%) and SF3 (95.3%). With respect to ages 10 – 12, SF3 (95.9%) was found to classify a greater proportion of subjects as opposed to SF1 (93.2%) and SF2 (91.9%). Table 14 represents the classification results of the analysis sample by age.

Table 12

Classification Results for the Analysis Sample (n = 192)

SF	Classification	Predicted Group Membership		Total Sample
		Non-gifted	Gifted	
1	Non-gifted	84 87.5%	12 12.5%	93.2%
	Gifted	1 1%	95 99%	
2	Non-gifted	89 92.7%	7 7.3%	94.8%
	Gifted	3 3.1%	93 96.9%	
3	Non-gifted	85 88.5%	11 11.5%	93.2%
	Gifted	2 2.1%	94 97.9%	
4	Non-gifted	88 91.7%	8 8.3%	92.7%
	Gifted	6 6.3%	90 93.8%	
5	Non-gifted	82 85.4%	14 14.6%	90.1%
	Gifted	5 5.2%	91 94.8%	
6	Non-gifted	84 87.5%	12 12.5%	90.1%
	Gifted	7 7.3%	89 92.7%	

Table 13

Classification Results for the Analysis Sample by Gender (Male n = 102 and Female n = 90)

SF	Classification	Predicted Group Membership				Total Sample	
		Non-gifted		Gifted		Male	Female
		Male	Female	Male	Female		
1	Non-gifted	42 82.4%	43 95.6%	9 17.6%	2 4.4%	89.2%	97.8%
	Gifted	2 3.9%	0 0%	49 96.1%	45 100%		
2	Non-gifted	43 84.3%	45 100%	8 15.7%	0 0%	91.2%	98.9%
	Gifted	1 2%	1 2.2%	50 98%	44 97.8%		
3	Non-gifted	42 82.4%	43 95.6%	9 17.6%	2 4.4%	90.2%	97.8%
	Gifted	1 2%	0 0%	50 98%	45 100%		
4	Non-gifted	44 86.3%	43 95.6%	7 13.7%	2 4.4%	89.2%	97.8%
	Gifted	4 7.8%	0 0%	47 92.2%	45 100%		
5	Non-gifted	43 84.3%	40 88.9%	8 15.7%	5 11.1%	89.2%	93.3%
	Gifted	3 5.9%	1 2.2%	48 94.1%	44 97.8%		
6	Non-gifted	40 78.4%	43 95.6%	11 21.6%	2 4.4%	83.3%	96.7%
	Gifted	6 11.8%	1 2.2%	45 88.2%	44 97.8%		

Table 14

Classification Results for the Analysis Sample by Chronological Age (6 – 7, n = 54;

8 – 9, n = 64; 10 – 12, n = 74)

SF	Class	Predicted Group Membership						Total Sample		
		Non-gifted			Gifted			6-7	8-9	10-12
		6-7	8-9	10-12	6-7	8-9	10-12			
1	Non-gifted	26	29	33	1	3	4	96.3%	95.3%	93.2%
	Gifted	1	0	1	26	32	36			
2	Non-gifted	26	32	33	1	0	4	94.4%	98.4%	91.9%
	Gifted	2	1	2	25	31	35			
3	Non-gifted	27	30	35	0	2	2	96.3%	95.3%	95.9%
	Gifted	2	1	1	25	31	36			
4	Non-gifted	24	30	34	3	2	3	90.7%	93.8%	93.2%
	Gifted	2	2	2	25	32	35			
5	Non-gifted	23	28	33	4	4	4	90.7%	93.8%	91.9%
	Gifted	1	0	2	26	32	35			
6	Non-gifted	25	28	34	2	4	3	90.7%	90.6%	94.6%
	Gifted	3	2	1	24	30	36			

Linear Equating Technique

In order to make practical use of a short form, the sum of the scaled scores on the abbreviation must be converted into a more meaningful unit. The linear equating technique developed by Tellegen and Briggs (1967) is one method for obtaining short form intelligence quotient (SFIQ) scores. This procedure represents the corrected version of the McNemar (1950) formula and is recommended by Kaufman (1977) when the WISC-III short form is used as a replacement for the original scale. In essence, the corrected formula removes the error variance that is shared between the short and long forms of the WISC-III. Consistent with the Wechsler scales, the linear equating technique yields SFIQ scores that have a mean of 100 and a standard deviation of 15. Appendix B provides the linear equating mathematical formula as outlined by Tellegen and Briggs, in addition to the actual equations that were used for deriving SFIQ scores for the six short forms based on the analysis sample.

Multiple Regression Analysis

While linear equating represents one method for converting short form scores into standardized quantities (i.e., SFIQ), multiple regression analysis can also be used to achieve this objective. Multiple regression provides a linear equation for estimating the value of the criterion based on several predictor variables. Although regression analysis requires additional time to compute, when used with specific populations it has been found to yield SFIQ scores that are slightly closer to FSIQ scores than linear equating (Beck et al., 1983; Donders, 1992; Kennedy & Elder, 1982; Zimet & Adler, 1990). In order for the multiple regression equation to provide the most accurate predictions, however, it is essential that the regression equation be derived from a sample that is

similar to the population on which the formula is to be applied (Vollmerhausen et al., 1986; Zimet & Adler, 1990).

In the present investigation, both linear equating and multiple regression were performed in order to compare the value of both techniques. A standard multiple regression was conducted with the analysis sample in order to derive a formula for estimating FSIQ for each of the six short forms. For this type of analysis, FSIQ was the dependent variable and the WISC-III subtests that comprised each short form were the independent or predictor variables. All data was entered simultaneously for the standard multiple regression. For each of the short forms, beta values were calculated to determine the weighted regression equations so that the estimated FSIQ scores could be obtained. Appendix C provides the multiple regression equations (raw score form) that were used to estimate FSIQ scores for the six short forms based on the analysis sample.

It was noted that all six short forms in the current investigation were found to be significant predictors of FSIQ ($p < .05$). The percentage of variance in the FSIQ that was accounted for by the predictor variables ranged from 92% in SF1 ($F [6, 185] \text{ df} = 364.990$) to 84% in SF6 ($F [4, 187] \text{ df} = 241.958$). Table 15 provides the standard multiple regression summary based on the analysis sample.

Pearson Product-Moment Correlation Coefficients

Pearson product-moment correlation coefficients were computed to determine whether the correlations between SFIQ scores and actual FSIQ scores were positive and significant. Regardless of the procedure used to obtain SFIQ scores (i.e., linear equating or multiple regression), statistically significant correlations were found between the estimated and the actual scores ($p < .01$) for all six short forms. It was concluded that for

Table 15

Standard Multiple Regression Summary for the Analysis Sample (n = 192)

SF	Multiple R	R Squared	Adjusted R Square	p
1	.960	.922	.920	.0001
2	.935	.874	.871	.0001
3	.948	.900	.897	.0001
4	.920	.846	.843	.0001
5	.920	.847	.843	.0001
6	.915	.838	.835	.0001

this sample of subjects, the SFIQ scores obtained from the six short forms were excellent predictors of FSIQ scores.

With respect to SFIQ scores obtained from linear equating, the correlation coefficients for the SFIQ and the FSIQ were found to range from .959 (SF1) to .909 (SF6). Table 16 furnishes a detailed account of the correlation coefficients that were derived from the linear equating technique for the analysis sample. When estimated FSIQ scores were calculated based on regression equations, the correlation coefficients for the SFIQ and the FSIQ ranged from .960 (SF1) to .915 (SF6). Table 17 summarizes all correlation coefficients obtained for the analysis sample utilizing regression equations.

Paired t-Tests

Paired t-tests were performed in order to confirm that no significant differences existed between the mean SFIQ scores and the mean FSIQ scores. The t-test assumes that the researcher is conducting only one test, hence the probability of finding a significant difference by chance alone increases rapidly with the number of tests. To resolve this problem of an inflated Type I error due to multiple testing, more stringent significance levels are required (Tabachnick & Fidell, 1996). In this investigation, the Bonferroni correction for six simultaneous comparisons was applied to all paired t-tests to control for Type I error.

With respect to SFIQ scores derived via linear equating with the analysis sample, one of the six short forms yielded SFIQ scores that were significantly different from the FSIQ scores. In the case of SF1 ($t = -3.202$, $p = .012$), the FSIQ was overestimated by a mean of -1.15 ($SD = 4.99$). The remaining five short forms yielded mean SFIQ scores

Table 16

Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Analysis Sample (n = 192)

	SF1 IQ	SF2 IQ	SF3 IQ	SF4 IQ	SF5 IQ	SF6 IQ
FSIQ	.959*					
FSIQ		.927*				
FSIQ			.946*			
FSIQ				.919*		
FSIQ					.913*	
FSIQ						.909*

* Correlation is significant at the .01 level (two-tailed).

Table 17

Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Analysis Sample (n = 192)

	SF1 IQ	SF2 IQ	SF3 IQ	SF4 IQ	SF5 IQ	SF6 IQ
FSIQ	.960*					
FSIQ		.935*				
FSIQ			.948*			
FSIQ				.920*		
FSIQ					.920*	
FSIQ						.915*

* Correlation is significant at the .01 level (two-tailed).

that were not significantly different from the mean FSIQ scores. Table 18 provides a complete summary of the results from these paired t-tests.

The SFIQ scores obtained from multiple regression were also subjected to paired t-tests. The results indicated that all six short forms yielded SFIQ scores that were not significantly different from the mean FSIQ scores. Moreover, these scores were observed to be within one hundredth of a point of the actual FSIQ scores. Table 19 reports the results from the paired t-tests between FSIQ scores and SFIQ scores obtained from multiple regression.

Although paired t-tests refer to the statistical significance of the results in this investigation, they provide little insight into the practical significance of the differences between the SFIQ and FSIQ mean scores. In order to determine the magnitude of these differences between mean scores, the findings from each short form were first converted into an effect size estimate and then adjusted for paired t-tests (Cohen, 1977; refer to Appendix D for the effect size formula). With respect to interpretation, a smaller effect size is associated with a smaller group difference.

For SFIQ scores derived via linear equating, SF4 (.056) yielded the smallest value while SF1 (-.421) was found to render the largest value. SFIQ scores obtained from multiple regression were also examined for effect size. All six short forms were observed to yield extremely small values (SF5 = .0014, SF1 = .0015, SF3 = .0018, SF6 = -.0024, SF4 = -.0025, and SF2 = .0027). An overview of the effect size values for the analysis sample is provided in Table 20.

Table 18

Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from Linear

Equating for the Analysis Sample (n = 192)

SF	Scores	<u>M</u>	<u>SD</u>	<u>t-test</u>	Mean difference (md)	<u>SD</u> of md	p (Bonferroni correction)
1	FSIQ	114.18	13.48	-3.202	-1.15	4.99	.002
	SFIQ	115.33	16.14				(.012)*
2	FSIQ	114.18	13.48	.977	.426	6.04	.330
	SFIQ	113.75	15.78				
3	FSIQ	114.18	13.48	-2.260	-.831	5.10	.025
	SFIQ	115.01	15.36				(.150)
4	FSIQ	114.18	13.48	.503	.213	5.87	.615
	SFIQ	113.96	14.91				
5	FSIQ	114.18	13.48	1.960	.864	6.10	.051
	SFIQ	113.31	14.92				(.306)
6	FSIQ	114.18	13.48	2.133	.935	6.08	.034
	SFIQ	113.24	14.57				(.204)

*p < .05; df = 191.

Table 19

Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from MultipleRegression for the Analysis Sample (n = 192)

SF	Scores	<u>M</u>	<u>SD</u>	<u>t-test</u>	Mean difference (md)	<u>SD</u> of md	p
1	FSIQ	114.18	13.48	.014	.004	3.76	.989
	SFIQ	114.17	12.94				
2	FSIQ	114.18	13.48	.026	.009	4.78	.979
	SFIQ	114.17	12.60				
3	FSIQ	114.18	13.48	.017	.005	4.27	.986
	SFIQ	114.17	12.78				
4	FSIQ	114.18	13.48	-.022	-.009	5.29	.982
	SFIQ	114.19	12.40				
5	FSIQ	114.18	13.48	.017	.006	5.28	.987
	SFIQ	114.17	12.40				
6	FSIQ	114.18	13.48	-.026	-.010	5.42	.980
	SFIQ	114.19	12.34				

*p < .05; df = 191.

