NOTE TO USERS

msg pg - acknowledgement

University of Alberta

Development of a New Teaching Method to Reduce Scoring Errors on the WISC-III and WAIS-III

BY

Yvonne Legris 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY

IN

Counselling Psychology

Department of Educational Psychology

Edmonton, Alberta

Fall 2003

Your file Votre référence
ISBN: 0-612-88011-7
Our file Notre référence
ISBN: 0-612-88011-7

University of Alberta

Library Release Form

Name of Author	Yvonne Lorraine Marie Legris
Title of Thesis	Development of A New Teaching Method To Reduce Scoring Errors On The WISC-III and WAIS-III
Degree	Doctor of Philosophy

Year this Degree Granted 2003

RR#4 Tofield, Alberta T0B 4J0

August 25, 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommended to the Faulty of Graduate Studies and Research for acceptance, a thesis entitled DEVELOPMENT Of A NEW TEACHING METHOD To REDUCE SCORING ERRORS On The WISC-III and WAIS-III submitted by Yvonne Lorraine Legris in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Counselling Psychology.

Dr. J. G. Paterson- Committee Chair

Dr. H. L. Janzen- Supervisor

Dr. W. T. Rogers

Dr. B. L. Paulson

Dr. M. F. Juliebo

Dr. L. Derevensky

Date: May 5/03

ABSTRACT

The purpose of the present study was to determine if the experimental teaching method was more effective than the regular teaching method in reducing the number of administration and scoring errors made by student examiners on the WISC-III and the WAIS-III. The experimental teaching method contained two teaching interventions, a specific lecture on administration and scoring errors and an observation of a Wechsler demonstration. These two interventions replaced two of the three practice sessions included in the regular teaching method. The number of administration and scoring errors committed by both groups of students on both tests was calculated using checklists developed for this study. The subtest errors were the sums of errors made on each of the subtests while the total errors were the sums of errors made on the cover page and the back of the cover page of the record form. The results indicated that both types of errors were haphazardly committed. For the WISC-III, there was generally little discrepancy between the number of errors made by the experimental and control groups on the performance subtests. In contrast, the control group committed fewer verbal subtest errors than the experimental group. For the WAIS-III, the control group committed fewer subtest errors. For both the WISC-III and the WAIS-III, the experimental group committed a greater number of total errors. The teaching intervention was not found to reduce scoring errors for the experimental group. Administration and scoring errors continued to surface from one test (WISC-III) to the next (WAIS-III) and the control group committed fewer overall subtest and total errors. Conclusions about the study were made such as the need to maintain the use of practice administrations and other interventions like formalized, immediate feedback to students about their errors in order to provide a greater teaching impact that could result in fewer administration and scoring errors. Implications for practice included using the checklist from this study to identify and calculate the number of administration and scoring errors. Recommendations for future research were centered on reducing the high number of administration and scoring errors on both tests.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

Introduction

The Wechsler Intelligence tests (WISC-III, 1984 and WAIS-III, 1998) are clinical instruments used to assess the intellectual ability of individuals. These tests and their predecessors (WISC, WISC-R, WAIS) have a long history as a decision-making tool in the realm of psychology and education. Scores from these intelligence tests provide data about a person's intellectual performance, which are then transformed into meaningful information through a process of interpretation (Saklofske & Prifitera, 1998). This interpretation is then used as part of the assessment process for identifying such things as suitable educational programming for children and adults, increased accessibility to appropriate educational and emotional services, and the promotion of mental health (Sattler, 1988). Psychological assessment is a problem-solving activity and includes " the process of evaluating the characteristic strengths and weaknessess of an individual as a basis for informal decision making" (Gregory, 1998 p.27). The accuracy of the data obtained from these tests is important since erroneous test scores can lead to faulty interpretations. Thus, the incorrect administration and scoring of intelligence tests affect the reliability and validity of test scores, which in turn may result in misleading and potentially harmful consequences. For example, improper labeling of individuals can result in internalization of negative self-image or unrealistic expectations of future performance (Reynolds & Kamphaus, 1990). Therefore, determining accurate intellectual testing results is of paramount importance in the assessment process.

The Occurrence of Administration and Scoring Errors

Speculation about the possibility of testing errors occurring during the administration and scoring of IQ protocols was initially voiced by Goodenough (1940) and later by Littell (1960) who both raised concerns about the effects of such errors and their negative impact on the reliability of intellectual testing results. For example, the accuracy, or lack of such, of the examiners' administration techniques, and speculation about other possible variables that might affect the accuracy of reported results were instrumental in spearheading research about scoring errors.

Alfonso and Pratt (1997) summarized the scoring and administrative errors on Wechsler tests in their review of relevant studies spanning the period from 1970 to 1995.

The types of errors made were found to be mainly clerical in nature: incorrectly adding subtest scores and failing to record responses in a verbatim manner. Incorrectly assigning point values to responses, in particular with the subtests that involved verbal expressions, was noticed by other researchers (Belk, LoBello, Ray, & Zachar, 2002; Franklin, Stillman, Burpeau, & Sabers, 1982; Slate & Jones, 1990a). The Information, Similarities, and Comprehension subtests are considered more difficult to score because of the difficulty in accurately scoring ambiguous responses. Other, less frequently noted sources of error included incorrect calculations of chronological age, improper discontinuation of subtests, and inappropriate probing or questioning (Franklin, Stillman, Burpeau, & Sabers, 1982; Slate, Jones, Murray, & Coulter, 1993).

In discussing the negative impact of administration and scoring errors made on tests, Sattler (1988) explained that students in the process of learning how to administer, score, and interpret test scores should become aware of the common errors in test scoring. He highlighted the importance for students to learn the administration and scoring principles and criteria as outlined in test manuals.

The Importance of Improving Administration and Scoring Accuracy

The desirable consequence of improved administration and scoring accuracy on Wechsler intelligence tests is a more precise assessment process with the end result being increased diagnostic accuracy. As stated by Woody and Robertson (1988), "Assessment and diagnosis form the distinguishing cornerstone for the practice of clinical psychology" (p. 210). With properly administered and scored tests, the obtained information can be interpreted in conjunction with other test data as part of the whole assessment process. Shaffer, Lucas, and Richters (1999) described assessment as "a process whose products precipitate decision making, action taking, and policy formulation" (p. 312). As such, psychological assessment results provide important information and the recommendations made from these results can have considerable and serious consequences for an individual. For example, assigning an incorrect diagnostic classification may result in inappropriate patient treatment and care and erroneous labeling of a student.

Scores from the WISC-III and WAIS-III IQ tests provide basic useful information about an individual's intellectual functioning. The IQ scores obtained from the Wechsler

tests are valuable information that provide an overall impression of a person's cognitive ability and specific intellectual strengths and weaknesses. This information is used in conjunction with other test results (e.g., academic achievement) and assessment procedures (e.g., interview and observations) to arrive at important conclusions about an individual such as cognitive impairments or gifted ability. However, when test scores and assessment information are error-laden from the onset, the results are misleading and the long-term impact can be a distorted and damaging representation of the individual tested.

Sattler (1992) states "Test and other assessment procedures are powerful tools, but their effectiveness will depend on your [the examiners'] skill and knowledge" (p.5). Therefore, it is crucial that students who want to learn to administer and score intelligence tests must be aware of the need for accuracy in both the administration and scoring of these tests. Not only is this a necessary in terms of the accuracy of their work that will lead to more reliable and valid test results and better diagnostic accuracy, but also in terms of their training to become competent professionals in the area of intellectual testing and psychological assessment. Therefore, a reduction in IQ scoring errors leads to increased scoring accuracy, a subsequently useful assessment process, and more accurate diagnoses and recommendations to assist the individual being tested.

The need for accuracy in the administration and scoring of intelligence tests also has implications for psychologists' standard of ethical conduct. These standards require that professionals in the assessment field be capable of understanding, administering, and scoring IQ tests with a high level of proficiency. A number of ethical guidelines, including *The Ethical Principles of Psychologists and Code of Conduct* (APA, 1992) and the College of Alberta Psychologists' *Code of Conduct* (1997), state that psychologists are responsible for maintaining the highest standards of conduct and competency. Additionally, the *Principles for Fair Student Assessment Practices for Education in Canada* (1993) addresses issues related to judging and scoring of student performance, including errors in scoring that could negatively influence assessment results. Likewise, the *Standards for Educational and Psychological Testing* (1999), outlines testing practices, the criteria for evaluating tests, and the effects of test use. Standard 11.10 of *The Responsibility of Test Users* within this latter set of standards states that "test users should be alert to the possibility of scoring errors; they should arrange for rescoring if

individual scores or aggregated data suggest the need for it" (p.115). The implications for student examiners and psychologists who do not adhere to these ethical standards is a lack of competency leading to unfair assessment practices.

The Need for Effective Teaching Methods

In their review of administration and scoring errors on Wechsler tests, Slate and Hunnicutt (1988) indicated that inadequate instruction and training on the part of test administrators was responsible for the high incidence of scoring errors. According to the *Standards of Educational and Psychological Testing* (1999), university instructors are responsible for the competent use of cognitive assessment instruments and for providing instructional programs that will help student examiners achieve the highest possible level of competence in the field of assessment. For their part, student examiners are responsible for committing themselves to spend the time and effort required for learning and becoming proficient in this field. Alfonso and Pratt (1997) emphasized the responsibility of university instructors to design suitable cognitive assessment courses because "ultimately, students rely on the expertise of university trainers in the development of their assessment skills [and for students in training] to administer, score, and interpret properly the assessment instruments they will use in practice" (pp. 339).

Several attempts have been made to design appropriate teaching methods that will result in increased accuracy of administration and scoring with a decreased number of errors and faulty results (Moon, Fantuzzo, & Gorsuch, 1986; Slate, Jones, & Cover, 1992). This research has demonstrated that increased focus on pre-learning of material, observation of a Wechsler administration, use of quizzes and tests of proficiency, and provision of feedback to students were teaching interventions that helped reduce the number of administration and scoring errors.