Table 20

Effect Size for Group Differences Based on the Analysis Sample (n = 192)

Technique	Short Form	Effect Size Estimate	Effect Size Adjusted for Paired t-Tests
Linear Equating	1	-.085	-.421
	2	.032	.119
	3	-.062	-.267
	4	.016	.056
	5	.064	.217
	6	.069	.228
Multiple Regression	1	.0003	.0015
	2	.0007	.0027
	3	.0004	.0018
	4	-.0007	-.0025
	5	.0004	.0014
	6	-.0007	-.0024

Psychometric Properties of the Short Forms

Based on the standardization data reported in the WISC-III Canadian manual, it was determined that the reliability coefficients for the six short forms ranged from .922 (SF1) to .885 (SF5). With respect to validity, coefficients were found to range from .922 (SF1) to .876 (SF5 and SF6). Table 21 provides a summary of the reliability and validity coefficients of the six short forms.

Determining the reliability of a short form is necessary in order to calculate the standard error of measurement (SEM) of an instrument. SEM is an estimate of the amount of error associated with a subject's score. The SEM for the short forms in this investigation ranged from 4.19 (SF1) to 5.09 (SF5), as compared to 3.49 for the complete WISC-III battery (Wechsler, 1996). Table 21 also reports the SEM of the six short forms based on the standardization data reported in the WISC-III Canadian manual.

As it is virtually impossible for any test to produce scores that are perfectly reliable, confidence intervals are typically used to aid the interpretation of test results. A smaller standard error is related to a smaller confidence interval (i.e., the range of scores around the obtained score that likely includes the true score), hence a more precise measurement. In the current investigation, a confidence level of 95% was selected. This confidence interval includes the range in which a subject's true score would be 95% of the time. This implies that the statistical chances would be 5 out of 100 that a subject's true score would fall outside this confidence interval.

While the SEM is typically used to determine the confidence intervals for obtained scores (i.e., FSIQ), it is recommended that the standard error of estimate (SE_{est}) be used to compute the confidence intervals for estimated scores (i.e., SFIQ; Glutting,

Table 21

Psychometric Properties of the Six Short Forms Derived from the Standardization Data
Reported in the WISC-III Canadian Manual (Wechsler, 1996)

SF	Reliability	Validity	<u>SEM</u>
1	.922	.922	4.19
2	.905	.889	4.62
3	.921	.906	4.22
4	.918	.880	4.30
5	.885	.876	5.09
6	.904	.876	4.65

McDermott, & Stanley, 1987; Sattler, 2001). The difference between these two procedures is that reliability is used to calculate the $\underline{SEM}$ while the correlation between the SFIQ and the FSIQ scores is required for the $\underline{SE}_{est}$. It is important to note that although the conventional formula for calculating $\underline{SE}_{est}$ is appropriate when regression equations are used, this is not the case for linear equating. According to Kaufman (1976), one of the problems with the linear equating technique is that it does not permit the calculation of $\underline{SE}_{est}$ in the standard way. To ameliorate this situation, Silverstein (1985b) offered a revised $\underline{SE}_{est}$ formula that is suitable for use with the linear equating technique.

The $\underline{SE}_{est}$ values for the analysis sample derived via linear equating and multiple regression were observed to be very similar. Although not statistically significant, the multiple regression technique was found to yield smaller $\underline{SE}_{est}$ values for all six short forms. Table 22 contains a summary of the $\underline{SE}_{est}$ values and confidence intervals for the six short forms based on the analysis sample. Appendix D provides the formulae used to derive reliability and validity coefficients, standard error of measurement, standard error of estimate, and confidence intervals.

Frequency of Underestimated and Overestimated Scores

Each short form was additionally evaluated in terms of the extent to which SFIQ scores either underestimated or overestimated actual FSIQ scores. The number of cases where the SFIQ scores fell above or below their respective 95% confidence intervals, based on the $\underline{SE}_{est}$ of the short form, were tallied in the analysis sample. As a whole, all six short forms were found to yield relatively small percentages of cases falling outside of their respective confidence intervals.

Table 22

Short Form Standard Error of Estimate (SE_{est}) Values and Confidence Intervals By
Technique Based on the Analysis Sample (n = 192)

Technique	Calculation	SF1	SF2	SF3	SF4	SF5	SF6
Linear Equating	SE_{est}	4.30	5.73	4.93	6.04	6.26	6.40
	95%	± 8.4	± 11.2	± 9.7	± 11.8	± 12.3	± 12.5
	90%	± 7.1	± 9.4	± 8.1	± 9.9	± 10.3	± 10.5
Multiple Regression	SE_{est}	4.20	5.32	4.77	5.88	5.88	6.05
	95%	± 8.2	± 10.4	± 9.3	± 11.5	± 11.5	± 11.9
	90%	± 6.9	± 8.7	± 7.8	± 9.6	± 9.6	± 9.9

With respect to SFIQ scores derived via linear equating, SF3 (± 9 confidence interval) was observed to yield the lowest percentage of underestimated cases (0%), while SF2 (± 11 confidence interval) and SF4 (± 11 confidence interval) were both found to render the largest percentage of underestimated cases (3.6%). SF5 (± 12 confidence interval) and SF6 (± 12 confidence interval) each produced the smallest number of overestimated cases (1%), while SF1 (± 8 confidence interval) was found to yield the highest number of overestimated cases (6.3%). Table 23 provides the percentages and the actual number of cases for SFIQ scores falling outside of their respective confidence bands.

The frequency of underestimated and overestimated cases was also calculated based on SFIQ scores that were obtained from multiple regression. Once again, SF3 (± 9 confidence interval) was found to produce the smallest number of underestimated cases (0.5%) as compared to the other short forms. While SF4 (± 11 confidence interval) was observed to yield the largest number of underestimated cases (2.6%) it also rendered the smallest percentage of overestimated cases (0.5%). SF6 (± 11 confidence interval) demonstrated the largest percentage of overestimated cases (2.6%). An overview of the frequency of underestimated and overestimated cases based on SFIQ scores obtained from regression equations is provided in Table 24.

Cross Validation

After the accuracy of the six WISC-III short forms was determined, a cross validation study was conducted in order to strengthen the findings from the analysis sample.

Table 23

Frequency of Linear Equating Derived SFIQ Scores Falling Outside the 95% Confidence Interval for the Analysis Sample (n = 192)

SF	Confidence Interval	Percentage of Underestimates	Percentage of Overestimates
1	± 8	2.6% (5 cases)	6.3% (12 cases)
2	± 11	3.6% (7 cases)	3.1% (6 cases)
3	± 9	0% (0 cases)	5.2% (10 cases)
4	± 11	3.6% (7 cases)	2.1% (4 cases)
5	± 12	3.1% (6 cases)	1% (2 cases)
6	± 12	1% (2 cases)	1% (2 cases)

Table 24

Frequency of Multiple Regression Derived SFIQ Scores Falling Outside the 95%Confidence Interval for the Analysis Sample (n = 192)

SF	Confidence Interval	Percentage of Underestimates	Percentage of Overestimates
1	± 8	1.6% (3 cases)	1% (2 cases)
2	± 10	1.6% (3 cases)	1.6% (3 cases)
3	± 9	0.5% (1 case)	2.1% (4 cases)
4	± 11	2.6% (5 cases)	0.5% (1 case)
5	± 11	2.1% (4 cases)	1.6% (3 cases)
6	± 11	1% (2 cases)	2.6% (5 cases)

Discriminant Analysis Classification Results for the Cross Validation Sample

To examine how well the discriminant functions identified for the six short forms in the initial analysis generalized to a new sample of cases, a classification analysis was performed with the cross validation sample. As with the analysis sample, all six short forms were observed to perform exceptionally well when applied to the cross validation sample. Although not statistically significant, SF1 and SF3 were both observed to render the highest overall classification rates (97.7%). More specifically, SF1 correctly classified 98.1% of the subjects as gifted and 97.4% as non-gifted, while SF3 correctly predicted 97.7% as gifted and 97.7% as non-gifted. The total classification rates for the remaining short forms ranged from 96.6% to 95.7%.

Overall, a small percentage of the subjects in the cross validation sample were misclassified by the discriminant function. SF5 was observed to yield the highest number of cases (4.3%) where gifted subjects were falsely identified as being non-gifted. The highest incidences of non-gifted subjects classified erroneously as being gifted were found in SF6 (6.4%) and SF2 (6.2%). Table 25 summarizes the cross validation sample classification rates for subjects in the gifted and non-gifted groups.

Pearson Product-Moment Correlation Coefficients

The subtest scaled scores in the cross validation sample were converted into SFIQ scores based on linear equating and multiple regression equations obtained from the analysis sample. With respect to SFIQ scores derived from linear equating, the correlation coefficients for the SFIQ and the FSIQ in the cross validation sample ranged from .975 (SF1) to .941 (SF5). When SFIQ scores were obtained via multiple regression for the cross validation sample, coefficients were observed to range from .975 (SF1) to

Table 25

Classification Results for the Cross Validation Sample (n = 1,058)

SF	Classification	Predicted Group Membership		Total Sample
		Non-gifted	Gifted	
1	Non-gifted	515 97.4%	14 2.6%	97.7%
	Gifted	10 1.9%	519 98.1%	
2	Non-gifted	496 93.8%	33 6.2%	95.9%
	Gifted	10 1.9%	519 98.1%	
3	Non-gifted	517 97.7%	12 2.3%	97.7%
	Gifted	12 2.3%	517 97.7%	
4	Non-gifted	501 94.7%	28 5.3%	96.6%
	Gifted	8 1.5%	521 98.5%	
5	Non-gifted	511 96.6%	18 3.4%	96.1%
	Gifted	23 4.3%	506 95.7%	
6	Non-gifted	495 93.6%	34 6.4%	95.7%
	Gifted	11 2.1%	518 97.6%	

.948 (SF5). Overall, both sets of correlation coefficients were found to be very similar to that of the analysis sample. The correlation coefficients that were derived from linear equating and multiple regression for the cross validation sample are provided in Tables 26 and 27.

Paired t-Tests

With respect to SFIQ scores derived via linear equating with the cross validation sample, only SF3 ($t = -.640$, $p = .522$) yielded SFIQ scores that did not differ significantly from the mean FSIQ scores ($M = -.092$, $SD = 4.69$). The remaining five short forms all produced statistically different mean SFIQ scores as compared to the mean FSIQ scores. Table 28 contains a complete summary of the results from these paired t-tests for the cross validation sample.

The SFIQ scores obtained from multiple regression were also subjected to paired t-tests with the cross validation sample. The results indicated that only SF3 and SF5 yielded mean SFIQ scores that were not significantly different from the mean FSIQ scores. It was noted that SF3 ($t = .424$, $p = .672$) estimated the FSIQ scores with near accurate precision ($M = .062$, $SD = 4.78$), while SF5 ($t = -1.41$, $p = .158$) also yielded estimated FSIQ scores that were found to closely approximate the FSIQ scores ($M = -.252$, $SD = 5.80$). The means derived from SF1, SF2, SF4, and SF6 were all found to be significantly different from the mean FSIQ scores. Table 29 reports the results from the paired t-tests between FSIQ scores and SFIQ scores obtained from multiple regression with the cross validation sample.

Table 26

Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Linear Equating for the Cross Validation Sample (n = 1,058)

	SF1 IQ	SF2 IQ	SF3 IQ	SF4 IQ	SF5 IQ	SF6 IQ
FSIQ	.975*					
FSIQ		.958*				
FSIQ			.966*			
FSIQ				.950*		
FSIQ					.941*	
FSIQ						.949*

* Correlation is significant at the .01 level (two-tailed).

Table 27

Pearson Product-Moment Correlations Between FSIQ Scores and SFIQ Scores Obtained from Multiple Regression for the Cross Validation Sample (n = 1,058)

	SF1 IQ	SF2 IQ	SF3 IQ	SF4 IQ	SF5 IQ	SF6 IQ
FSIQ	.975*					
FSIQ		.961*				
FSIQ			.967*			
FSIQ				.950*		
FSIQ					.948*	
FSIQ						.953*

* Correlation is significant at the .01 level (two-tailed).