Training students to properly administer and score intelligence tests is of pivotal importance in regard to obtaining accurate results and has the effect of increasing students' competence in the area of assessment. With the increase in students' competence in producing accurate IQ results, more precise and accurate diagnoses can be rendered. Improving the field of diagnosis, where psychologists can provide an invaluable service, is connected to improving training methods for students to ensure the accuracy of their results.

Purpose of the Study

Accountability to both the profession of psychology in terms of diagnostic accuracy and to the public in upholding high ethical standards and principles underlies and emphasizes the need for student examiners to competently administer and score intelligence tests. Therefore, the purpose of this study was to determine if the application of a specific teaching method designed to train graduate students who are learning to administer and score the Wechsler child and adult tests of intelligence (Wechsler Intelligence Scale for Children-Third Edition, 1984 [WISC-III] and the Wechsler Adult Intelligence Scale-Third Edition, 1998[WAIS-III]) would lead to a reduction in administration and scoring errors committed by them.

Research Objectives

In order to achieve the purpose of this study, the following objectives were followed:

1. Development of a new method of teaching the administration and scoring of the Wechsler Intelligence tests.
2. Assessment of the new teaching method by comparing the number and types of administration and scoring errors committed by the student examiners who received the new teaching method with the number and types of administration and scoring errors from the student examiners who received the usual teaching method (i.e., the instructional format historically used in the assessment class).

Organization of the Dissertation

This dissertation includes eight chapters. The first chapter contains a brief introduction followed by a statement of purpose and delineation of the research design. The second chapter contains a review of the literature with an emphasis placed on previous studies that led to the formulation of the method developed for this study. Chapter Three details the empirical methods and procedures followed, including a description of the research design, participants, instruments used, and teaching methods used. Chapters Four, Five, Six, and Seven provide the results of this study. Chapter Eight includes a summary of the study and findings, the conclusions drawn in light of the limitations of this study, and commentary about practice and recommendations for future research.

CHAPTER TWO

Literature Review

The review of literature is divided into three sections. In the first section, the rationale for choosing the Wechsler tests used in this study and a history and description of them is presented. The second section follows with a description of the administration and scoring errors committed on the Wechsler tests. In the third section, topics relevant to training the student examiners, such as training programs, models, specific teaching interventions, and student examiner characteristics, are described.

Rationale for choosing the Wechsler Intelligence Tests

The Wechsler scales were chosen as instruments for this study because of the extensive research background supporting their usefulness in the assessment process. Over 1,100 research publications have been reported and attest to the WISC-R's clinical utility and validity (Reynolds & Kaufman, 1990). Their importance to psychology has been heralded by knowledgeable researchers in the area of intellectual assessment: "The Wechsler scales enjoy unprecedented popularity and have a rich clinical and research tradition""(Kamphaus, 2001).

Two of the Wechsler tests were examined for scoring errors as a part of this study: the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) and the Wechsler Adult Intelligence Scale- Third Edition (WAIS-III). The Wechsler tests were chosen as instruments for this study partly because they are well-known and frequently used intelligence tests. During the 1960' and 1970's the WISC and the revised edition (WISC-R) gained in popularity (Aiken, 1996). The WISC continues to be the most popular test worldwide with little decline foreseen in the future widespread use of the test (Kamphaus, 2001). Reasons for its popularity include the extensive amount of research conducted to attest for its clinical usefulness and validity (Reynolds & Kaufman, 1990) and the number of standardization samples (Canadian, Australian, and United Kingdom, as well as the United States), and the 11 language translations of the test (Prifitera & Saklofske, 1998). The Wechsler scales "are uncontested as the primary cognitive measures of adolescent and adult intelligence" (Kaufman & Lichtenberger, 2002, p. 3) and the WAIS-III was described as the most popular intelligence test used in assessing adult intelligence (Camara, Nathan, & Puente, 2000). Other well known intelligence tests

are available and in use such as the Stanford-Binet: Fourth Edition, which has been described as "The oldest individual intelligence test in existence" (Gregory, in Cullari p.61). The Kaufman tests are also popular and include the Kaufman Assessment Battery for Children (K-ABC) (Kaufman, Kamphaus, & Kaufman, 1985) and the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990). However, the WISC-III and the WAIS-III were used in this study because of their established popularity and documented utility and also because they were both taught during the Individual Assessment Course to the participants of this study.

The Wechsler Intelligence tests

A brief history of the Wechsler test is presented in order to place the latest versions of these tests within a historical context. The first intelligence test developed by Wechsler was the Wechsler-Bellevue Intelligence test used for adolescents and adults (Wechsler, 1939). He reportedly borrowed test items from The Stanford Binet Intelligence Scale (Binet & Simon, 1916), the point scale format scoring system from Robert Yerkes (1917), and the deviation score format for the Intelligent Quotient (IQ) developed by Otis (Sattler, 1992).

Wechsler viewed the construct of intelligence in terms of a continuum of abilities (Sattler, 1992). Consequently, he formed two scales, the Verbal and Performance scales. The Verbal scale consisted of subtests that assessed verbal comprehension, concept formation, and reasoning. The Performance scale consisted of subtests that assessed perceptual organizational abilities, visual and motor abilities, and abstract nonverbal reasoning. The total scores for each scale corresponded, respectively, to Verbal IQ and Performance IQ. The sum of the two IQ's yielded the Full Scale IQ (FSIQ). Wechsler stated that the FSIQ represented an index of general mental ability but did not equal a person's level of intelligence. Rather, he defined intelligence as " The aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment" (Wechsler, 1958). He considered other factors like motivation, drive, and persistence, also contributed to intelligent behaviour. However, he did not include their measurement in his scales because these constructs could not be measured through the Verbal and Performance subtests of the Wechsler tests. He advocated weighing all factors in the assessment of intelligence (Armour-Thomas &

Gopaul-McNicol 1998). A description of the different versions of the Wechsler tests follows.

Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949) and the Wechsler Intelligence Scale for Children - Revised (WISC-R) (Wechsler, 1974)

The WISC was developed to assess intelligence in children and adolescents and was described as a downward extension of the original Wechsler-Bellevue Intelligence Scale (Sattler, 1992). Easier items were added to the test to make it more suitable for children and adolescents. The WISC-R was published in 1974, twenty-five years after the original WISC (Wechsler, 1949). Almost a third (72%) of the test items from the WISC were retained intact or modified (8%) for the WISC-R with one subtest, Coding, left unchanged from the WISC to the WISC-R (Sattler, 1992). The WISC-R spans the ages from 6 years, 6 months to 16 years 6 months and, the standardization group contained a cross section of children (i.e., Caucasian, American Indian, Asian, Puerto Ricans, and Mexican Americans) with eleven different age groups. This standardization process was more extensive and representative of the American population in comparison to the WISC where only Caucasians were included in the standardization sample. There are 12 subtests in total with an even division of 6 subtests for the Verbal scale and 6 subtests for the Performance scale. The examinee's scores are converted to three deviation IQ scores: the Verbal scale IQ, Performance scale IQ, and Full Scale IQ. Three factors scores can also be calculated in addition to the Verbal, Performance, and Full Scale IQ scores. The first factor, Verbal Comprehension, consists of the sum of scaled scores for the Information, Similarities, Vocabulary, and Comprehension subtests that are then converted to a deviation IQ. This factor measures verbal knowledge and comprehension. The second factor, Perceptual Organization, consists of the sum of scaled scores for the Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests. This factor measures visual perception and organization. The third factor, Freedom from Distractibility, consists of the sum of scaled scores for the Arithmetic, Digit Span, and Coding subtests. This factor measures attention and concentration.

Internal consistency reliabilities for the Verbal, Performance, and Full Scale scores based on the eleven age groups were 0.94, 0.90 and 0.96, respectively (Sattler, 1992). Normative data for the indexes was not included in the WISC-R manual.

Individual subtest reliabilities were less satisfactory and ranged from 0.70 for the Object Assembly subtest to 0.86 for the Vocabulary subtest. The standard errors of measurement were 3.60 for the Verbal scale, 4.66 for the Performance scale and 3.19 for the Full Scale. For the individual subtests, they ranged from 1.15 for the Vocabulary subtest to 1.70 for the Object Assembly subtest.

In terms of validity measures, Sattler (1992) reported that the WISC-R had acceptable criterion validity with median correlations between the WISC-R and school grades and other measures of achievement ranging from 0.30's to a low of 0.80. Also, satisfactory concurrent validity was noted between the WISC-R and other measures of intelligence [i.e., Wechsler Preschool and Primary Scale of Intelligence (WPPSI, 1967), WAIS-R, Standford-Binet: Fourth Edition (SBIV, 1986)] with correlations between 0.70 and 0.80.

Wechsler Intelligence Scale for Children- Third Edition (WISC-III) (Wechsler, 1991)

The WISC-III is the most recent version of the Wechsler scales for children and was published 17 years after the WISC-R. Updating the norms was the main reason for the most recent revision of the test (Sattler, 1992). Similar to the WISC-R, the WISC-III spans the ages from 6 years, 6 months to 16 years 6 months with eleven age groups in the standardized sample. The WISC-III has been translated into 11 different languages (Prifitera & Saklofske, 1998). There are 13 subtests in total with 6 for the Verbal Scale and 7 for the Performance Scale. A new Performance subtest, Symbol Search, was added to this latest version. Four-factor based index scores can be calculated in addition to the Verbal, Performance, and Full Scale IQ scores. These four indexes are the Verbal Comprehension Index, Perceptual Organization Index, Freedom from Distractibility Index, and Processing Speed Index. The Processing Speed Index was added to the WISC-III and reflects a perceptual and speed-related ability. Only the Freedom from Distractibility Index, which was reported to assess attention and concentration, was noted as unsubstantiated as an independent factor (Sattler, 1992), although the four factor structure has been supported by others (Prifitera & Saklofske, 1998).

Internal consistency reliabilities for the Verbal, Performance and Full Scale scores based on the eleven age groups were 0.95, 0.91 and 0.96, respectively (Sattler, 1992). Likewise, the reliability coefficients for the four indices are 0.94 for Verbal

Comprehension, 0.90 for Perceptual Organization, 0.87 for Freedom from Distractibility, and 0.85 for Processing Speed. Individual subtest reliabilities were less satisfactory and range from 0.69 for the Object Assembly subtest to 0.87 for the Vocabulary and Block Design subtests. The standard errors of measurement were 3.53 for the Verbal scale, 4.54 for the Performance scale and 3.20 for the Full Scale. For the indices, the standard errors of measurement were 3.78 for Verbal Comprehension, 4.68 for Perceptual Organization, 5.43 for Freedom from Distractibility, and 5.83 for Processing Speed. For the individual subtests, they ranged from 1.08 for the Vocabulary subtest to 1.67 for the Object Assembly subtest.

According to the Manual for the WISC-III (1991) validity results indicated acceptable construct and predictive validity. Additionally, criterion validity was viewed as satisfactory according to Sattler (2001) based on studies where the WISC-III FSIQ was compared to other intelligence measures with a mean correlation of 0.72. Specific correlations with other Wechsler tests included the WPPSI-R (0.85), the WAIS-R (0.88), and the WAIS-III (0.93) and the SBIV (0.74), K-ABC (.70), and K-BIT (0.80). Sattler (2001) stated that there was strong evidence from the research literature that the WISC-III provides a measure of general intelligence although there remains some uncertainty regarding the existence of three or four factor indices (i.e., the existence of a Freedom from Distractibility factor).

Wechsler Adult Intelligence Scale- WAIS and the Wechsler Adult Intelligence Scale-Revised Edition (WAIS-R) (Wechsler, 1981).

The WAIS-R is a revision of the 1981 version of the Wechsler-Bellevue Intelligence Scale-Form 1 for adolescents and adults (Wechsler, 1939). The original form was revised twice, once in 1955 (WAIS, Wechsler, 1955) and again in 1981 as the WAIS-R. The WAIS-R is an adult intelligence test that consists of verbal and performance scales. The Verbal Scale includes six subtests that assess verbal ability and comprehension and the ability to process verbal information and apply verbal skills to the solution of new problems. The Performance Scale includes five subtests that assess perceptual ability and organization and the ability to think in visual images, manipulate stimuli, and apply non-verbal reasoning to problem solving. The order of subtest administration begins with a verbal subtest followed by a performance subtest weaving

between the two areas until all subtests are administered. The WAIS-R covers the age ranges from 16 years, 0 months to 74 years, 11 months with nine different age groups in the standardization sample.

A three-factor structure was reported for the WAIS-R resulting in three factors that are scored in the interpretation of the results. The Verbal Comprehension Factor measures verbally acquired knowledge and verbal reasoning and consists of the Information, Vocabulary, Comprehension, and Similarities subtests. The Perceptual Organization Factor measures nonverbal, fluid reasoning, attentiveness to detail, and visual-motor integration and consists of the Block Design, Object Assembly, and Picture Completion subtests. The Freedom from Distractibility Factor measures concentration, attention, short-term memory, and numerical ability and consists of the Digit Span and Arithmetic subtests.

The WAIS-R was reported to have technical problems in the standardization data, particularly for the 16-to 17-year-old and the 18-to 19-year-old groups. While it was expected that performance would increase with increasing age for each of the two groups, the IQ distribution for these two groups did not follow this developmental trend (Kaufman, 1990). Instead, highly similar overall IQ scores were reported for the 16 to 17 and the 18 to 19 year old groups where differences, based on educational level, would have been expected. In addition, the scaled scores for all age groups of the WAIS-R were based on the scaled scores of a reference group of 20 to 34 year olds. Therefore, meaningful comparisons of younger and older adults to a different age reference group was a problem only partially corrected through the use of age-corrected scores. Kamphaus (1983) also noted that scoring errors occurred during the procedure of determining age-corrected scores.

In terms of reliability data for the WAIS-R, reliability coefficients ranged from 0.95 to 0.97 for the Verbal scale, from 0.88 to 0.94 for the Performance scale, and from 0.96 to 0.98 for the Full Scale. Individual subtest reliabilities ranged from 0.96 for the Vocabulary subtest to 0.52 for the Object Assembly subtest. The standard error of measurement based on the average of the nine age groups, was 2.53 for full Scale, 2.74 for Verbal scale, and 4.14 for Performance scale. The Verbal scale subtest standard errors of measurement ranged from 0.61 to 1.24 scaled score points and the Performance

scale subtest standard errors of measurement ranged from 0.98 to 1.54. Construct and concurrent and validity studies were reported as uniformly positive (Sattler, 1992). Construct validity was assessed using factor analysis "because it provides a method for determining the structure and components of intelligence measured by a given test (Sattler, 1992, p.224). All 11 subtests were found to measure general intelligence within a moderate to high degree of success with support for the FSIQ score as an aggregate measure of intelligence.

Wechsler Adult Intelligence Scale- Third Edition (WAIS-III) (Wechsler,1997)

The WAIS-III is the latest version of the Wechsler Adult Intelligence test. Gregory (1999) described this version as a significant improvement over the WAIS-R in a number of important areas. Subtest scaled scores are no longer based on the performance of a reference group of subjects from the standardization sample. The standardization sample was broadened to include an extended age range from 16 to 89. Additional changes included an increase in the number of non-biased items (i.e.,items considered more easily answered by North Americans), more contemporary items were included, and a new record form was developed. A positive change for the WAIS-III was the extension of the IQ score range with lower level IQ's possible (Saklofske & Hildebrand, 1997). Two new subtests, Letter-Number Sequencing and Matrix Reasoning, were added. The Letter-Number Sequencing subtest is a new supplementary subtest whose scores were not required to obtain the VIQ, PIQ, or the FSIQ but it does contribute to the Working Memory Index score. The Matrix Reasoning subtest replaces the Object Assembly subtest and is used in the calculation of the PIQ. Optional procedures called Incidental Learning and Digit Symbol Copy were added to the Digit Symbol Coding subtest "to help the examiner rule out potential problems or to identify weak areas if the examinee does not perform well on the subtest" (p.15, WAIS-IIII Manual, 1997).

Unlike the three factors of the WAIS-R, the WAIS-III reports four Indexes. These indexes were formed by summing scaled scores from particular subtests. For the Verbal Comprehension Index, the Vocabulary, Similarities, and Information subtest scaled scores are summed. For the Perceptual Organization Index, the Picture Completion, Block Design, and Matrix Reasoning scores are summed. The Arithmetic, Digit Span, and Letter-Number Sequencing form the Working Memory Index, and the

Digit Symbol-Coding and Symbol Search scores from the Processing Speed Index. Both the Verbal Comprehension and Perceptual Organization indexes remain similar to the WAIS-R except for the inclusion of the Matrix Reasoning subtest that replaces the Object Assembly subtest to form the Perceptual Organization Index. The new indexes are the Working Memory index and the Processing Speed Index. The Working Memory Index measures attention to information and the ability to briefly hold and process information in memory, then formulate a response. The Processing Speed Index measures the ability to process visual information quickly.

The reliability coefficients for the Verbal, Performance, and Full Scale Scores are 0.97, 0.94, and 0.98 respectively (WAIS-III, WMS-III Technical Manual, 1997). Likewise, the reliability coefficients for the four indices are 0.96 for Verbal Comprehension, 0.93 for Perceptual Organization, 0.94 for Working Memory, and 0.88 for Processing Speed. Individual subtest reliabilities range from 0.70 to 0.93. Stability reliability coefficients was 0.96 for Full Scale scores, 0.96 Verbal scale scores, 0.91 for Performance scale scores. Likewise, the stability reliability coefficients for the four indices are 0.95 for Verbal Comprehension, 0.88 for Perceptual Organization, 0.89 for Working Memory, and 0.89 for Processing Speed. The standard errors of measurement were 2.55 for the Verbal scale, 3.67 for the Performance scale and 2.30 for the Full Scale. For the four indices, the standard errors of measurement were 3.01 for Verbal Comprehension, 3.95 for Perceptual Organization, 3.84 for Working Memory, and 5.13 for Processing Speed. Individual subtest standard errors of measurement ranged from 0.79 for the Vocabulary subtest to 1.66 for the Object Assembly subtest.

In terms of validity research, content validity, defined as "the degree to which the test items adequately represent and relate to the trait or function that is being measured" (p. 75, Technical Manual, 1997), was assessed by a number of means. These included comprehensive literature views to identify problems with the WAIS-R and make changes, consultation with psychologists to review items for deletion, revision, or inclusion, additional revisions based on surveys and focus groups, and extensive validation studies with the WAIS-R, WISC-III, WIAT (Wechsler Individual Achievement Test, 1992), the WMS-II (Wechsler Memory Scale-Revised, 1987), and the SBIV (1986) occurred to assure concurrent validity. Factor analysis was used to assess construct validity and the

14 subtests were all found to be moderate to high degree of success in measuring general intelligence with support for the VIQ and PIQ as separate features of the test and the FSIQ as an aggregate measure of intelligence (Sattler, 2001). Intercorrelations between the subtests and the scales revealed higher correlations between the Verbal subtests and the Verbal Scale (Mdn r = .76) than between the Performance subtests and the Performance Scale (Mdn r = .64) (WAIS-III WMS-III Technical Manual) but that the pattern of intercorrelations among the subtests suggests good discriminant (i.e., "tasks that purport to measure different functions yield relatively low or nonsignificant correlations" p.384, Sattler, 2001) and convergent validity (i.e., "tasks that theoretically tap similar functions correlate more highly with each other than with tasks that theoretically measure different functions" (p.384, Sattler, 2001).