Table 28

Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from LinearEquating for the Cross Validation Sample (n = 1,058)

SF	Scores	<u>M</u>	<u>SD</u>	<u>t-test</u>	Mean difference (md)	<u>SD</u> of md	p (Bonferroni correction)
1	FSIQ	113.59	17.60	-10.10	-1.45	4.67	.0001
	SFIQ	115.04	19.70				(.0006)*
2	FSIQ	113.59	17.60	-9.59	-1.66	5.62	.0001
	SFIQ	115.25	19.26				(.0006)*
3	FSIQ	113.59	17.60	-.640	-.092	4.69	.522
	SFIQ	113.69	18.20				
4	FSIQ	113.59	17.60	-6.72	-1.17	5.67	.0001
	SFIQ	114.77	18.02				(.0006)*
5	FSIQ	113.59	17.60	10.77	2.02	6.09	.0001
	SFIQ	111.58	17.82				(.0006)*
6	FSIQ	113.59	17.60	-3.42	-.590	5.61	.001
	SFIQ	114.18	17.69				(.006)*

*p < .05; df = 1057.

Table 29

Paired t-Tests Between Mean FSIQ Scores and SFIQ Scores Obtained from MultipleRegression for the Cross Validation Sample (n = 1,058)

SF	Scores	<u>M</u>	<u>SD</u>	<u>t-test</u>	Mean difference (md)	<u>SD</u> of md	p (Bonferroni correction)
1	FSIQ	113.59	17.60	-5.40	-.674	4.06	.0001
	SFIQ	114.27	15.99				(.0006)*
2	FSIQ	113.59	17.60	-13.34	-2.06	5.03	.0001
	SFIQ	115.66	15.71				(.0006)*
3	FSIQ	113.59	17.60	.424	.062	4.78	.672
	SFIQ	113.53	15.43				
4	FSIQ	113.59	17.60	-7.86	-1.39	5.74	.0001
	SFIQ	114.98	15.08				(.0006)*
5	FSIQ	113.59	17.60	-1.41	-.252	5.80	.158
	SFIQ	113.85	15.16				
6	FSIQ	113.59	17.60	-11.95	-2.03	5.53	.0001
	SFIQ	115.63	15.34				(.0006)*

*p < .05; df = 1057.

As a follow up to the paired t-tests, calculations were also performed in order to compare the effect size of the six short forms for the cross validation sample. Based on linear equating derived SFIQ mean scores, SF3 rendered the smallest effect size value (-.027) while SF1 (-.519), SF5 (.473), and SF2 (-.459) demonstrated the largest values. With respect to SFIQ scores obtained via multiple regression, SF3 (.022) again emerged as having the smallest effect size value while SF2 (-.594) and SF6 (-.530) were found to yield the largest values. Table 30 illustrates the effect size values for group differences based on the cross validation sample.

Frequency of Underestimated and Overestimated Scores

The cross validation sample was also evaluated in terms of the extent to which SFIQ scores either underestimated or overestimated actual FSIQ scores. Using the SE_{est} values for each short form, the number of cases where the SFIQ scores fell outside of their respective 95% confidence interval were tallied. Overall, each of the six short forms rendered relatively few occurrences of underestimated or overestimated scores.

With respect to SFIQ scores that were derived via linear equating, SF4 (± 11 confidence interval) was found to underestimate the smallest percentage of cases (1.2%) falling below the 95% level of confidence. Although SF5 (± 12 confidence interval) demonstrated the largest number of underestimated cases (4.3%), it was also found to render the smallest number of overestimated cases (0.9%). SF1 (± 8 confidence interval) was observed to yield the largest percentage of overestimated cases (7.0%). Table 31 outlines the percentages and the actual number of cases in the cross validation sample where linear equating derived SFIQ scores fell outside of their respective confidence bands.

Table 30

Effect Size for Group Differences Based on the Cross Validation Sample (n = 1,058)

Technique	Short Form	Effect Size Estimate	Effect Size Adjusted for Paired t-tests
Linear Equating	1	-.082	-.519
	2	-.094	-.459
	3	-.005	-.027
	4	-.066	-.295
	5	.115	.473
	6	-.034	-.150
Multiple Regression	1	-.038	-.241
	2	-.117	-.594
	3	.004	.022
	4	-.079	-.353
	5	-.014	-.061
	6	-.115	-.530

Table 31

Frequency of Linear Equating Derived SFIQ Scores Falling Outside the 95% Confidence

Interval for the Cross Validation Sample (n = 1,058)

SF	Confidence Interval	Percentage of Underestimates	Percentage of Overestimates
1	± 8	1.3% (14 cases)	7.0% (74 cases)
2	± 11	1.5% (16 cases)	2.6% (28 cases)
3	± 9	2.5% (26 cases)	2.4% (25 cases)
4	± 11	1.2% (13 cases)	4.2% (44 cases)
5	± 12	4.3% (46 cases)	0.9% (9 cases)
6	± 12	1.6% (17 cases)	2.2% (23 cases)

An examination of SFIQ scores that were obtained from multiple regression indicated that SF2 (± 10 confidence interval) and SF6 (± 11 confidence interval) yielded the smallest percentage of underestimated cases (1.0% each). SF5 (± 11 confidence interval) was observed to render the largest number of underestimated cases (2.8%). SF3 (± 9 confidence interval) was found to produce the fewest number of overestimated cases (1.6%), while SF6 yielded that greatest number of overestimated cases (4.0%). Table 32 provides an overview of the frequency of underestimated and overestimated cases based on SFIQ scores obtained from regression equations with the cross validation sample.

Summary

All six short forms investigated in this study were effective in discriminating between gifted and non-gifted elementary students in Canada. This finding is not unexpected, given that the short forms for this research were chosen based on their level of success using American normative data. Although small differences were observed with respect to the classification rates for each short form, these differences were not statistically significant. Gender and chronological age did not effect the accuracy of the short forms in discriminating between gifted and non-gifted students. Depending on the technique used to obtain SFIQ scores, coupled with the size of the sample, it was found that certain short forms yielded statistically significant results with respect to the mean differences between SFIQ scores and FSIQ scores.

Table 32

Frequency of Multiple Regression Derived SFIQ Scores Falling Outside the 95%

Confidence Interval for the Cross Validation Sample (n = 1,058)

SF	Confidence Interval	Percentage of Underestimates	Percentage of Overestimates
1	± 8	1.7% (18 cases)	2.6% (28 cases)
2	± 10	1.0% (11 cases)	3.5% (37 cases)
3	± 9	2.4% (25 cases)	1.6% (17 cases)
4	± 11	1.5% (16 cases)	3.7% (39 case)
5	± 11	2.8% (30 cases)	2.8% (30 cases)
6	± 11	1.0% (11 cases)	4.0% (42 cases)

CHAPTER FIVE: DISCUSSION

Summary

Discriminant Analysis

This research study investigated the accuracy of six short forms of the WISC-III in estimating the Full Scale IQ of potentially gifted elementary school students in Canada. In order to accomplish this task, predictive discriminant analysis was performed with each of the short forms to determine how well they differentiated the gifted and non-gifted subjects in the analysis sample. Overall, the discriminant function separated the gifted and non-gifted groups with a high degree of accuracy for all short forms examined in this investigation.

The discriminant function for each of the short forms produced large canonical correlations that were found to be statistically significant. An examination of the discriminate weights and loadings of the six short forms revealed that Similarities and Block Design were the two subtests that consistently contributed the most to group separation. Vocabulary was also found to be an excellent predictor for distinguishing between the two groups in three of the four short forms. Picture Arrangement was observed to be an excellent predictor of group membership in the one short form that employed this subtest. Arithmetic and Coding were generally found to be fair discriminators, while Picture Completion and Information contributed less in terms of discriminating between the gifted and non-gifted groups.

In order to explore whether gender might effect the utility of the short form in differentiating between gifted and non-gifted students, six different discriminant analyses were conducted for male and female sample groups. It was found that male and female

groups produced similar canonical correlations as those of the total analysis sample. Although females yielded higher correlations as compared to males, these differences were not statistically significant. Consequently, the same short form may be used with confidence for both male and female students. Overall, these findings are consistent with previous research which has suggested the presence of only modest gender differences (Cravens, 1999; Jensen & Reynolds, 1982; Kaufman et al., 1988; Thompson et al., 1986).

A review of the literature failed to uncover any existing short form research comparing the performance of students across the ages of 6 through 12. Nevertheless, given that 6 and 7-year-old children are administered a different form for the Coding subtest as compared to older children, coupled with the fact that three of the six short forms included the Coding subtest, it was decided that chronological age merited further investigation. Discriminant analyses was conducted for each of the six short forms according to three age groups (i.e., 6 – 7, 8 – 9, 10 – 12). It was found that the canonical correlations for the 10 - 12 group were virtually identical to the total sample, while the 6 – 7 group had lower correlations and the 8 – 9 group was had higher correlations. Nevertheless, these differences were not found to be statistically significant. Overall, the results of the three age groups were very similar to the total analysis sample and did not support the need to employ a different short form based on the chronological age of the subject.

With respect to accurately classifying subjects as being either gifted or non-gifted, the percentage of subjects that would have been sorted correctly utilizing a chance model was 50% (i.e., $1 \div \text{number of groups}$). According to Hair et al. (1998), the discriminant function should classify at least 25% more subjects than what would have been classified

due to chance alone. In the current investigation, all six short forms exceeded the proportional chance criterion for 75%, hence they were deemed to be more efficient than chance in distinguishing between the gifted and non-gifted groups. Moreover, the discriminant functions identified in the initial analysis generalized well to the new cases in the large cross validation sample. In fact, the cross validation sample classification rates for the six short forms were higher overall than those from the analysis sample.

Although not statistically significant, SF2 (94.8% overall) was found to classify the largest proportion of gifted (96.9%) and non-gifted (92.7%) subjects in the analysis sample. SF1 correctly classified the greatest number of gifted subjects (99%), but it also misclassified a large portion of the non-gifted subjects as being gifted (12.5%). SF3 also correctly identified a high number of gifted subjects (97.9%), but misclassified 11.5% of non-gifted subjects. By way of comparison, SF2 demonstrated the lowest number of misclassifications (7.3%).

In the cross validation sample, SF1 (97.7%) and SF3 (97.7%) were found to provide the best overall classification results for both gifted (98.1% and 97.7%, respectively) and non-gifted (97.4% and 97.7%, respectively) subjects. Nevertheless, these results were not found to be statistically significant. Although SF4 emerged as being the best classifier of gifted subjects (98.5%), it also misclassified 5.3% of the non-gifted subjects. In contrast, SF1 and SF3 misclassified non-gifted subjects as being gifted at a rate of 2.6% and 2.3%, respectively. While SF2 accurately classified a large proportion of the gifted subjects (98.1%), it also misclassified 6.2% of non-gifted subjects.

Validity of the Short Forms

In order to further compare the utility of the six short forms, the validity of each abbreviation was also evaluated. Based upon the standardization data reported in the WISC-III Canadian manual (Wechsler, 1996), short form validity coefficients were calculated. All six short forms yielded validity coefficients that were found to exceed .80, the common accepted standard (Sattler, 2001). SF1 (.922) and SF3 (.906) were observed to demonstrate the highest validity coefficients, while SF5 (.876) and SF6 (.876) had the lowest.

With respect to additional short form validation guidelines, the criteria set forth by Resnick and Entin (1971) were also considered. The first of these criteria was met as the Pearson product-moment correlations between SFIQ and FSIQ scores for the analysis sample were observed to be both positive and significant. All correlation coefficients were found to be above .90, thus indicating a strong relationship between the SFIQ scores and the actual FSIQ scores. A comparison of the correlations obtained via linear equating (ranging from .959 for SF1 to .909 for SF6) and multiple regression (ranging from .960 for SF1 to .915 for SF6) suggested that the regression technique yielded marginally higher correlations overall. With respect to the cross validation sample, correlation coefficients between SFIQ and FSIQ scores were found to be similar, albeit slightly larger, than the correlations obtained by the analysis sample.

In terms of the second Resnick-Entin criterion, differences between the mean SFIQ and FSIQ scores were small and non-significant for all six short forms when scores were obtained via multiple regression for the analysis sample. When SFIQ scores were derived via linear equating, however, SF1 was observed to yield statistically significant mean

differences compared to the actual FSIQ scores. Based on SFIQ scores obtained from multiple regression in the cross validation sample, the differences between mean SFIQ and FSIQ scores were found to be non-significant for SF3 and SF5. When the linear equating procedure was employed, only SF3 yielded non-significant differences. Finally, the effect size was also calculated in order to rule out the possibility that significant scores might have been the result of trivial differences that were emphasized by the large sample size. SF3 was observed to yield the smallest effect size value when the cross validation sample was examined. This implies that all things being equal, the difference between the mean SFIQ and FSIQ scores were smaller in SF3 than they were in any of the other short forms. Consequently, the significant differences observed between the mean SFIQ and FSIQ scores of the other short forms were real differences that cannot be attributed to the large sample size.

The findings from this research study indicate that the performance of the short form varies with the technique used to obtain the SFIQ scores. For example, five short forms in the analysis sample and one short form in the cross validation sample had non-significant results with respect to the mean differences between SFIQ scores derived via linear equating and actual FSIQ scores. When multiple regression was used to obtain SFIQ scores, all six short forms in the analysis sample and two short forms in the cross validation sample had non-significant scores. Regardless of the method used to obtain SFIQ scores, SF3 was the only abbreviation that consistently demonstrated no statistically significant differences between estimated SFIQ scores and actual FSIQ scores across both samples.

The third Resnick-Entin criterion states that there must be a high degree of correspondence between SFIQ and FSIQ classification labels. Owing to the strong likelihood that errors in classifying a subject's level of intelligence can occur even if the SFIQ score is within one point of the FSIQ (e.g., 119 vs. 120), this criterion was not evaluated in the present study. The decision to reconsider this criterion is well supported in the literature (Bersoff, 1971; Kaufman, 1990; Silverstein, 1990a).

The number of cases where SFIQ scores either underestimated or overestimated FSIQ scores were calculated as an alternative to comparing the correspondence of classification labels between the short form and the full form (Cravens, 1999). Based on the SE_{est} of the short form, confidence intervals (95% level) were developed around each SFIQ score. FSIQ scores that fell outside of the band of scores surrounding the SFIQ were considered to be either an underestimate or an overestimate. In this study, underestimated scores were deemed to be more important than overestimated scores in order to minimize the possibility of screening out any gifted students.

All six short forms were observed to have low rates of underestimation (ranging from 0% for SF3 to 4.3% for SF5) and overestimation (ranging from 0.5% for SF3 to 7.0% for SF1). This implies that there is a strong likelihood that the confidence interval surrounding the SFIQ will encompass the actual FSIQ score.

Reliability of the Short Forms

Reliability of the abbreviation was also considered when evaluating the six short forms. Short form reliability coefficients were calculated based on the standardization data reported in the WISC-III Canadian manual (Wechsler, 1996). As a reliability coefficient greater than .90 is generally considered to be acceptable (Sattler, 2001), this

guideline was utilized in the current investigation. Not surprisingly, short forms that were comprised of a larger number of subtests (e.g., SF1 and SF3) yielded higher reliability coefficients (.922 and .921, respectively) and lower SEM values as compared to the short forms made up of fewer subtests. Although five of the six short forms were found to have reliability coefficients greater than .90, SF5 fell below this level with a reliability coefficient of .885. This finding suggests that one would have less confidence in the reliability of SF5 as compared to the other short forms.

Clinical Value of the Short Forms

Given that the psychometric properties of the six short forms are generally very similar, the clinical value of each abbreviation (e.g., the breadth and depth of information yielded from the abbreviation, the amount of time required to administer the short form) was also considered.

Information Yielded from the Short Forms. The Cattell-Horn-Carroll (CHC) theory of intelligence, the most comprehensive and empirically supported model of cognitive ability currently available (Flanagan et al., 2000; McGrew, 1997; McGrew & Flanagan, 1998), was used as a benchmark for determining the extent to which each short form measured a variety of mental processes. The complete WISC-III battery is capable of measuring five of the eight broad CHC cognitive areas (i.e., crystallized intelligence, visual processing, quantitative ability, short-term memory, and processing speed). It was noted that SF1 and SF5 each assessed four of the eight broad areas (i.e., crystallized intelligence, visual processing, quantitative ability, and processing speed). SF2, SF6 (i.e., crystallized intelligence, visual processing ability, and quantitative ability), and SF3 (i.e., crystallized intelligence, visual processing ability, and processing speed) each measured

three of the broad CHC cognitive areas. SF4 was found to be somewhat limited in its range as it only tapped into two broad CHC areas (i.e., crystallized intelligence and visual processing ability).

The inclusion of Similarities or Block Design in a short form represents an excellent choice as both subtests are strong measures of “g” and also load highly on their respective indices (i.e., Verbal Comprehension and Perceptual Organization; Sattler, 2001). Moreover, these subtests are excellent measures of their respective broad cognitive areas (i.e., crystallized intelligence and visual processing ability). One or both of these subtests were present in all six short forms that were investigated. The inclusion of Vocabulary in a short form (SF1, SF2, SF3, and SF4) also appears to be a good choice as this subtest is the best measure of “g” on the entire WISC-III battery (Sattler, 2001). Furthermore, Vocabulary contributes substantially to the Verbal Comprehension Index and it represents an excellent measure of crystallized cognitive ability. While the combination of Similarities and Vocabulary in SF1 may seem redundant because they both represent strong measures of crystallized intelligence, the Similarities subtest taps into language development while Vocabulary is a measure of lexical knowledge. Within the current investigation, Similarities, Block Design, and Vocabulary yielded the highest discriminant weights and loadings, thus supporting their ability to discriminate between gifted and non-gifted subjects. The apparent value of these three subtests comes at a high cost, however, as they require substantially more time to administer and score as compared to other subtests.

The Information subtest was included in both SF3 and SF4. This subtest is a strong measure of “g” and it loads highly on the Verbal Comprehension Index (Sattler,

2001). Furthermore, the nature of the Information subtest makes it relatively quick to administer and score. As this subtest taps into experience and school-based knowledge, however, it may place students from impoverished backgrounds at a disadvantage. With respect to the discriminant weights and loadings of Information in this study, this subtest was found to be a good predictor of group membership. One may question the utility of combining Information and Vocabulary in SF3 and SF4, as both subtests are strong measures of crystallized intelligence within the CHC model. Nevertheless, Information assesses verbal information while Vocabulary measures lexical knowledge.

The inclusion of Arithmetic in SF1, SF2, SF5, and SF6 was not unexpected given that this subtest is quick to administer and score. It is also a fair measure of “g” and it contributes moderately to the Freedom from Distractibility Index (Sattler, 2001). Given that Arithmetic tends to be more representative of an achievement task as opposed to a reasoning task, coupled with its potential for being affected by distractibility on the part of the subject, it may not provide as much clinical information as other subtests (Kline, 2001). Nevertheless, within CHC theory, Arithmetic is a strong measure of quantitative ability and for this reason it may yield potentially useful information. The discriminant weights and loadings for Arithmetic within the current investigation revealed that it added substantially to the discrimination process in SF5 and SF6, while it appeared to contribute to a lesser degree in SF1 and SF2.

The Picture Completion subtest was used in SF1, SF3, SF4, SF5, and SF6, likely due to its relative ease of administration and scoring. This subtest is considered to be a fair measure of “g” and to contribute moderately to the Perceptual Organization Index (Sattler, 2001). Based on CHC theory, Picture Completion seems to offer only modest

information in terms of visual processing ability (i.e., visualization). Moreover, the discriminant weights and loading associated with Picture Completion in the current investigation suggest that it offered less than other subtests in terms of discriminating between the gifted and non-gifted groups.

The inclusion of Picture Arrangement in SF2 represents a task that requires substantially more time to administer as compared to other subtests, but one that may offer additional insight into nonverbal reasoning involving planning and sequencing. It can also provide an opportunity to gather further information regarding the student's awareness of social situations. This subtest, which is considered to be a fair measure of "g", contributes only minimally to the Perceptual Organization Index (Sattler, 2001). Within CHC theory, Picture Arrangement has been found to be only a moderate measure of visual processing ability. Notwithstanding, the discriminant weight and loading demonstrated by Picture Arrangement in the current investigation suggested that this subtest was an excellent predictor in terms of discriminating between gifted and non-gifted subjects. SF2 is composed of Block Design and Picture Arrangement which both measure the same broad cognitive area (i.e., visual processing ability). Nevertheless, Block Design is a strong measure of spatial relationships while Picture Arrangement is a moderate test of visualization.

The Coding subtest is quick to administer and score, thus making it a good candidate for use in a short form. In the present study, Coding is used in SF1, SF3, and SF5. Although Coding is thought to be poor measure of "g", it contributes substantially to the Processing Speed Index (Sattler, 2001). Moreover, it is viewed as being a strong indicator of processing speed within CHC theory. There is evidence to suggest that the

faster an individual's processing speed, the greater his or her intellectual power (Fry & Hale, 2000; Kail, 2000; Kail & Salthouse, 1994). More specifically, faster processing speed results in a better working memory that in turn leads to enhanced fluid reasoning. According to K. S. McGrew (personal communication, September 5, 2001), processing speed has recently been found to be one of the most important contributors to academic learning for children between the ages of six through 12. He added that after the age of 12, other abilities within the CHC model of intelligence (e.g., short term memory) seem to become more important.

Despite the apparent value of including the Coding subtest in a short form, there are some that question its use (Kaufman, 1994; Wechsler, 1991). Kaufman (1994) suggested that Symbol Search (a supplemental subtest) should be routinely substituted for Coding when using the WISC-III battery. Based on an inspection of the mean scores for the WISC-III subtests in the analysis sample of this study, however, very little difference was detected between Coding and Symbol Search for the gifted and non-gifted subjects. With respect to the cross validation sample, gifted and non-gifted subjects were both found to perform approximately one scaled score higher on Symbol Search as compared to Coding. There is also WISC-R research available which suggests that gifted students tend to demonstrate lower scores on the Coding subtest as compared to other subtests (Karnes & Brown, 1980). Nevertheless, an examination of the WISC-III mean scores for the gifted subjects in the analysis sample of this study indicated that Digit Span rendered the lowest mean score while the mean score for Coding was similar to other subtests (e.g., Information, Arithmetic, Similarities, Symbol Search, Picture Completion, Picture Arrangement, Comprehension). With respect to the cross validation sample, gifted

students were also found to perform the lowest on Digit Span, while the mean score for Coding was comparable to other subtests (e.g., Object Assembly, Picture Completion). Finally Sattler (2001) speculated that Coding A (for ages 6 to 7) and Coding B (for ages 8 to 16) might involve separate information processing modes. Nevertheless, the discriminant weights and loadings from the current investigation found Coding to have good (SF1 and SF3) to excellent (SF5) discriminating ability in terms of separating the gifted and non-gifted groups, regardless of chronological age.

Short Form Administration Time. The clinical value of a short form was also evaluated in terms of the length of time required to administer the instrument. Using the work of Terminie (1997) to estimate the amount of time necessary to administer the WISC-III subtests to gifted elementary students, it would seem that SF5 (17 minutes), SF4 (20 minutes) and SF6 (21 minutes) may be somewhat more desirable as they require slightly less time as compared to SF2 (24 minutes), SF3 (24 minutes), and SF1 (28 minutes). Despite these minor differences, even the most time consuming short form in this study is estimated to require less than half the amount of time to administer as compared to the complete WISC-III test battery. Moreover, short forms that require more time to administer typically include a greater number of subtests that will likely add to the breadth and depth of information rendered. Clearly, a balance must be struck between the amount of information obtained from the short form and the amount of time required to administer the instrument.

Best Short Form for the Population in the Current Investigation

Based on an evaluation of the discriminating power, validity, reliability, and clinical value of the six short forms, it is concluded that SF3 represents the best

compromise between all factors. This finding supports the research of Mark et al. (1998) which reported SF3 to be appropriate for determining gifted eligibility for Canadian elementary students.

With respect to the effectiveness of the discriminant function of SF3 in predicting group membership, 93.2% of the total analysis sample was accurately classified. More specifically, 97.9% of the gifted subjects and 88.5% of the non-gifted subjects were correctly identified. When the results of the discriminant function were applied to the larger cross validation sample, however, it was found that SF3 performed even better. Of the 1,058 total subjects, 97.7% were accurately classified overall with equal percentages (97.7%) of gifted and non-gifted subjects being correctly identified.

The reliability (.921) and validity (.906) coefficients for SF3 were observed to far exceed the minimum required values. Moreover, these findings were very similar to the reliability (.93) and validity (.89) coefficients obtained in the original SF3 research (Dumont & Faro, 1993). Based on regression equations, statistical analysis with SF3 resulted in correlation coefficients of .948 (analysis sample) and .967 (cross validation sample). These findings indicated a strong relationship between SFIQ scores obtained by SF3 and the actual FSIQ scores. Furthermore, the difference between the mean SFIQ and FSIQ scores were found to be small and non-significant. It is important to note that this non-significant difference was observed both with the analysis sample and the cross validation sample. SF5 was the only other abbreviation in the study to demonstrate this characteristic. Due to the lower reliability of SF5 (.885), however, it was deemed to be a less useful short form. Finally, it was noted that SF3 demonstrated a relatively low

number of occurrences where SFIQ scores underestimated or overestimated actual FSIQ scores.