Scoring and Administration Errors on the Wechsler Tests

The following section contains information about scoring errors on the Wechsler tests presented in table format and summarized. The first table contains scoring and administration errors on the WISC, WISC-R, and WISC-III

Table 1.
Summary of Studies Investigating Scoring and Administration Errors on the WISC, WISC-R, and WISC-III.

Study	Wechsler Test	Sample	Number of Protocols	Major Findings
Miller, Chansky, & Gredler (1970)	WISC	32 graduate students	32	Total errors = 68, ξ= 2.12 errors per protocol (2 protocols free of errors)
Miller & Chanksy (1972)	WISC	64 Doctoral and Masters level psychologists	64	Total errors = 152, ξ= 2.37 errors per protocol (10 protocols free of errors)
Warren & Brown (1972)	WISC	40 graduate students	240 (6 protocols each)	Total errors =1, 939, ξ = 8.1 errors per protocol

Study	Wechsler Test	Sample	Number of Protocols	Major Findings
Sherrets, Gard, & Langner (1979)	WISC & WISC-R	39 examiners (included a combination of Psychologists, and graduate students)	200 (randomly chosen)	Total errors = 119, ξ= 3.05 errors per protocol
Conner & Woodall (1983)	WISC-R	10 graduate students	150 (15 protocols each)	A decrease in administrative and total errors over 15 administrations.
Beasley, Lobasher, Henley, & Smith (1988)	WISC & WISC-R	Undetermined number of psychologists	457	24% of protocols contained calculation errors.
Slate & Chick (1989)	WISC-R	14 graduate students	112 (8 protocols each)	ξ= 8.1 errors per protocol and 15.2 errors per protocol when failure to record responses included as error type.
Slate & Jones (1990a)	WISC-R	26 graduate students	217 (approx. 8 protocols each)	ξ= 11.3 errors per protocol. No improvement over 5 protocols, but reduction in errors from 6-10 protocols.
Slate, Jones, Coulter, & Covert (1992)	WISC-R	9 Psychologists	56 (approx. 6 protocols each)	ξ= 38.4 errors per protocol, ξ= 8.7 errors per protocol when failure to record responses not counted as an error type.
Klassen & Kishor (1996)	WISC-R and WISC-III	7 School Psychologists	252 protocols (18 WISC-R and 18 WISC-III	ξ= 6.86 errors per WISC-R protocol, ξ= 7.57 errors per WISC-III protocol.
Alfonso, Johnson, Patinella, & Rader (1998)	WISC-III	15 graduate students	60 protocols (4 protocols each)	ξ= 7.8 errors per protocol. Reduction in errors from protocol 1 (14.4 errors) to protocol 4 (5.4 errors).

Scoring Errors on the WISC

Miller, Chansky, and Gredler (1970) examined the scoring errors committed by 32 psychology graduate students. The students scored the same fabricated protocol and only 2 of the 32 students did not commit errors. The remaining 30 students committed a combination of mechanical (administration) errors, clerical (calculation errors and errors in locating scores in the tables), and errors made as a result of scoring ambiguity (i.e., errors in assigning credit to the verbal responses). Most of the errors made on the performance subtests were mechanical errors, with the most frequent error being exceeding the discontinuance criteria. Most of the errors made on the verbal subtests were due to assigning incorrect credit to the Comprehension and Vocabulary subtests in particular and to a lesser extent the Information and Similarities subtests.

Miller and Chansky (1972) found a similar number of errors when professional psychologists were asked to score the same fabricated WISC protocol. Only 10 of the 64 psychologists did not commit any errors. The frequency and type of errors made were similar to those committed by the graduate students. They recommended that the WISC be revised with modifications to the administration and scoring procedures. They suggested a need for better training for psychologists in all phases of scoring the WISC and added that the Wechsler manual needed to be revised in order to clarify the scoring criteria for the verbal subtests in particular.

Warren and Brown (1972) examined the scoring errors made by 40 graduate student trainees who scored real as opposed to fabricated protocols. An average of 8.1 errors per WISC protocol was cited and there was no significant decrease in errors between their first and last three WISC administrations. The most frequent errors were failing to record responses, failing to follow correct procedures, and calculation errors.

In summary, scoring errors on the WISC might possibly reflect a lack of clarity with respect to administration and scoring guidelines. Administration and calculation errors categorized under the umbrella term of mechanical scoring errors were also reported as responsible for many of the obtained errors.

Scoring Errors on the WISC and WISC-R

Subsequent studies involved the use of both the WISC and WISC-R and then focused exclusively on the WISC-R as the first edition of the WISC was replaced.

Sherrets, Gard, and Langner (1979) examined 200 real WISC and WISC-R protocols randomly selected from a psychiatric setting and a local public school. Ph.D. psychologists and graduate students scored the protocols from the psychiatric setting and school psychologists and graduate students scored the protocols from the school. The WISC and WISC-R protocols were grouped together for this analysis and possible error differences between the two versions of the test were not examined. Sherrets et al. examined the number of clerical errors such as transformation of scores and calculation errors and found that 89% of all examiners made at least one error. Failure to add scores correctly, improper calculation of scaled scores, and incorrectly transferring scaled scores from the tables were the most frequent errors noted. Almost half (46.5 %) of the protocols revealed at least one scoring error. Neither the level of training or setting were related to the errors committed.

Beasley, Lobasher, Henley, and Smith (1988) focused exclusively on calculation errors of the WISC and WISC-R. They found calculation errors in 111 of 457 WISC and WISC-R protocols. These errors were made by an undetermined number of professional psychologists who had provided IQ assessments to over 600 children as part of a longitudinal study in the United Kingdom. A series of 457 protocols were checked for calculation errors and they included, in order of prevalence, incorrect conversion of raw scores to scaled scores (68%), incorrect age calculation (14%), incorrect addition of scaled scores errors, prorating score errors (6%), and transferring scores from the tables (2%). They advocated an automated process, like computer scoring, to calculate scores that could help to reduce the number of calculation errors.

Scoring Errors on the WISC-R

In their study, Conner and Woodall (1983) found that a combination of structured feedback about scoring errors on the WISC-R, in the form of a checklist and individual conferences, and practice with administration of the WISC-R resulted in a decrease in scoring errors. Of the five error types (total, administrative, response scoring, IQ, and mathematical) only the administrative and total errors decreased. The administrative errors consisted of errors in obtaining basal and ceiling levels in addition to failing to record responses verbatim and incorrectly cueing responses. The total error rate was a calculation of combined errors. Despite the structured feedback received by the ten

graduate students, the IQ, mechanical and response scoring errors continued to significantly surface after 15 administrations. Response scoring errors accounted for 58% of all errors made on each protocol and represented a higher percentage of errors than the other three error types combined (i.e., administrative, mathematical, and IQ errors). They speculated that individual scoring patterns might surface over time for examiners who may be prone to relying more on their experience and memory rather than consulting the manual to score responses.

Slate and Chick (1989) investigated the frequency and types of errors made by 14 graduate students who scored 112 WISC-R protocols. The graduate students had engaged in extensive training that included studying the test manual, discussion of problematic administration and scoring errors, pairing of students with each checking the others' protocols for errors, and receipt of written and verbal feedback from the instructor to the students following each WISC-R submission. The students administered eight WISC-R's and the instructor observed the eighth administration; unfortunately, no information was available about whether errors decreased with practice administrations. No subtest was reported to be free from error and 8.1 errors were noted for each protocol that increased to 15.2 total errors when combined with subsequent changes in raw scores, standard scores, and IQ scores. They noted that mechanical errors and administration errors such as difficulty establishing correct basal and ceiling levels occurred most often on the performance subtests. They attributed these errors to carelessness on the part of the student examiner. The most frequent errors were observed on the verbal subtests such as incorrect point assignments for the Vocabulary, Comprehension, and Similarities subtests with the authors commenting that " there still remains a substantial amount of 'grey area' in which responses fall between scoring categories" (P. 82). Students tended to err by assigning too many points (scoring generosity) as opposed to assigning too few points. They called for clearer instructions for scoring the tests (i.e.: determining correct point values for the verbal subtests). Inappropriate questioning such as failing to question a response when needed and questioning a response when prohibited and failing to record examinee responses were the next two most frequently committed error types. They stated that a student's judgment and experience may need to be developed to increase scoring accuracy on the more ambiguous verbal subtests.

Slate and Jones (1990a) investigated the number of errors made on the WISC-R by graduate students after 5 and 10 practice administrations. They randomly assigned 16 students to complete 5 administrations and 20 students to complete 10 administrations. In total, 26 students completed 217 WISC-R administrations. The students received both written and oral feedback after each practice administration from a graduate student who had previously completed the course requirements. There was no reported decrease in the number of errors after five administrations, but a significant decrease in errors was reported across administrations six through ten. They commented that practice with feedback seemed to help students improve scoring accuracy. Overall, the students averaged 11.3 errors per WISC-R protocol and none of the protocols were reported to be without errors. The most frequently occurring error was failure to record the examinees' responses followed by incorrect point assignments, and inappropriate questioning of examinee responses, particularly on the Information, Similarities, Vocabulary, and Comprehension subtests. Mechanical errors on the performance subtests also occurred including difficulty in establishing correct basal and ceiling levels and calculation errors such as incorrectly adding subtest raw scores and miscalculation of the chronological age. They suggested that students check over their protocols for errors considering this high degree of carelessness. They also suggested a focus on specific instruction to help students score the more difficult to score verbal subtests.