In terms of clinical value, SF3 measures three broad areas of intelligence according to the CHC model. This short form includes Vocabulary, which is the best measure of “g” on the entire WISC-III battery and is also a strong measure of crystallized intelligence (i.e., lexical knowledge). The Information subtest also symbolizes a strong measure of “g”, while at the same time providing an excellent measure of crystallized intelligence (i.e., general information). Block Design provides the strongest measure of “g” amongst the Performance subtests and is also considered to be an excellent measure of visual processing ability (i.e., spatial relations). Picture Completion is thought to be a fair measure of “g” and a moderate representation of visual processing ability (i.e., visualization). Although the Coding subtest renders a poor measure of “g”, it provides a strong measure of the broad cognitive area of processing speed. Based upon the discriminant weights and loadings observed in the current investigation, Vocabulary and Block Design were the subtests in SF3 that contributed most to discriminating between subjects in the gifted and non-gifted groups. Information was also found to be a good predictor of group membership, while Coding and Picture Completion seemed to provide the smallest amount of discriminating power.

Given that SF3 is estimated to require 24 minutes to administer, this short form has the potential to cut the administration time of the WISC-III in half without sacrificing a significant amount of reliability or validity. Although three of the other short forms (i.e., SF4, SF5, SF6) are estimated to require slightly less time to administer, the breadth and depth of information provided by SF3 is considered to be more valuable than

additional time saved. In sum, the findings of this study indicate that SF3 can be used as a screening instrument to identify potentially gifted elementary students in Canada.

Linear Equating vs. Multiple Regression

A comparison of the accuracy of SFIQ scores derived via linear equating and those obtained by multiple regression was also considered in this investigation. An examination of the mean differences between SFIQ scores and FSIQ scores in the analysis sample indicated that multiple regression consistently yielded SFIQ scores that were closer to the actual scores than those of linear equating. Although not statistically significant, SFIQ scores were observed to be within one hundredth of a point of the actual FSIQ scores when multiple regression was used. Furthermore, multiple regression uniformly demonstrated a smaller range of variance in the SFIQ scores as compared to linear equating. In the cross validation sample, smaller mean differences were also observed between FSIQ scores and SFIQ scores obtained via multiple regression in SF1, SF3, and SF5. On the other hand, smaller mean differences were noted between FSIQ scores and linear equating derived SFIQ scores in SF2, SF4, and SF6. These results were not statistically significant. Based on the gifted population in the current investigation, it was concluded that either linear equating or multiple regression could be used to obtain SFIQ scores. This finding is consistent with previous research which has suggested that either technique may be used with equal accuracy for high ability students (Dumont & Faro, 1993).

Although developing regression equations tends to be somewhat more difficult to compute, this technique has been previously demonstrated to be more precise than linear equating when used with specific populations (Beck et al., 1983; Donders, 1992;

Kennedy & Elder, 1982; Vollmerhausen et al., 1986; Zimet & Adler, 1990). Given that the linear equating technique considers the intercorrelations between the subtests in the short form, coupled with the reliabilities of these subtests, it has been found to be more stable when the data from one population is applied to data from a different population (Beck et al, 1983; Sattler, 2001; Silverstein, 1987). This was not an issue in the present investigation, however, as the derived regression equations are intended to be applied to a similar population (e.g., potentially gifted students).

Subtest Scatter

As the presence of subtest scatter on a short form may be indicative of cognitive strengths or deficits, the SFIQ score might not be an accurate reflection of the student's actual ability. Consequently, test protocols must be scrutinized for scatter in order to determine whether the administration of the remaining WISC-III subtests is warranted. With respect to the practical application of the findings in this study, LoBello's (1991) recommendation of subtest scatter greater than 6 points (i.e., more than two standard deviations on the subtest scale) was used as a benchmark for determining when the remaining subtests should be administered.

Practical Application of the Findings from the Current Investigation

Calculating SFIQ Scores. In order to calculate SFIQ scores, the student's raw scores on the subtests in the short form must first be transformed into scaled scores according to the appropriate age group in the WISC-III Canadian manual. If the difference between the lowest and highest subtest scale scores is greater than six points, the remaining WISC-III subtests should be administered to determine the student's FSIQ. In the absence of significant subtest scatter, the scaled scores are inserted into the

regression equation for SF3 (refer to Appendix C). The outcome of the regression equation, rounded to the nearest whole number, is the student's SFIQ score.

Accepting the Student as Being Gifted. The optimal cutoff score is the criterion against which each student's SFIQ score is judged to determine whether he or she should be accepted as being gifted without the need to administer the remaining WISC-III subtests. Based on the data in the current investigation, the cutoff score for SF3 was set at 130 (i.e., Very Superior range). This score was chosen as it seemed to provide the best compromise between effectiveness and efficiency. According to Pegnato and Birch (1959), the effectiveness of a gifted screening instrument is the ratio of students identified by the instrument to the total number of gifted students, while efficiency is the ratio of the number of students identified by the screening instrument to the number of students referred.

When a SFIQ cutoff score of 130 was applied to the large cross validation sample, a total of 190 students were found to have a score of 130 or above. It was noted that 169 of these cases had actual FSIQ scores on the WISC-III of 130 or above. Of the remaining 21 cases where the FSIQ was below 130, it was noted that 20 fell within the 95% confidence interval for the FSIQ (i.e., 130 ± 7 based on the SEM of the Full Scale IQ). Given the band of error, these cases were viewed as being gifted rather than non-gifted. Based on this reasoning, the use of SF3 resulted in the accurate identification of 189 (99.5%) students as being gifted. The SFIQ cutoff score of 130 resulted in the misclassification of only one non-gifted student (i.e., FSIQ below 130 ± 7) as being gifted (0.5%). With this degree of accuracy, there would be no need to administer the remaining WISC-III subtests to a student who receives a SFIQ score of 130 or above.

Based on a total of 241 students in the cross validation sample that had a FSIQ score of 130 or above, SF3 was successful in identifying 189 gifted students. This implies that SF3 was 78.4% effective in identifying gifted students during the first phase of screening. It should be noted that although a lower cutoff score would have identified more gifted students, this would have also increased the proportion of non-gifted students that would have been falsely identified as gifted. By setting the SFIQ cutoff score at this conservative level, greater accuracy was ensured.

Administering the Remaining WISC-III Subtests. Based on the SE_{est} for SF3, a range of scores between 130 ± 9 (95% confidence level) was calculated. This range was used as a guideline for identifying gifted students who might be overlooked due to receiving a SFIQ score within the “grey area” below the 130 cutoff. Consequently, students who receive SFIQ scores between 121 to 129 move to the second phase of screening whereby the remaining subtests of the WISC-III are administered in order to determine the actual FSIQ score. When applied to the cross validation sample, 255 students were identified who required additional testing. Of these students, 244 were found to have an actual FSIQ of at least 130 ± 7 , thus confirming the value of further assessment. Hence, the efficiency rate for SF3 was found to be 87.8%. Moreover, only 31 students (12.2%) that were recommended for further testing based on SFIQ scores within the 121 to 129 range were found to be non-gifted when the full battery was administered.

Further investigation found that an even greater number of gifted students would have eventually been identified from additional testing if the band of scores that constituted the “grey area” was extended downward by two additional points thus ranging

from 119 to 129. With respect to the cross validation sample, this meant that an extra 50 students were identified as requiring additional testing. Of these students, 22 were found to have an actual FSIQ of at least 130 ± 7 , thus confirming the value of expanding the lower limit of the range by two points. As a result of this change, the efficiency rate for SF3 was found to be reduced only slightly (83.7%).

Discontinuing Further Assessment. At the third phase of screening, a student who receives a SFIQ score of 118 or less on SF3 should be excluded from further testing as it is highly unlikely that the actual FSIQ would be 130 ± 7 . An examination of the cross validation sample indicated that based on a total of 456 cases where a FSIQ of at least 130 ± 7 was possible, there were 33 gifted students that were found to have a SFIQ of 118 or less. Upon further examination, however, it was noted that 22 of these cases contained subtest scatter greater than six points. As this amount of scatter would have lead to the administration of the remaining WISC-III subtests, these gifted students would have presumably been identified. Eleven gifted students would not have been detected due to the decision to discontinue further assessment based on a SFIQ score of 118 or less. This error represented only 2.4% of the total cases.

Summary. Based on the findings of this study, the following procedure is recommended. At the first phase of screening, SF3 should be administered to all elementary students referred for gifted eligibility purposes. If the subtest scatter on the short form is greater than six points, the administration of the remaining WISC-III subtests is warranted. Students with SFIQ scores of 130 and above should be accepted as gifted without the need to administer the remaining subtests. Students who demonstrate SFIQ scores between 119 to 129 move into the second phase of screening whereby the

remaining WISC-III subtests are administered in order to determine the actual FSIQ scores. At the third phase of screening, SFIQ scores of 118 or less result in the discontinuation of the assessment process and the student is considered to be non-gifted. The flowchart in Figure 1 depicts the practical application of SF3 for elementary students who are referred for giftedness.

Implications for School Psychology Practice

The six WISC-III short forms studied in the current investigation appear to be valid when the purpose of assessment is to provide an estimate of FSIQ for gifted eligibility purposes. If a potentially gifted elementary student is referred for intellectual assessment based on nomination, achievement test results, or class grades, a WISC-III short form could provide a global estimate of the student's cognitive ability. Given that the short form takes less than half the time to administer as the complete battery, a great deal of time may be saved. School psychologists may use the time saved to conduct more in-depth assessments in other areas (e.g., behavioural, creativity, etc.) or to assess additional students.

Despite the apparent utility of short forms, a great deal of clinical information may be lost when the full WISC-III battery is abbreviated. Furthermore, school psychologists must be aware of factors that would indicate the need to pursue a more thorough assessment based upon short form results (e.g., subtest scatter, a significantly low score on a subtest of particular clinical relevance, suspicion of problems aroused by the student's behaviour during testing, situational factors, etc.). Clearly, school psychologists need to use sound clinical judgment in deciding when further assessment of a student's cognitive functioning is warranted.

Figure 1. Flowchart depicting the practical application of the Dumont-Faro short form for elementary students who are referred for giftedness.

Limitations of the Current Investigation

A limitation of this research study is that the SFIQ scores were extracted from the full WISC-III battery as opposed to the short form being administered separately.

Although rescoring WISC-III protocols successfully avoids the complication of retesting effects, it also makes it difficult to interpret the psychometric properties of the abbreviation as it is not independent from the full form.

This study was also limited in that the cross validation sample was drawn from a single geographic region in Canada. Moreover, the focus of the investigation was restricted to elementary children as opposed to all school-aged students.

Directions for Future Research

Further research is needed in order to confirm whether the separate administration of the six short forms in the current investigation yield the same results as those obtained from rescored WISC-III protocols. To accomplish this task, counterbalanced administrations of the WISC-III full battery and the short form would be beneficial.

As the current investigation was limited to gifted elementary students between the ages of 6 to 12, WISC-III short form research conducted with a similar population of students older than the age of 12 would be valuable. Although there is some research available which examines the time required to administer each of the WISC-III subtests, there does not appear to be studies measuring the amount of time required to set-up and score each subtest. This type of information would be helpful when evaluating the actual amount of time saved from the use of a short form. Finally, the development of standardized tables for converting the scaled score sums of the short form into SFIQ scores would serve to further reduce the amount of time required to use the abbreviation.

Conclusion

This research study explored the accuracy of six short forms of the WISC-III in estimating the Full Scale IQ of potentially gifted elementary students in Canada. The results of the discriminant analyses found that all six short forms demonstrated a high degree of accuracy in separating the gifted students from the non-gifted students. When applied to the cross validation sample, all short forms produced impressive classification rates. It was also discovered that gender and chronological age of the subject did not effect the accuracy of the short forms investigated. With respect to the technique used for determining short form IQ scores, both linear equating and multiple regression were found to be equally effective with the gifted students in this study.

When psychometric soundness and clinical utility were considered together with discriminating power, the Dumont-Faro short form (SF3) emerged as providing the best compromise between all factors in terms of the gifted population in the current investigation. Based on discriminant analysis conducted with the cross validation sample, the Dumont-Faro abbreviation had an overall classification rate of 97.7%. With respect to the psychometric properties of this short form, it demonstrated strong reliability (.921) and validity (.906). Moreover, short form IQ scores obtained from the Dumont-Faro abbreviation were found to correlate highly (.967) with the full WISC-III battery. Based upon the 95% confidence interval for this short form, there were relatively few cases where the short form IQ scores underestimated or overestimated the actual Full Scale IQ scores. In terms of clinical value, this short form provides strong measures of three broad areas of intelligence (i.e., crystallized intelligence, visual processing ability, processing speed) based on the Cattell-Horn-Carroll theory of intelligence. Finally, the

Dumont-Faro short form was estimated to require only 24 minutes to give, hence reducing the typical WISC-III administration time by at least half.

The practical application of the Dumont-Faro was also explored. At the first phase of screening, a cutoff score of 130 correctly identified 99.5% of the students in the cross validation as being gifted. With this level of accuracy there would be little need to administer the remaining WISC-III subtests. At the second phase of screening, short form IQ scores ranging between 119 to 129 were used as a benchmark for deciding when the remaining WISC-III subtests should be administered. When applied to the cross validation sample, approximately 84% of the students were identified as being gifted based on the administration of the remaining subtests. At the third phase of screening, students with a short form IQ score of 118 or less were excluded from further assessment and classified as non-gifted. Based on the cross validation sample, this cutoff score signified a 2.4% risk of classifying a gifted student as being non-gifted.

In conclusion, the Dumont-Faro short form of the WISC-III may be used as an alternative to full battery assessment for screening potentially gifted elementary students in Canada. The use of this short form could result in a substantial saving of school psychology time and resources, without sacrificing accuracy. This saved time may be invested into other school psychology activities, such as pre-referral interventions, consultations, social-emotional assessments, or additional cognitive assessments. Despite the apparent value of any short form, school psychologists must be aware of the limitations of their use and be prepared to administer the remaining subtests if there is any reason to suspect that further assessment is warranted. Finally, although a short form should not be used as the sole criterion for determining a student's eligibility for gifted

programming, it may be used in conjunction with other selection criteria (i.e., teacher and parent nomination, achievement tests, school grades, portfolios, etc.) for screening potentially gifted students. With these reservations in mind, the Dumont-Faro short form of the WISC-III can be used as a screening instrument for potentially gifted elementary students in Canada.

References

- American Psychological Association. (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Assouline, S. G. (1997). Assessment of gifted children. In N. Colangelo & G. A. Davis (Eds.), Handbook of gifted education (2nd ed., pp. 89-108). Needham Heights, MA: Allyn & Bacon.
- Assouline, S. G., & Lupkowski-Shoplik, A. (1997). Talent searches: A model for the discovery and development of academic talent. In N. Colangelo & G. A. Davis (Eds.), Handbook of gifted education (2nd ed., pp. 170-179). Needham Heights, MA: Allyn & Bacon.
- Bashaw, W. L., & Anderson, H. E. (1967). A correction for replicated error in correlation coefficients. Psychometrika, 32, 435-441.
- Beal, A. L, Dumont, R., Branche, A. H., & Cruse, C. L. (1996). Validation of a WISC-III short form for Canadian Students with learning disabilities. Canadian Journal of School Psychology, 12, 1-6.
- Beal, A. L., Dumont, R., Cruse, C. L., & Branche, A. H. (1996). Practical implications of differences between the American and Canadian norms for WISC-III and a short form for children with learning disabilities. Canadian Journal of School Psychology, 12, 7-14.

Beck, N. C., Ray, J. S., Seidenberg, M., Young, T. R., & Gamache, M. P. (1983). Development and cross-validation of a short form for the WISC-R. Journal of Consulting and Clinical Psychology, 51, 864-869.

Bersoff, D. N. (1971). Short forms of individual intelligence tests for children: Review and critique. Journal of School Psychology, 9, 310-320.

Boatman, T. A., Davis, K. G., & Benbow, C. P. (1995). Best practices in gifted education. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology (3rd ed., pp. 1083-1095). Washington, DC: The National Association of School Psychologists.

Borland J. H., & Wright, L. (1994). Identifying young, potentially gifted, economically disadvantaged students. Gifted Child Quarterly, 38, 164-171.

Braden, J. P. (1995). Review of the Wechsler Intelligence Scale for Children-Third Edition. In J. C. Conoley & J. C. Impavra (eds.), The twelfth mental measurements yearbook (pp. 1098-1103). Lincoln, NE: Buros Institute.

Browne, J. L. (1995). Validity of the K-BIT and a WISC-III short form with a sample of adolescent young offenders. Unpublished master's thesis, Lakehead University, Thunder Bay, Ontario, Canada.

Calgary Board of Education. (1999). The GATE program (Available from the Calgary Board of Education in Calgary, Canada).

Campbell, J. M. (1998). Internal and external validity of seven Wechsler Intelligence Scale for Children-Third Edition short forms in a sample of psychiatric inpatients. Psychological Assessment, 10, 431-434.

Carleton, F. O., & Stacey, C. L. (1954). Evaluation of selected short forms of the Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 10, 258-261.

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

Clarizio, H., & Veres, V. (1984). A short-form version of the WISC-R for the learning disabled. Psychology in the Schools, 21, 154-157.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cole, S. N., Williams, R. L., Nix, A. P., & Litaker, R. G. (1967). Validity of an abbreviated WISC scale for retarded and borderline children. Psychological Reports, 21, 571-572.

Coleman, L. J. (1994). Portfolio assessment: A key to identifying hidden talents and empowering teachers of young children. Gifted Child Quarterly, 38, 65-69.

Comninel, M. E., & Bordieri, J. E. (2001). Estimating WISC-III scores for special education students using the Dumont-Faro short form. Psychology in the Schools, 38, 11-16.

Connery, S., Katz, D., Kaufman, A. S., & Kaufman, N. L. (1996). Correlations between two short cognitive tests and a WISC-III short form using a sample of adolescent inpatients. Psychological Reports, 78, 1373-1378.

Cravens, T. B. (1999). Evaluation of Kaufman, Ishikuma, and Kaufman-Packer's short forms for use with psychiatric African American children. Journal of Psychoeducational Assessment, 17, 332-342.

Cummings, S. B., MacPhee, H. M., & Wright, H. F. (1946). A rapid method of estimating the IQs of subnormal white adults. Journal of Psychology, 21, 81-89.

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

Dean, R. S. (1977). The validity and reliability of abbreviated versions of the WISC-R. Educational and Psychological Measurement, 37, 1111-1116.

DiPasquale, G., Rosehart, D., Kempa, L., Rooney, C., & Steinman, A. (1992). Short form tests for gifted screening: Comparison of the WISC-R and the Stanford-Binet: Fourth Edition. Canadian Journal of School Psychology, 8, 120-125.

Dirks, J., Wessels, K., Quarfoth, J., & Quenon, B. (1980). Can short-form WISC-R IQ tests identify children with high full scale IQ? Psychology in the Schools, 17, 40-46.

Doll, E. A. (1917). A brief Binet-Simon Scale. Psychological Clinic, 11, 197-211, 254-261.

Donders, J. (1992). Validity of two short forms of the WISC-R in children with traumatic brain injury. Journal of Clinical Psychology, 48, 364-370.

Donders, J. (1997). A short form of the WISC-III for clinical use. Psychological Assessment, 9, 15-20.

Donders, J., & Warschausky, S. (1996). Validity of a short-form of the WISC-III in children with traumatic head injury. Child Neuropsychology, 2, 227-232.

Doppelt, J. E. (1956). Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores of four subtests. Journal of Consulting Psychology, 20, 63-66.

Dumont, R., & Faro, C. (1993). A WISC-III short form for learning-disabled students. Psychology in the Schools, 30, 212-219.

Edwards, R., & Edwards, J. L. (1993). The WISC-III: A practitioner perspective. Journal of Psychoeducational Assessment Monographs-Advances in Psychological Assessment: WISC-III, 144-150.

Elman, L., Blixt, S., & Sawicki, R. (1981). The development of cutoff scores on the WISC-R in the multidimensional assessment of gifted children. Psychology in the Schools, 18, 426-428.

Enburg, R., Rowley, V. N., & Stone, B. (1961). Short forms of the WISC for use with emotionally disturbed children. Journal of Clinical Psychology, 17, 280-284.

Erikson, R. V. (1967). Abbreviated form of the WISC: A reevaluation. Journal of Consulting Psychology, 31, 641.

Feldhusen, J. F. (1998). Identification and assessment of talented learners. In J. VanTassel-Baska (Ed.), Excellence in educating gifted and talented learners (3rd ed., pp. 193-210). Denver, CO: Love Publishing.

Feldhusen, J. F., & Jarwan, F. A. (1993). Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Monks, & A. H. Passow (Eds.), International handbook of research and development of giftedness and talent (pp. 233-251). Tarrytown, NY: Pergamon Press.

Fell, L., & Schmidt-Fell, S. (1982). Effectiveness of WISC-R short-forms in screening gifted children. Psychological Reports, 51, 1017-1018.

Finch, A.,J., Childress, W. B., & Ollendick, T. H. (1973). Comparison of separately administered and abstracted WISC short forms with the full scale WISC. American Journal of Mental Deficiency, 77, 755-756.

Finch, A. J., Childress, W. B., Wilkins, K. A., & Kendall, P. C. (1974). WISC short forms with emotionally disturbed children. Journal of Abnormal Child Psychology, 2, 337-341.

Finch, A. J., Kendall, P. C., Spirito, A., Entin, A., Montgomery, L. E., & Schwartz, D. J. (1979). Short form and factor-analytic studies of the WISC-R with behavior problem children. Journal of Abnormal Child Psychology, 7, 337-344.

Finch, A. J., Ollendick, T. H., & Ginn, F. W. (1973). WISC short forms with mentally retarded children. American Journal of Mental Deficiency, 78, 144-149.

Fineman, C. A., & Carran, D. T. (1986). An epidemiological approach to screening gifted students utilizing WISC-R subtests. Psychology in the Schools, 23, 142-147.

Finley, C. J., & Thompson, J. M. (1958). An abbreviated Wechsler Intelligence Scale for Children for use with educable mentally retarded. American Journal of Mental Deficiency, 63, 473-480.

Fishkin, A. S., Kampsnider, J. J., & Pack, L. (1996). Exploring the WISC-III as a measure of giftedness. Roeper Review, 18, 226-231.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). The Wechsler intelligence scales and CHC theory: A contemporary approach to interpretation. Boston: Allyn & Bacon.

- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. Psychological Bulletin, 95, 29-51.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101, 171-191.
- Fry, A. F., & Hale, S. (2000). Relationship among processing speed, working memory, and fluid reasoning. Biological Psychology, 54, 1-34.
- Gallagher, J. J. (1996). A critique of critiques of gifted education. Journal for the Education of the Gifted, 19, 234-249.
- Gayton, W. F., Wilson, W. T., & Bernstein, S. (1970). An evaluation of an abbreviated form of the WISC. Journal of Clinical Psychology, 26, 466-468.
- Geil, G. A. (1945). A clinically useful abbreviated Wechsler-Bellevue Scale. Journal of Psychology, 20, 101-108.
- Glascoe, F. P. (1996). Can the Brigance Screens detect children who are gifted and academically talented? Roeper Review, 19, 20-24.
- Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. Educational and Psychological Measurement, 47, 607-614.
- Goh, D. S. (1979). Empirical versus random item selection in the design of intelligence test short forms-the WISC-R example. Applied Psychological Measurement, 3, 75-82.
- Gross, M. U. M. (1992). The early development of three profoundly gifted children of IQ 2000. In P. S. Klein & A. J. Tannenbaum (Eds.), To be young and gifted (pp. 94-140). Norwood, NJ: Ablex Publishing.

- Gross, M. U. M. (1999). Small poppies: Highly gifted children in the early years. Roeper Review, 21, 207-214.
- Hadaway, N., & Marek-Schroer, M. F. (1997). Multidimensional assessment of the gifted minority student. Roeper Review, 15, 73-77.
- Hagen, E. (1980). Identification of the gifted. New York: Teachers College.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, A. (1998). Multivariate Data Analysis (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Harrison, P. L. (1997). Editor's Comments. School Psychology Review, 26, 143-145.
- Haynes, J. P. (1982). Validity of a WISC-R short form with male delinquents. Journal of Clinical Psychology, 38, 406-408.
- Haynes, J. P. (1985). Comparative validity of two WAIS-R short forms with clients of low IQ. Journal of Clinical Psychology, 41, 282-284.
- Haynes, J. P., & Bensch, M. (1982). Validity of a WISC-R short form for female delinquents. Criminal Justice and Behavior, 9, 115-119.
- Herrera-Graf, M., Dipert, Z. J., & Hinton, R. N. (1996). Exploring the effective use of the vocabulary/block design short form with a special school population. Educational and Psychological Measurement, 56, 522-528.
- Hilden, A. H., Taylor, J. W., & Dubois, P. H. (1952). Empirical evaluation of short W-B Scales. Journal of Clinical Psychology, 8, 321-331.
- Hobby, K. L. (1982). WISC-R split-half short form manual. Los Angeles, CA: Western Psychological Services.