In examining the scoring accuracy of nine professional psychologists with extensive training and experience in using the WISC-R (range = 4-14 years), Slate, Jones, Coulter, and Covert (1992) found an average of 38.4 errors per protocol on 56 randomly selected protocols. When failure to record responses verbatim was not counted as an error, only 8.7 errors per protocol surfaced, although no protocol was free of errors. The most frequent error types were failure to record responses, circle scores, or record the time to complete items of a subtest followed by incorrect point assignment and inappropriate questioning. Most errors were committed on the verbal subtests (i.e., Vocabulary, Comprehension, and Similarities subtests) with the Vocabulary subtest remaining the most error prone even when failure to record responses was excluded as an error type. Failing to obtain correct basal and ceiling levels were commonly occurring errors and the Picture Completion subtest was listed as one of the performance subtests

where inappropriate questioning occurred most frequently. Examiners were more likely to overestimate IQ scores (68%) than to underestimate IQ scores (13%) indicating scoring generosity. In regard to this scoring generosity mistake made by examiners, the authors stated that it "may reflect a sincere desire to help a child/client that creates a subtle pressure to 'read into answers'" (p. 81).

In summary, both student and professional psychologists continued to make errors on the revised edition of the WISC. Some evidence of a decrease in scoring errors across practice administrations with feedback about scoring errors was noted; however, both administrative and response scoring errors continued to be in evidence. Difficulties in scoring the verbal subtests continued to be a concern with the WISC-R with the additional error of failing to record examinee responses.

Scoring Errors on the WISC-III

Klassen and Kishor (1996) compared the number of clerical errors committed by seven Master's level psychologists on 18 WISC-R'S and 18 WISC-III'S. Clerical errors included errors of addition and transformation of raw, standard, and IQ scores, and calculation errors like incorrect chronological age calculation. They noted that 86% of examiners made errors on both tests (39% of errors on the WISC-R and 42% of errors on the WISC-III). The overall error rate did not decrease significantly over an 18 month period with an average of 6.86 errors per protocol for the WISC-R and 7.57 errors per protocol for the WISC-III. Errors in adding raw scores was listed as the most common error for both tests and made more frequently on the performance subtests, particularly the Coding subtest. The WISC-III was described as less vulnerable to clerical errors than the WISC-R in one area only - the transformation of raw scores to scaled scores, which may have reflected the better design and clarity of the WISC-III scoring protocol.

On the most recent edition of the Wechsler Intelligence Scale for Children, the WISC-III, Alfonso, Johnson, Patinella, and Rader (1998) investigagted the number of errors made by fifteen graduate students who completed 60 WISC-III protocols (four protocols per student). The average number of errors on the WISC-III was 7.8 errors per protocol similar to previous findings on the WISC and WISC-R. Additionally, the types of errors were also consistent and included the failure to record responses verbatim and calculation errors. The Comprehension subtest produced the most errors followed by

errors on the Similarities subtest. However, errors on the Vocabulary subtest occurred with similar frequency to errors made on other subtests suggesting that the verbal subtests may be somewhat easier to score on the WISC-III. The researchers speculated that the placement of the correct item responses within the administration portion of the manual and not in the appendix (as it was for versions of the WISC) may help, over time and with practice, to facilitate scoring. Importantly, a significant decrease in scoring errors occurred across student protocols with the mean number of errors decreasing from protocol one (14.4 errors) to protocol four (5.4 errors).

In summary, while changes in the manual of the WISC-III appear to have increased the potential for more accurate administration and scoring, these errors continue to surface.

Scoring Errors on the WAIS and WAIS-R

Table 2 contains information about scoring and administration errors on the WAIS and the WAIS-R.

Table 2.
Summary of Studies Investigating Scoring and Administration Errors on the WAIS and WAIS-R

Study	Wechsler Test	Sample	Number of Protocols	Major Findings
Franklin,Stillman Burpeau, & Sabers (1982)	WAIS	33 graduate students and psychologists	33	Examiner errors on Information, Comprehension, and Vocabulary subtests significantly different than obtained scaled scores
Ryan, Prifitera, & Powers (1983)	WAIS-R	39 graduate students and psychologists	78 (2 protocols each)	Standard deviations ranged from 1.9 to 3.5 across PIQ & VIQ scores and from 1.7 to 2.3 across FSIQ scores
Slate & Jones (1990b)	WAIS-R	26 graduate students	180	ξ= 8.8 errors per protocol. No decrease in errors after five administrations
Slate & Jones (1990c)	WAIS-R	22 graduate students	149(approx 7 protocols each)	ξ = 7.9 errors per protocol
Slate, Jones, Murray, and Coulter (1993)	WAIS-R	8 Master's level practitioners	50 (randomly selected)	ξ= 36.9 errors per protocol ξ= 15.4 errors when failure to record responses not counted as errors

Franklin, Stillman, Burpeau, and Sabers (1982) had three school psychologists develop four fabricated WAIS scripts and protocols and had four "clients" memorize the responses on the scripts. Thirty-three practicing school psychologists and school psychology students were combined together as a group that assessed the clients resulting in a total of 33 WAIS administrations. Differences between the number of errors made by the professionals versus those made by students were not addressed. Errors made were reported to cancel out over all subtests; however, the number of errors on the Information, Comprehension, and Vocabulary subtests varied significantly from the actual scaled scores determined for the fabricated protocols. The most common errors were failure to accurately assign point values to responses and improper discontinuation of subtests. The performance subtests, such as the Digit Symbol and Object Assembly, were also error prone with the scores determined by the examiners differing significantly from the actual scaled scores. The authors speculated that such errors might "well result

in misplacement or exclusion of an individual from a special program, and may make subtest profile analysis suspect" (p. 568). They recommended better training procedures for students and increased continuing education for professional psychologists to ensure greater proficiency in scoring accuracy of the new edition of the WAIS (WAIS-R).

Ryan, Prifitera, and Powers (1983) used WAIS-R protocols from two actual clients, as opposed to fabricated protocols, and had 20 graduate students and 19 psychologists (with an average of 7.3 years of testing experience) score them. They wanted to determine the scoring reliability of the WAIS-R and any scoring differences based on level of experience. The results indicated that 77% of the scores were reported to be within one SEM of the true scores for the psychologists and 88% for the students. Between the two groups there were no significant scoring differences on either protocol for the subtest or IQ means. However, greater scoring variance on the Performance scale results was noted for the psychologists and was attributed to poor attention to detail and a lack of clerical precision. The results suggested that scoring errors occurred frequently despite the examiner's level of experience in using the WAIS-R.

Slate and Jones (1990b) reported that when 26 graduate students scored 180 WAIS-R protocols an average of 8.8 errors per protocol was noted with errors reported on 177 (98%) of the protocols. There was no significant decrease in scoring errors over five practice administrations but improvement in the form of reduced scoring errors was noted after eight administrations, even though students still continued to make errors (mean error on the eighth protocol was still moderately high at 5.2 errors). The most frequent error was failing to record the examinee's response verbatim followed by incorrect point assignment, and inappropriate questioning. Other errors included incorrectly calculating subtest raw scores and chronological age and failing to obtain appropriate basal and ceiling levels. Most errors occurred on the Vocabulary, Comprehension and Similarities subtests although carelessness was cited as responsible for errors like circling incorrect point values.

Similarly, Slate and Jones (1990c) found that when 22 graduate students scored a total of 149 WAIS-R protocols, they made an average of 7.95 errors per protocol with errors on 145 of the protocols (97.3%). Most errors occurred on the Vocabulary,

Comprehension, and Similarities subtests including incorrect point assignments, failure to record responses verbatim, and inappropriate questioning of examinee responses.

A random selection of 50 WAIS-R protocols administered and scored by eight professional psychologists with extensive training in testing with the WAIS-R (range = 4-14 years) revealed errors on all protocols (Slate, Jones, Murray, and Coulter, 1993). An average of 36.9 errors per protocol was calculated when failure to record responses verbatim, circle scores or record times were all counted as errors. When they were not counted as errors, there were 15.4 errors per protocol. Most errors occurred on the Vocabulary subtest regardless of how the errors were counted. The most frequent error was failing to record examinee responses followed by inappropriate questioning, failing to query responses, and incorrect point assignment. Errors related to achieving correct basal and ceiling levels, calculating raw to subtest scores, and age calculation errors were also reported.

In summary, both professional psychologists and students were observed to make errors in administration and scoring on the WAIS and WAIS-R. Errors continued to surface on the adult version of the Wechsler Intelligence tests.

Summary of Scoring Errors on the Wechsler tests

In summary, studies examining scoring errors on the Wechsler tests reveal that numerous scoring errors are being made (to date, there are no studies investigating scoring errors for the latest version of the Wechsler Adult Intelligence Scale for Adults, the WAIS-III). Failure to record responses verbatim was cited many times. In addition, administration errors such as inappropriately questioning examinee responses and calculation errors such as, transforming raw scores to scaled scores and scaled scores to IQ scores, were frequently noted. Other administration errors noted somewhat less frequently included inaccurately beginning or discontinuing subtests. The verbal subtests were consistently reported as difficult to score, although the performance subtests were also reported to be error-prone. A tendency for examiners to engage in scoring generosity or to award too many points was noted, although less so for the WISC-III as compared to earlier versions of the test. The results of these studies suggest that professional psychologists are making as many, if not more, errors than student examiners in administration and scoring of Wechsler tests. Recommendations have

consistently included a need for modified and improved training practices for learning to administer and score the Wechsler tests.

The Need for Training Programs

The 1980's were a time when the value of psychological assessment training was questioned even though such training remained a core component of many clinical programs (Elbert, 1984; Kolbe, Shemberg, & Leventhal; 1985; Piotrowski & Keller, 1984). The type of training has also been subject to debate and there has been a continued emphasis on developing student examiner's testing and assessment abilities. Brabender (1991) noted a renewed interest in the field of psychological testing in terms of the training models used in psychological assessment with assessment and testing hold a central role in graduate psychology programs. During the 1991 conference in of The National Council of Schools of Professional Psychology, Peterson, McHolland, Bent, Davis-Russell, Edwall, Polite, Singer and Stricker (1991) presented a paper entitled *The Core Curriculum in Professional Psychology*. They emphasized the pivotal role of testing and assessment training in the preparation of psychologists. In 1993, Piotrowski and Zalewski found that 94% of the programs surveyed required some intellectual assessment coursework to be taught, and 88% of those surveyed predicted that little change would occur with respect to intelligence testing and that it would maintain a dominant role in the academic preparation. In commenting on the results, they stated that "these results clearly show that assessment is, and will continue to be, a critical factor in professional practice:" (p. 400). The American Psychological Association's Psychological Assessment Work Group (PAWG) responded to concerns by managed care groups in the United States who questioned the usefulness of psychological assessments. The PAWG group emphasized the need for such assessments and for highly trained psychologists to conduct them (Meyer et al., 1998). According to a survey of 374 school psychologists (37% response rate) in the state of Illinois, the WISC-R was viewed as the most useful intellectual assessment tool (Giordano, Schwiebert & Brotherton, 1997). The school counselors' responses to the survey indicated that they recognized the usefulness of standardized assessment instruments and perceived their primary role as administering tests and interpreting assessment results. They also noted a need for additional training in the area of assessment including assistance with scoring, administration, and

interpretation. These reports suggest that psychological assessment continues to be an important clinical activity that requires extensive training in order to produce well-trained professionals.

Training the Student Examiner

This next section contains information about training models and procedures specific to the administration and scoring of the WISC-R and WAIS-R that have been highlighted in the literature. Studies exploring the Competency-based Mastery Model are presented first in table format and summarized in Table 3.

Table 3.
Summaries of Studies Investigating the Competency-Based Mastery Model

Study	Wechsler Test	Sample	Details of interventions	Major Findings
1. Fantuzzo, Sisemore, & Spradline (1983)	WISC-R	8 graduate students	Mastery Model. Approx. 15 hrs. instruction time: -2 hrs study -2 hrs. lecture -3 observations of admin. (1 video, 2 live) rated by students using CCWA -3 admin. -2 feedback sessions	Pre-test admin. accuracy = 60% increased to 97% after training
2. Fantuzzo & Moon (1984)	WAIS-R	31 graduate students	8-10 hrs. instruction time: -2 hrs. study -1 hr. lecture -1 observation of video admin. -2 admin. -1 feedback session	Pre-test admin. accuracy = 60% increased to 95% after training.
3. Blackey, Fantuzzo, & Moon (1985)	WAIS-R	22 graduate students	8-10 hrs. instruction time: -1 hr. study -1 hr videotaped lecture -1 observation of video admin. -2 admin.	Pre-test admin. accuracy = 62% increased to 95% after training.
4. Moon, Fantuzzo & Gorsuch (1986)	WAIS-R	33 graduate students (exp. Grp.) 13 grad students (cntrl grp.)	Exp. grp = 8-10 hrs. instruction time: -2 hrs. study -1 hr. lecture -observation of video admin. -2 admin Cntrl grp = variance in training -2 admin.	Comparison Group accuracy = 59% increased to 67%; Mastery Group accuracy increased to 94%.

Study	Wechsler Test	Sample	Details of interventions	Major Findings
5. Blakey, Fantuzzo, Gorsuch, & Moon (1987)	WAIS-R	32 graduate students	Exp grp = -1 hr study WAIS-R manual -unltd study competency-based admin. training manual -past written admin test (1st post test) -peer training in admin. -unltd study competency-based admin scoring manual -pass written scoring test (2nd post test) -score WAIS-R with CCWS feedback -post test Cntrol grp = -2 hrs. study WAIS-R manual -1st post test -trained with exp. teaching method -2nd post test	Post-test 1 exp grp = 93% accuracy, cntrl grp = 63% accuracy. After receiving same peer training as exp.grp Post-test 2 = 96% accuracy.
6. Moon, Blakey, Gorsuch, & Fantuzzo (1991)	WAIS-R	33 graduate students	66 (2 protocols each)	20 items in CCWA failed by 50% and after more training, 11 items failed by 24%

The Mastery Model

The American Psychological Association's (1981) specialty standards in the disciplines of counseling, clinical, and school psychology called for a demonstrated level of competency within the area of psychological testing and assessment. In response, Fantuzzo, Sisemore, and Spradlin (1983) also emphasized the need for competency in the area of psychological assessment and, more specifically, in the administration of intellectual tests due to increased test use. They developed a model for teaching the necessary skills to accurately administer the WISC-R to graduate students. Following a task analysis of what was needed, they developed what they called the Criteria for Competent WISC-R Administration (CCWA). The CCWA is a performance checklist

consisting of 198 tasks distributed across 15 sections: introduction, conclusion, general considerations, and one section for each of the 12 subtests. Each task statement includes specific administration instructions for a particular subtest or a testing consideration (such as appropriately developing rapport and minimizing distractions). Accurate administration is calculated by dividing the number of tasks correctly performed by the total number of tasks per section. According to the authors, "this instrument details every aspect of the standardized administration procedure, including specific subtest-related points and general testing considerations" (p.229).

The authors used the CCWA with eight graduate students. They developed a competency-based procedure that involved two hours of study of the WISC-R manual, and one hour of pre-training administration where the students administered their first WISC-R and received both verbal feedback and feedback using the CCWA rating from their instructor. Students then watched a video administration of the WISC-R, rated it using the CCWA, and attended a two hour lecture about the major pitfalls of administering the test. They then completed their second WISC-R administration, which was again rated using the CCWA feedback. A post-training administration constituted the final and third WISC-R administration. In total, this competency-based procedure included students studying that WISC-R manual, observing and rating a video administration, attending two hours of lecture time, conducting three of their own administrations and receiving feedback using the CCWA, all within a 15 hour time period. The results indicated good inter-observer agreement between two graduate students who observed both the student examiners administer the pre- and post-test administrations. Post-test accuracy increased to above the set minimal competency criterion of 90%. A significant positive difference was reported between pretest (mean score = 60%, SD = 12.5) and post-test (mean score = 97%, SD = 2.06) administrations. Areas of difficulty for the students included establishing and maintaining a functional testing atmosphere. The Comprehension and Vocabulary subtests were the most difficult verbal subtests to administer mainly due to failures in cueing ambiguous responses. The Block Design and Picture Arrangement subtests were the most inaccurately administered performance subtests due to a lack of adherence to standardized instructions and procedures for manipulating the stimulus materials. Even though the sample size was

small (n = 8), the increase in accurate administration and scoring after use of this competency-based training model that included the use of the CCWA and a 15 hour instructional program, suggested its potential use as a WISC-R training instrument.

The competency-based training procedure (now called the Mastery model) was modified by Fantuzzo and Moon (1984) and used to teach graduate students administration skills for the WAIS-R. The CCWA (Criteria for Competent WAIS Administration) instrument was used. The training time was reduced to eight hours with only two WAIS-R administrations. As part of the training procedure, students were required to study the WAIS-R manual for two hours, administer the WAIS-R and receive both verbal feedback and feedback on the CCWA rating from their instructor. Students then attended a one hour lecture focused specifically on major administration difficulties, and then observed and rated (using the CCWA) one video-taped administration of an advanced student modeling an accurate administration of the WAIS-R, followed by a second WAIS-R administration. Similar to the competency-based model used by Fantuzzo et al., significant positive differences were reported between the pretest accuracy results (mean = 59.5%, SD = 8.44) and post-test accuracy (mean = 95%, SD = 2.46). The sample size used in this study was larger (n=31). A combination of observation, instruction, and practical application of accrued knowledge stemming from the feedback from the CCWA seemed to have increased students' skills in correctly administering and accurately scoring the WAIS-R.

Refinement of the Mastery Model

Blakey, Fantuzzo and Moon (1985) subsequently automated the Mastery model and investigated its effectiveness. A video-taped lecture replaced the live instructor delivering the lecture about the major difficulties in administering intelligence tests and a video-taped administration of the WAIS-R replaced the instructor conducting a live WAIS-R administration. Students studied the WAIS-R manual for only one hour prior to their first administration. After receiving feedback from graduate laboratory assistants who used the CCWA, the students viewed the video-taped WAIS-R administration and a video-taped lecture and were then required to demonstrate their accuracy in scoring by detecting 80% of the errors on the videotaped demonstration. A fellow student, using the CCWA, evaluated their second WAIS-R administration. Overall, similar improvements

in administration and scoring accuracy were reported through the use of this automated model. Pretest percentages of accurate administration were an average of 62% with post-test accuracy percentages reaching an average of 94.5%. Establishing and maintaining a functional testing atmosphere and departing from standardized procedures were the main sources of error. The Comprehension and Vocabulary subtests remained difficult to accurately administer due to students' uncertainty in scoring ambiguous responses. The Block Design, Picture Arrangement, and Picture Completion subtests were cited as the most difficult performance subtests to administer. The authors sought student feedback to the automated training program through a student examiner questionnaire and favorable results were reported. The automated format appeared to be a cost-effective, efficient way to help students achieve a good level of competency in administering the WAIS-R.

A comparison was made between the Mastery model and other existing training models by Moon, Fantuzzo, and Gorsuch (1986). The experimental group of 33 students followed the Mastery model while the control group of 13 students adhered to training models available from their APA approved internship training sites. Both groups received a similar amount of overall training and study time (eight to ten hours) and administered the WAIS-R (Pre-test) prior to any intervention. There were no reported significant Pre-training effects as assessed through accuracy of administration for both groups of students using the CCWA for their first WAIS-R administration. The Mastery group received the lecture focused on particular errors of the WAIS-R and rated one video administration of the WAIS-R using the CCWA. The control group conducted more WAIS-R administrations overall, received more structured didactic training regarding the WAIS-R, and also experienced a greater time lag between their Pre and Post-test administrations. The Mastery group showed significant improvements across all areas when compared to the comparison group and achieved a higher level of Post-Test accuracy (94% for the Mastery group compared to 67% for the control group). The Mastery model was also found to be more cost-effective than the internship training models.