Howley, C. B., Howley, A., & Pendarvis, E. D. (1995). Out of our minds: Anti-intellectualism and talent development in American schooling. New York: Teachers College Press.

Huberty, C. J. (1994). Applied discriminant analysis. New York: John Wiley & Sons.

Jensen, A. R., & Reynolds, C. R. (1982). Race, Social class and ability patterns on the WISC-R. Personality and Individual Differences, 3, 423-438.

Jin, S., & Feldhusen, J. F. (2000). Parent identification of the talents of gifted students. Gifted Education International, 14, 230-236.

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. Journal of School Psychology, 38, 51-61.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. Acta Psychologica, 86, 199-225.

Karnes, F. A., & Brown, K. E. (1980). Factor analysis of the WISC-R for the gifted. Journal of Educational Psychology, 72, 197-199.

Karnes, F. A., & Brown, K. E. (1981). A short form of the WISC-R for gifted students. Psychology in the Schools, 18, 169-173.

Kaufman, A. S. (1972). A short form of the Wechsler Preschool and Primary Scale of Intelligence. Journal of Consulting and Clinical Psychology, 39, 361-369.

Kaufman, A. S. (1976). A four-test short form of the WISC-R. Contemporary Educational Psychology, 1, 180-196.

Kaufman, A. S. (1977). Should short-form validity coefficients be corrected? Journal of Consulting and Clinical Psychology, 45, 1159-1161.

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

Kaufman, A. S. (1992). Evaluation of the WISC-III and WPPSI-R for gifted children. Roeper Review, 14, 154-158.

Kaufman, A. S. (1994). Intelligent testing with the WISC-III. New York: Wiley.

Kaufman, A. S., & Harrison, P. L. (1986). Intelligence tests and gifted assessment: What are the positives? Roeper Review, 8, 154-159.

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

Kaufman, A. S., Kaufman, J. C., Balgopal, R., & McLean, J. E. (1996). Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. Journal of Clinical Child Psychology, 25, 97-105.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAS-R subtests. Journal of Clinical Psychology, 44, 231-248.

Kennedy, L. P., & Elder, S. T. (1982). WISC-R: An abbreviated version. Journal of Clinical Psychology, 38, 174-178.

Killan, J. B., & Hughes, L. C. (1978). A comparison of short forms of the Wechsler Intelligence Scale for Children-Revised in the screening of gifted referrals. Gifted Child Quarterly, 22, 111-115.

King, L. A. & King, D. W. (1982). Wechsler short forms: A brief status report. Psychology in the Schools, 19, 433-438.

Kirschenbaum, R. J. (1998). Dynamic assessment and its use with underserved gifted and talented populations. Gifted Child Quarterly, 42, 140-147.

Klausmeier, K., Mishra, S. P., & Maker, C. J. (1987). Identification of gifted learners: A national survey of assessment practices and training needs of school psychologists. Gifted Child Quarterly, 31, 135-137.

Kline, R. B. (2001). Brief cognitive assessment of children: Review of instruments and recommendations for best practice. In J. J. W. Andrews, D. H. Saklofske, & H. L. Janzen (Eds.), Handbook of psychoeducational assessment (pp. 103-132). San Diego, CA: Academic Press.

Kluever, R. C., & Green, K. E. (1990). Identification of gifted children: A comparison of the scores on Stanford-Binet 4th Edition and Form LM. Roeper Review, 13, 16-20.

Kramer, J. H. (1993). Interpretation of individual subtest scores on the WISC-III. Psychological Assessment, 5, 193-196.

Kramer, J. J., Shanks, K., Markley, R. P., & Ryabik, J. E. (1983). The seductive nature of WISC-R short forms: An analysis with gifted referrals. Psychology in the Schools, 20, 137-141.

Kriegman, G., & Hansen, F. W. (1947). VIBS: A short form of the Wechsler-Bellevue Intelligence Scale. Journal of Clinical Psychology, 3, 209-216.

Levy, P. (1967). The correction for spurious correlation in the evaluation of short form tests. Journal of Clinical Psychology, 23, 84-86.

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

- Lichtenstein, R. (1998). How long does a psychoeducational evaluation take? An urban Connecticut study. Professional Psychology: Research and Practice, 29, 144-148.
- Linn, M., & Lopatin, E. (1990). A simultaneous screening/assessment procedure for identifying the gifted student. Psychology in the Schools, 27, 303-309.
- Linville, J. N., Rust, J. O., & Kim, J. K. (1999). The information and picture completion dyad of the WISC-III as a screening test for gifted referrals. Journal of Instructional Psychology, 26, 98-104.
- LoBello, S. G. (1991). Subtest scatter as an indicator of the inaccuracy of short-form estimates of IQ. Psychological Reports, 68, 1115-1118.
- Louis, B. Subotnik, R. F., Breland, P. S., & Lewis, M. (2000). Establishing criteria for high ability versus selective admission to gifted programs: Implications for policy and practice. Educational Psychology Review, 12, 295-314.
- Lupart, J. L. & Pyryt, M. C. (1996). "Hidden gifted" students: Underachiever prevalence and profile. Journal for the Education of the Gifted, 20, 36-53.
- Lustberg, R. S., Motta, R., & Naccari, N. (1990). A model using the WISC-R to predict success in programs for gifted students. Psychology in the Schools, 27, 126-130.
- Maker, C. J. (1996). Identification of gifted minority students: A national problem, needed changes and a promising solution. Gifted Child Quarterly, 40, 41-50.
- Mantzicopoulos, P. Y. (2000). Can the Brigance K&1 Screen detect cognitive/academic giftedness when used with preschoolers from economically disadvantaged backgrounds? Roeper Review, 22, 185-191.

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.

Mark, R., Beal, A. L., & Dumont, R. (1998). Validation of a WISC-III short-form for the identification of Canadian gifted students. Canadian Journal of School Psychology, 14, 1-10.

Matarazzo, J. D. (1992). Psychological testing and assessment in the 21st century. American Psychologist, 47, 1007-1018.

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

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a process comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (151-180). New York: Guilford.

McGrew, K. S., & Flanagan, D. P. (1998). The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment. Boston: Allyn & Bacon.

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

McNemar, Q. (1974). Correction to a correction. Journal of Consulting and Clinical Psychology, 42, 145-146.

McNemar, Q. (1975). McNemar replies. Journal of Consulting and clinical Psychology, 43, 424-425.

Mogel, S. & Satz, P. (1963). Abbreviation of the WAIS for clinical use: An attempt at validation. Journal of Clinical Psychology, 19, 298-300.

Mumpower, D. L. (1955). The fallacy of the short form. Journal of Clinical Psychology, 11, 275-277.

Ortiz, V. Z., & Gonzalez, A. (1989). Validation of a short form of the WISC-R with accelerated and gifted Hispanic students. Gifted Child Quarterly, 33, 152-155.

Patchett, R. F., & Stansfield, M. (1992). Subtest scatter on the WISC-R with children of superior intelligence. Psychology in the Schools, 29, 5-11.

Pegnato, C. & Birch, J. (1959). Locating gifted children in junior high school: A comparison of methods. Exceptional Children, 48, 340 – 367.

Phelps, L., & Rosso, M. (1986). Evaluation of the Kennedy-Elder Abbreviated IQ formula with emotionally disturbed adolescents. Journal of Clinical Psychology, 42, 979-982.

Piirto, J. (1999). Talented children and adults: Their development and education (2nd ed.). Columbus, OH: Merrill.

Piotrowski, R. J. (1976). The effect of omitting a limited number of subtests on the full scale reliability of the WISC-R. Psychology in the Schools, 13, 298-301.

Prifitera, A., Weiss, J. G., & Saklofske, D. H. (1998). The WISC-III in context. In A. Prifitera & D. Saklofske (Eds.), WISC-III clinical use and interpretation: Scientist-practitioner perspectives (pp. 1-36). San Diego, CA: Academic Press.

Pyryt, M. C. (1996). IQ: Easy to bash, hard to replace. Roeper Review, 18, 255-258.

Pyryt, M. C. & Heck, R. H. (1991). Multivariate techniques for gifted education. In N. K. Buchanan & J. F. Feldhusen (Eds.), Conducting research and evaluation in gifted education (pp. 201-219). New York: Teachers College Press.

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

Rasbury, W. C., Falgout, J. C., & Perry, N. W. (1978). A Yudin-type short form of the WISC-R: Two aspects of validation. Journal of Clinical Psychology, 34, 120-126.

Reid, W. B., Moore, D., & Alexander, D. (1968). Abbreviated form of the WISC for use with brain-damaged and mentally retarded children. Journal of Consulting and Clinical Psychology, 32, 236.

Reschly, D. J. (1997). Diagnostic and treatment utility of intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp.437-456). New York: The Guilford Press.

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.

Reynolds, C. R., Willson, V. L., Clark, P. L. (1983). A 4-test short form of the WAIS-R for clinical screening. Clinical Neuropsychology, 5, 111-116.

Robinson, N. M., & Robinson, H. (1992). The use of standardized tests with young gifted children. In P. S. Klein & A. J. Tannenbaum (Eds.), To be young and gifted (pp. 141-170). Norwood, NJ.: Ablex Publishing.

Runco, M. (1997). Is every child gifted? Roeper Review, 19, 220-224.

Ryan, J. J. (1981). Clinical utility of a WISC-R short form. Journal of Clinical Psychology, 37, 389-391.

Ryan, J. J., & Rosenberg, S. J. (1984). Administration time estimates for WAIS-R subtests and short forms in a clinical sample. Journal of Psychoeducational Assessment, 2, 125-129.

Sabatino, D. A., Spangler, R. S., & Vance, H. B. (1995). The relationship between the Wechsler Intelligence Scale for Children-Revised and the Wechsler Intelligence Scale for Children-III scales and subtests for gifted children. Psychology in the Schools, 32, 18-23.

Saccuzzo, D. P., & Johnson, N. E. (1995). Traditional psychometric tests and proportionate representation: An intervention and program evaluation study. Psychological Assessment, 7, 183-194.

Sandoval, J. (1995). Review of the WISC-III. In J. C. Conoley & J. J. Kramer (Eds.), The tenth mental measurements yearbook (pp. 1102-1106). Lincoln, NE: Buros Institute.

Sarouphim, K. M. (1999). DISCOVER: A promising alternative assessment for the identification for gifted minorities. Gifted Child Quarterly, 43, 244-251.

Sattler, J. M. (1992). Assessment of children (3rd ed.). San Diego, CA: Jerome M. Sattler Publisher.

Sattler, J. M. (2001). Assessment of children: Cognitive applications (4th ed.). San Diego, CA: Jerome M. Sattler Publisher.

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

Satz, P. & Mogel, S. (1967). An abbreviation of the WISC-R for clinical use. Journal of Consulting Psychology, 31, 321-324.

Satz, P., Van de Riet, H., & Mogel, S. (1967). An abbreviation of the WISC for clinical use. Journal of Consulting Psychology, 31, 108.

Schwartz, L., & Levitt, E. E. (1960). Short-forms of the Wechsler Intelligence Scale for Children in the educable, non-institutionalized, mentally retarded. Journal of Educational Psychology, 51, 187-190.

Sevier, R. C., Bain, S. K., & Hildman, L. K. (1994). Comparison of WISC-R and WISC-III for gifted students. Roeper Review, 17, 39-42.

Shore, B. M., Cornell, D. G., Robinson, A., & Wards, V. S. (1991). Recommended practices in gifted education: A critical analysis. New York: Teachers College Press.

Silverman, L. K. (1998). Through the lens of giftedness. Roeper Review, 20, 204-210.

Silverman, L. K., & Kearney, L. (1992). The case for the Stanford-Binet LM as a supplemental test. Roeper Review, 15, 34-37.

Silverstein, A. B. (1967a). Validity of WISC short forms at three age levels. Journal of Consulting Psychology, 31, 635-636.

Silverstein, A. B. (1967b). A short short form of the WISC and WAIS for screening purposes. Psychological Reports, 20, 682.

Silverstein, A. B. (1968). Validity of a new approach to the design of WAIS, WISC, and WPPSI short forms. Journal of Consulting and Clinical Psychology, 32, 478-479.

Silverstein, A. B. (1970). Reappraisal of the validity of WAIS, WISC, and WPPSI short forms. Journal of Consulting and Clinical Psychology, 34, 12-14.

Silverstein, A. B. (1971). A corrected formula for assessing the validity of WAIS, WISC, and WPPSI short forms. Journal of Clinical Psychology, 27, 212-213.

Silverstein, A. B. (1975). A reply to McNemar. Journal of Consulting and Clinical Psychology, 43, 423-425.

Silverstein, A. B. (1982). Two and four subtest short forms of the WAIS-R. 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. (1989a). Reliability and abnormality of scaled-score ranged. Journal of Clinical Psychology, 45, 926-929.

Silverstein, A. B. (1989b). Agreement between a short form and the full scale as a function of the correlation between them. Journal of Clinical Psychology, 45, 929-931.

Silverstein, A. B. (1990a). Short forms of individual intelligence tests. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 3-11.

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

- Simpson, W. H., & Bridges, C. C. (1959). A short form of the Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 15, 424.
- Spangler, R. S., & Sabatino, D. A. (1995). Temporal stability of gifted children's intelligence. Roeper Review, 17, 207-210.
- Sparrow, S. S., & Gurland, S. T. (1998). Assessment of gifted children with the WISC-III. In A. Prifitera & D. Saklofske (Eds.), WISC-III clinical use and interpretation: Scientist-practitioner perspectives (pp. 59-72). San Diego, CA: Academic Press.
- Stanley, J. C. (1997). Varieties of intellectual talent. The Journal of Creative Behavior, 31, 93-119.
- Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.
- Sternberg, R. J. (1993). Rocky's back again: A review of the WISC-III. Journal of Psychoeducational Assessment Monographs-Advances in Psychological Assessment: WISC-III, 161-164.
- Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.
- Tellegen, A. & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. Journal of Consulting Psychology, 31, 499-506.
- Terminie, T. J. (1998). Identifying high IQ students using short forms of the Wechsler Intelligence Scale for Children-Third Edition (Doctoral dissertation, Mississippi State University). Dissertation Abstracts International, 58, 3963.

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.

Thompson, A. P. (1992). Subtest scatter is not an indicator of inaccuracy for short-form estimates of IQ. Psychological Reports, 70, 889-890.

Thompson, A. P., Howard, D., & Anderson, J. (1986). Two and four subtest short forms of the WAIS-R: Validity in a psychiatric sample. Canadian Journal of Behavioural Science, 18, 287-293.

Thompson, J. M., & Finley, C. J. (1963). An abbreviated WISC for use with gifted elementary school children. California Journal of Educational Research, 14, 167-177.

Tyler-Wood, T., & Carri, L. (1991). Identification of gifted children: The effectiveness of various measures of cognitive ability. Roeper Review, 14, 63-64.

Volker, M. A., Guarnaccia, V., & Scardapane, J. R. (1999). Short forms of the Stanford-Binet Intelligence Scale: Fourth Edition for screening potentially gifted preschoolers. Journal of Psychoeducational Assessment, 17, 226-235.

Vollmerhausen, S. Elder, S. T., & Clark, P. (1986). Comparison of two abbreviated models of the WISC-R. Journal of Clinical Psychology, 42, 347-353.

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. (1958). The measurement and appraisal of adult intelligence. Baltimore: Williams & Wilkins.

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children-Third Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1996). Manual: Canadian Supplement to the Wechsler Intelligence Scale-Third Edition. Toronto, Canada: The Psychological Corporation.

Weiss, L. G., Saklofske, D. H., Prifitera, A., Chen, H., & Hildebrand, D. (1999). Calculation of the WISC-III General Ability Index Using Canadian Norms. Canadian Journal of School Psychology, 14, 1-9.

Wolfson, W., & Bachelis, L. (1960). An abbreviated form of the WAIS Verbal Scale. Journal of Clinical Psychology, 16, 421.

Yalowitz, J. M., & Armstrong, R. G. (1955). Validity of short forms of the Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 11, 275-277.

Yudin, L. W. (1966). An abbreviated form of the WISC for use with emotionally disturbed children. Journal of Consulting Psychology, 30, 272-275.

Zimet, S. G., & Adler, S. S. (1990). Methodological and clinical issues in studies of the performance of emotionally disturbed children and adolescents on abbreviated forms of the WISC-R: A review. Journal of School Psychology, 28, 133-146.

Zimet, S. G., Farley, G. K., & Dahlem, N. W. (1985). An abbreviated form of the WISC-R for use with emotionally disturbed children. Psychology in the Schools, 22, 19-22.

Zorman, R. (1997). Eureka: The cross-cultural model for identification of hidden talent through enrichment. Roeper Review, 20, 54-61.

Zytowski, D. G., & Hudson, J. (1965). The validity of split-half abbreviations of the WAIS. Journal of Clinical Psychology, 21, 292-294.

Appendix A

WISC-III Subtest Administration Times for Gifted Elementary Students (Terminie, 1998)

Subtest	Mean Administration Time in Seconds	Standard Deviation
Information	196	44.63
Coding	201	10.57
Similarities	216	44.00
Vocabulary	225	64.24
Arithmetic	258	86.49
Comprehension	296	85.03
Picture Completion	311	66.54
Picture Arrangement	448	126.57
Block Design	450	149.31
Object Assembly	636	89.39

Appendix B

Linear Equating

Linear Equating Formula for Computing SFIQ Scores (Tellegen & Briggs, 1967):

$$\text{SFIQ} = \left(\frac{15}{S_c} \right) (X_c - \overline{X_c}) + 100$$

Where: $S_c = S_s \sqrt{n + 2 \sum r_{jk}}$ (the standard deviation of the short form score)

S_s = the subtest standard deviation (which has the uniform value of 3)

n = is the number of subtests in the short form

$\sum r_{jk}$ = the sum of the intercorrelations between the subtests in the short form

X_c = the sum of the subtest scaled scores in the short form

$\overline{X_c}$ = the normative mean of the short form subtests, which is equal to $10n$.

Linear Equating Formulae Based on the Analysis Sample Data:

Linear equating formula for SF1:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{6 + 2(4.88)}} \right) \times (\text{Similarities} + \text{Arithmetic} + \text{Vocabulary} + \text{Picture Completion} + \text{Coding} + \text{Block Design} - 60) + 100$$

Linear equating formula for SF2:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{4 + 2(2.42)}} \right) \times (\text{Vocabulary} + \text{Arithmetic} + \text{Picture Arrangement} + \text{Block Design} - 40) + 100$$

Linear equating formula for SF3:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{5 + 2(3.60)}} \right) \times (\text{Information} + \text{Vocabulary} + \text{Picture Completion} + \text{Coding} + \text{Block Design} - 50) + 100$$

Linear equating formula for SF4:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{4 + 2(2.66)}} \right) \times (\text{Information} + \text{Vocabulary} + \text{Picture Completion} + \text{Block Design} - 40) + 100$$

Linear equating formula for SF5:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{4 + 2(2.00)}} \right) \times (\text{Similarities} + \text{Arithmetic} + \text{Picture Completion} + \text{Coding} - 40) + 100$$

Linear equating formula for SF6:

$$\text{SFIQ} = \left(\frac{15}{3\sqrt{4 + 2(2.67)}} \right) \times (\text{Similarities} + \text{Arithmetic} + \text{Picture Completion} + \text{Block Design} - 40) + 100$$

Appendix C

Regression Equations Based on the Analysis Sample Data

Regression equation for SF1:

$$\hat{Y}_{FSIQ} = 41.731 + 1.031_{\text{(Similarities)}} + 1.074_{\text{(Arithmetic)}} + 1.170_{\text{(Vocabulary)}} + .908_{\text{(Picture Completion)}} + .926_{\text{(Coding)}} + .919_{\text{(Block Design)}}$$

Regression equation for SF2:

$$\hat{Y}_{FSIQ} = 49.625 + 1.819_{\text{(Vocabulary)}} + 1.279_{\text{(Arithmetic)}} + .993_{\text{(Picture Arrangement)}} + 1.261_{\text{(Block Design)}}$$

Regression equation for SF3:

$$\hat{Y}_{FSIQ} = 42.470 + 1.364_{\text{(Information)}} + 1.349_{\text{(Vocabulary)}} + 1.102_{\text{(Picture Completion)}} + 1.043_{\text{(Coding)}} + 1.077_{\text{(Block Design)}}$$

Regression equation for SF4:

$$\hat{Y}_{FSIQ} = 48.278 + 1.403_{\text{(Information)}} + 1.440_{\text{(Vocabulary)}} + 1.282_{\text{(Picture Completion)}} + 1.311_{\text{(Block Design)}}$$

Regression equation for SF5:

$$\hat{Y}_{FSIQ} = 44.613 + 1.958_{\text{(Similarities)}} + 1.467_{\text{(Arithmetic)}} + 1.303_{\text{(Picture Completion)}} + 1.137_{\text{(Coding)}}$$

Regression equation for SF6:

$$\hat{Y}_{FSIQ} = 47.439 + 1.850_{\text{(Similarities)}} + 1.454_{\text{(Arithmetic)}} + 1.166_{\text{(Picture Completion)}} + 1.108_{\text{(Block Design)}}$$

Appendix D

Formulae Used in the Current Investigation

Formula for Computing the Validity Coefficient of the Short Form (Sattler, 2001):

$$\text{Validity} = \frac{\sum \sum r_{ji}}{\sqrt{k + 2 \sum r_{ij}} \sqrt{t + 2 \sum r_{im}}}$$

Where: $\sum \sum r_{ji}$ = the total of the following three sums (1) sum of the reliabilities of the subtests in the short form (2) twice the sum of the intercorrelations among the short form subtests (3) the sum of the intercorrelations of the short form subtests with the remaining subtests not in the short form

k = number of subtests in the short form

r_{ij} = sum of the intercorrelations between the subtests in the short form with all of the other subtests in the full battery

t = number of subtests included in the full battery, which has a uniform value of 13

r_{im} = sum of the intercorrelations between the subtests in the full battery.

Formula for Computing the Reliability Coefficient of the Short Form (Tellegen & Briggs, 1967):

$$\text{Reliability} = \frac{\sum r_{ji} + 2 \sum r_{jk}}{n + 2 \sum r_{jk}}$$

Where: r_{ji} sum of the reliabilities of the subtests in the short form

r_{jk} = sum of the intercorrelations between the subtests in the short form

n = number of subtests in the short form.

Formula for Computing the Standard Error of Measurement (Sattler, 2001):

$$SEM = SD \sqrt{1 - r_{xx}}$$

Where: SD = standard deviation, which has a uniform value of 15

r_{xx} = reliability coefficient.

Formula for Computing the Standard Error of Estimate (Sattler, 2001):

$$SE_{est} = SD_y \sqrt{1 - r_{xy}^2}$$

Where: SD_y = standard deviation of the predicted Y (i.e., SFIQ), which has a uniform value of 15

r_{xy}^2 = square of the correlation between the X (i.e., FSIQ) and Y (i.e., SFIQ) scores.

Formula for Computing the Standard Error of Estimate Specifically for Linear Equating (Silverstein, 1984):

$$SE_{est} = SD_y \sqrt{2(1 - r_{xy})}$$

Where: SD_y = standard deviation of the predicted Y (i.e., SFIQ), which has a uniform value of 15

r_{xy} = the correlation between the X (i.e., FSIQ) and Y (i.e., SFIQ) scores.

Formula for Computing Confidence Intervals Using Standard Error of Estimate (Sattler, 2001):

$$\text{Confidence Interval} = \text{Score} \pm z (SE_{est})$$

Where: $z = 1.64$ for 90% and 1.96 for 95%

SE_{est} = the standard error of estimate for the short form.

Formula for Calculating Effect Size Estimate (Cohen, 1977)

$$d' = \frac{\overline{X}_1 - \overline{X}_2}{SD}$$

Where: d' = effect size estimate

$\overline{X}_1$ = Mean of FSIQ scores

$\overline{X}_2$ = Mean of SFIQ scores

SD = the standard deviation of the FSIQ scores

Formula for Calculating the Adjusted Effect Size for Paired t-Tests (Cohen, 1977):

$$d = \frac{d'}{\sqrt{1-r}}$$

Where: d = adjusted effect size for paired t-tests

d' = effect size estimate

r = correlation between the short form and full form