To reduce instructional costs further, Blakey, Fantuzzo, Gorsuch, and Moon (1987) employed student peers instead of instructors to train students to competently

administer and score the WAIS-R. Students in the experimental group studied both the WAIS-R manual and the CCWA training manual for the for administration and had to pass a written test (with 90% accuracy) prior to being considered competent to act as peer examiners to their colleagues. The students in the experimental group then monitored each others' WAIS-R administration ability using the CCWA, thereby serving as each others' peer trainers. The authors developed a second performance checklist they called the Criteria for Competent WAIS-R Scoring (CCWS) used to assess adherence to standardized scoring procedures of the WAIS-R. The CCWS instrument consists of 80 tasks distributed across 14 sections- one section for each of the 11 subtests in addition to the introduction, conclusion, and general considerations sections. Each task statement included specific scoring instruction for a particular subtest or a testing consideration. Accurate administration was calculated by dividing the number of tasks correctly performed by the total number of tasks per section. Students in the experimental group studied the CCWS manual for scoring and had to pass a written test (with 90% accuracy) to show competency as a peer trainer. The students then evaluated each others' scoring of a sample WAIS-R using the CCWS. Finally, the experimental group completed a post-test. The control group of students read over the WAIS-R manual and completed a post-test.

The Experimental group achieved 93% scoring accuracy percentage on the post-test and the control group achieved 63% accuracy percentage. After the control group had received similar training as the experimental group, their accuracy percentage on the second post-test increased to 96%. In terms of scoring accuracy, the experimental group achieved 95% accuracy percentage and the control group 92% accuracy percentage indicating no significant differences between the CCWS scores for the experimental and control groups; however, the experimental group was noted to have achieved a higher level of scoring accuracy on the Comprehension and Similarities verbal subtests that were more prone to causing scoring errors according to previous results. In a questionnaire asking about their experience with the peer-mediated approach, the students indicated their satisfaction with peer-mediation as part of the teaching method. This study indicates the potential usefulness of the Mastery approach to assist students achieve a level of administration and scoring accuracy using a peer training approach. The cost

effectiveness was determined to be even better than the automated version noted previously.

Moon, Blakely, Gorsuch, and Fantuzzo (1991) reanalyzed the results from Moon, Fantuzzo et al. (1986) to determine the most common WAIS-R administration errors made by students from the control group versus students who had engaged in the training using the Mastery Model. All students from the control group failed at least 50% of the administration requirements listed in the WAIS-R manual with only two hours of studying the WAIS-R manual before test administration. After receiving 8 to 10 hours of extra training, significant improvement in administration resulted but 24% of students continued to fail 11 specific administration requirements. Examples of the most common administration errors for both groups of students included incorrect recording of time and timing errors, failing to maintain rapport, and incorrect placement of stimulus items.

Other training interventions

Other training interventions were discussed in the literature like the use of practice administrations to increase efficient and accurate administration and scoring of the Wechsler tests and a specific lecture intervention aimed at reducing the number of errors that seemed to emerge with regularity, as noted by previous studies. Table 4 contains a summary of these other training interventions.

Table 4. Other Training Interventions

Study	Wechsler Test	Sample	Details of interventions	Major Findings
1. Slate & Jones (1989)	WISC-R	19 graduate students	Approx. 15 hrs. instruction time Exp & Cntrl grps: -2 hr observation admin WISC-R -1 hr. lecture Cntrl grp: -practice admins and class discussion Exp. grp: -2 more hrs. lecture time -7 practice admins-peer checking protocols -8th admin observed & graded by instructor	Exp.grp ξ= 3.40 errors per protocol (SD = 2.74); cntrl grp ξ = 8.10 errors per protocol (SD = 6.05). No decrease in errors across 7 protocols.

Study	Wechsler Test	Sample	Details of interventions	Major Findings
2. Slate, Jones, & Murray (1991)	WAIS-R	20 graduate students	150 protocols (approx. 8 protocols each) Both exp & cntrl grps: -Observation of WISC-R and WAIS-R admin and scoring Exp grp: -5 WISC-R admins prior to 10 WAIS-R admins Cntrl grp: -10 WISC-R admins prior to prior to 5 WAIS-R admins	5 administrations = 39 errors per protocol. When failure to record excluded ξ = 16.96 errors per protocol (SD = 7.9). 10 administrations = 24.1 errors per protocol. When failure to record excluded ξ= 13.8 errors per protocol (SD = 5.6). No decrease in errors across 10 protocols.
3. Slate, Jones, & Covert (1992)	WISC-R	20 graduate students	150 protocols (approx. 8 protocols each) Both exp & cntrl grps: -Observation of WISC-R admin and scoring Exp grp: -5 WISC-R admins Cntrl grp: -10 WISC-R admins	5 administrations = 11.8 errors per protocol. When failure to record excluded ξ= 10.6 errors per protocol (SD = 4.9). 10 administrations = 43.8 errors per protocol. When failure to record excluded ξ= 16.5 errors per protocol (SD = 7.5). No decrease in errors across protocols for either groups.
4. McQueen, Meschino, Pike, & Poelstra (1994)	WISC-R	75 graduate students	Grp 1=1988,Grp 2=1989,Grp 3=1990 Lecture intervention= weekly quizzes Lab intervention= lecture, extra practical instruction, peer checking protocols for errors.	Lecture and lab helped to increase scoring and administration accuracy.

Similar to the Mastery model format, Smith and Harty (in Dana & May, 1987), provided suggestions for supervisors working with students in diagnostic testing. They emphasized the need for collaboration between student and supervisor with feedback given along all stages of learning. While they did not advance a structured model like the Mastery model, they did offer the suggestion that supervision of test administration and scoring occupy a greater amount of time at the onset of training, stating that "without good data, even the most sophisticated analysis is crippled, and correct techniques are essential to the validity of test data" (p. 413). They described how the supervisor could relinquish the didactic/supervisory role and assume more of a consultant role as the student becomes increasingly familiar and competent in the administration, scoring and interpretation of test data.

Slate and Jones (1989) training intervention was centered on a detailed lecture about frequent errors and ways to avoid making these errors. Both the experimental and control groups observed a live demonstration of a WISC-R administration and took part in a general lecture on administration and scoring procedures. The experimental group then received an extra two hour lecture on frequently committed administration and scoring errors made by students on the WISC-R with instruction about how to avoid making these errors. While the experimental group received the extra lecture time, the control group began engaging in practice administrations. Both groups completed eight practice administrations with the final administration observed and graded by the instructor. Both groups of students were also each assigned a partner who checked over their WISC-R protocols prior to submitting them for grading and each received written feedback and oral feedback in the form of classroom discussion. The experimental group ($n = 9$) committed fewer administration and scoring errors than the control group ($n = 14$) as determined by the percentage of IQs that students assigned which had to be changed in accordance with the instructor's corrections. However, none of the students from either group decreased the number of errors they made after seven administrations. This result led the authors to comment that "when giving practice administrations, students appear to practice mistakes rather then to improve skills" (p.409).

Slate, Jones, and Murray (1991) conducted an evaluation of the effectiveness of practice administrations for the WAIS-R to determine if more practice on the WISC-R

would result in fewer errors on subsequent administration of the WAIS-R. All students had previously received similar training that included an observation of both a WISC-R and WAIS-R administration and a demonstration of scoring for each test. All students received verbal and written feedback regarding their performance from a graduate assistant. Students were randomly assigned to one of two groups with the first group administering the WISC-R five times followed by ten WAIS-R administrations. The second group administered the WISC-R ten times followed by five WAIS-R administrations. The extra practice for those students who administered the WISC-R ten times did not result in a reduction of subsequent WAIS-R errors but rather a repetition of similar types of errors on the WAIS-R. This negative transfer of errors occurred as students continued to make errors in converting raw scores to standard scores. The belief that students would transfer knowledge from their WISC-R administrations to the WAIS-R administrations was not borne out. No significant decrease in errors occurred after the ten administrations, except that students improved their tendency to record their responses. Most errors were made on the verbal subtests, particularly the Comprehension and Similarities subtests. Mechanical/calculation or "careless" errors, such score conversion errors, errors in questioning and establishing basals and ceilings, and failing to record responses were also more frequently noted.

Slate, Jones, and Covert (1992) investigated the effectiveness of practice administrations to reduce administration and scoring errors on the WISC-R. Twenty students observed a WISC-R administration and received instruction regarding scoring procedures. Ten students complete five WISC-R's and ten students complete ten WISC-R's. Both groups received oral and written feedback from a graduate assistant for each of their practice administrations. There was no decrease in errors across five or ten WISC-R administrations. The most frequent errors were failing to record responses verbatim, incorrect point assignments, and carelessness in terms of calculation errors (e.g., incorrect total raw score and chronological age calculations). The majority of these errors occurred on the Vocabulary, Comprehension, and Similarities subtests.

McQueen, Meshino, Pike, and Poelstra (1994) compared the use of both a lecture and lab intervention to improve students' assessment performance. This study occurred over a three-year period with the first group in the first year serving as the control group,

the second group in the second year receiving the lecture intervention, and the third group in the third year receiving the lecture and lab intervention. The lecture intervention consisting of weekly quizzes to assess knowledge accrued during the lecture portion of the course. The laboratory intervention was multifaceted. It included specific instructional information for eliminating common WISC-R and WAIS-R scoring errors (in verbal and printed format with a focus on reducing the verbal subtests errors on the Vocabulary, Similarities, and Comprehension subtests), review of scoring criteria and tables used, practice with correct chronological age calculations and students using the *WISC-R Scoring Criteria Supplement* (Massey et al., 1978) to score the verbal subtests. The final laboratory intervention included the students pairing up with a partner who reviewed their protocols and reports for accuracy. The laboratory instructors used an instructional checklist to ensure adherence to the experimental intervention as designed.

Assessment performance was calculated by the percentage of total points earned by each group from a possible 500 points. Group two (lecture intervention) and group three (lecture and laboratory intervention) achieved higher means than group one (no intervention) and group three also achieved a higher mean compared to group one. These results suggest that the lecture and combination lecture and laboratory intervention were effective for improving scoring accuracy and administration of the WISC-R and WAIS-R.

Summary of Training the Student Examiner

A review of these studies provides suggestions for ways to increase the effectiveness of teaching methods that could result in more accurate administration and scoring practices on the Wechsler tests. The Mastery, competency-based model developed by Fantuzzo et al. (1983) offers a means to assess accuracy through the 0CCWA and CCWS performance checklists. The use of checklists appears helpful in drawing attention to error-prone areas. The use of practice administrations alone as a means of increasing administration and scoring proficiency has failed to produce such a result although, there was some success with practice administrations paired with immediate and consistent feedback to students. Specific and detailed lectures regarding the frequency of examiner errors and scoring pitfalls were reported to help increase correct administration and scoring accuracy. These lectures seem to have effectively

alerted students to these administration and scoring problems and reduced these types of errors. A separate laboratory component where instruction and discussion about such errors is made showed promise. The value of peer training was demonstrated with peers serving dual roles as co-instructors and overseers of correct administration and scoring. The addition of weekly quizzes was noted as effective in ensuring a consistent level of knowledge for students and competency exams were listed as helpful for ensuring that students reach a prescribed level of competency.

Characteristics of the Student Examiner

Ideally, the best training model should help a student clinician develop certain characteristics deemed important in conducting psychological assessment work. The nature of students desiring a career in the area of assessment and diagnosis and the characteristics viewed as necessary from the perspective of professionals in the field are explored in more detail in this section. There may be certain characteristics that make some examiners better suited to psychological assessment and diagnostic work. A better understanding of these traits and characteristics could assist clinical trainers to develop training programs best suited to these student examiners' needs and help them to hone their skills. Better-trained examiners can also perform more accurate assessment and diagnostic work with fewer scoring errors.

Critical student examiner characteristics, as deciphered from the literature, include an examiner's basic skills in testing and assessment. Of primary importance is the examiner's familiarity with test material, and this familiarity can be developed through reading, practice, and evaluation prior to actual administration (Kamphaus, 2001; Sattler, 1992). The student examiner must have a high interest in the assessment and diagnostic area and be willing to learn the information. Related to the examiners' administrative ability is the ability to select and administer tests appropriately (Aiken, 1996; Slatter & Thomas, 1983). In addition to competent testing and assessment skills and acquisition of knowledge is the examiner's capacity to communicate in a clear manner with others. Effective communication throughout the assessment process needs to occur with a variety of interested parties such as children, parents, school personnel, and other professionals. The student examiner should have the ability to communicate

assessment findings in a concise and understandable manner (Nietzel et al, 1991; Smith & Harty in Dana & May, 1987).

Achieving and maintaining a high level of creativity is another important characteristic for the student examiner for two reasons. The first is that diagnostic work can become repetitive and boring after administering the same tests on a frequent basis and using similar report formats over time (Woody & Robertson, 1988). The second reason is that an examiner needs to sustain and replenish creativity in order to ensure a high level of professionalism (Gregory in Cullari, 1998). For example, maintaining sensitivity toward the client with a renewed level of interest and energy in the assessment area ensures high quality outcomes. This form of creativity may emerge, to a certain extent, from training and experience, but may also be a unique trait that characterizes students particularly suited to this work. As Woody and Robertson stated, "Creativity is the hallmark of high quality diagnostics" (1988, p. 283).

With respect to working with children, Nietzel, Bernstein and Milich (1991) suggested that a non-judgmental, pleasant and neutral presentation by the student examiner is important in order to allow the child's attitude and reactions to the testing situation and materials to emerge. For novice examiners attempting to acquire clinical skills in working with children, Kamphaus (2001) suggested that they become familiar with children through a process of observation. In addition to observing children to recognize age-appropriate behavior and to become acquainted with developmental milestones, he suggested that students observe experienced psychologists administer intelligence tests to both normal and disabled children to better learn the nuances of testing required for both types of children. The student examiner should be patient in his or her approach with children and retain a sense of humor (Sattler, 1992). In sum, Sattler stated that "a competent examiner must be flexible, vigilant, and self-aware and must genuinely enjoy working with children" (p.112).

The need for an examiner to be self aware was emphasized by Nietzel et al. (1991) who stated that a process of self-evaluation could " help the student uncover biases, attitudes, and personality problems that might interfere with later clinical work" (p.404). Having an awareness of personal biases is important in terms of alerting student examiners to those influences that affect personal judgment (Sattler & Thomas, 1983).

Aiken (1996) stated that human beings are not completely objective when they score tests. The bulk of research about examiners making scoring errors on the Wechsler tests supports this assertion. Aiken (1996) asserts that the bulk of research about examiners making scoring errors on the Wechsler tests indicated that examiners are not completely objective when they score these tests. Increased awareness can assist student examiners to become aware of their own predictive inaccuracies and the limits of their clinical judgment. Hunt (1946) long ago emphasized the importance of clinical judgment for student examiners. He stated that there were both "good and bad clinicians" and that clinicians needed to be evaluated on a continuous basis to ensure their ability to make good judgments and to adequately perform diagnostic work. More recently, Garb (1998) noted that psychologists do not attend closely to empirical research findings and tend to rely on their own judgments, which are not always accurate. Furthermore, being mindful of one's responses to others was also listed as important for student examiners in terms of countertransference "in which personal reactions can influence, distort and derail the feedback process" (p.44) Gregory (in Cullari, 1998). A student examiner is therefore one who is willing to engage in a process of self- analysis and be open to critical self-examination in order to increase his or her ability to work efficiently with others.

Diagnosticians require intelligence and common sense to properly interpret the varied sources of data collected during the assessment process, as well as the competence to realistically understand their professional abilities and limitations. Linked to this level of competence is the need for confidence, which is necessary in order to make judgments that may have a significant and long-term impact on people's lives (Woody & Robertson, 1988). In sum, these authors stated that "Confidence, competence, and creativity join with academic knowledge, professional experience, and the clinical psychologist's personality to bring about the sagacity that distinguishes being a diagnostician from other psychological roles and from the opinions of laypersons" (p.283). Kamphaus (2001) mused about the clinical skill involved with assessment and diagnostics by commenting that "clinical skill may be a personality related factor leading to the old question 'are clinicians born or are they trained?'" (p.94).

In summary, this information suggests a number of characteristics that student examiners require to achieve a high level of competency and to adequately learn and

practice in the area of assessment and diagnostic work. Student examiners require a solid base of skills in testing and assessment, intelligence and common sense, all of which may be amenable to being developed and honed by thorough training and practice at the graduate level. In addition, they require a strong ability to communicate information to concerned parties and to feel confident in the material they present. Learning through lecture format, individual practice and through an interactive method might help student examiners to better amalgamate such information. Increased self-awareness can occur through feedback from supervisors but could occur on a more personal level by assisting student examiners to be aware of their personal limitations.

This review of the literature provided a summary of previous research of administration and scoring errors and on the WISC-R, WISC-III, WAIS-R, and WAIS-III in addition to teaching models and interventions to assist in reducing such errors. Student examiner characteristics deemed important to develop for professional competency in the area of assessment were also reviewed. The next chapter contains information about the methods and procedures used in this research study.

CHAPTER THREE

Methods and Procedures

The methods used and the procedures followed to address the research objectives are presented in this chapter. The participants, setting, and ethical considerations are first described followed by a description of the teaching methods used to teach the students. The research design employed to evaluate the effectiveness of the two teaching methods in terms of the objectives of the study is then described. Lastly, the procedures for calculating the scoring errors are described.

Participants and Setting

This study was conducted in the Education Clinic in the Faculty of Education at the University of Alberta. The Education Clinic is a training facility for graduate students. The participants were 30 graduate students enrolled in the Individual Assessment Course offered by the faculty in the clinic. The course instruction and the associated practicum are offered in the clinic. During the lecture component, students are taught about the broad area of psychological assessment, including the administration and scoring of tests. During the practicum component, students practice testing using actual clients, both children and adults, who pay for this service. The course capacity for the individual assessment course is 24 graduate students, and the course is offered once every calendar year during the full-year Winter session and every second year during the four-month Spring/Summer session.

Ethical Considerations

The Research and Ethics Committee judged this research to have met acceptable ethical requirements. An information sheet was developed by the researcher and read aloud to the student examiners to ensure that all necessary ethical information for requesting participation was shared with them. None of the student examiners declined to participate and the researcher was not contacted for further information or questions. A copy of the Ethics Sheet and the Ethics Approval Sheet are provided in Appendix A.

Sample

Approximately half of the student examiners for this study completed the course during the Winter 1998-99 session (n =16) and half completed the course during the Spring/Summer 1999 session (n =14). The sample of student examiners from each one of

these sessions was separated into either the control or experimental group (i.e., the usual teaching method or the experimental teaching method). A lack of Wechsler testing kits available at the Education Clinic meant that two students had to share one kit. When asked to separate into the different groups, some students remained with their partners in one group while others split from their partners in order to form equal groups. As such, there was no specific randomization of the groups.

During data processing procedure, nine students were deleted because of missing protocols. Four students from the Winter 1998-99 session were deleted. Two of these students had exceeded the number of required Stanford-Binet Fourth Edition (SBIV) administrations (i.e., five administrations instead of the required four administrations). When a student completes more of one intelligence test than another, the student is allowed to complete fewer of the other test. The remaining two students did not complete one WAIS-III protocol. One of these students had previously registered in the 1997-98 session and had completed half of her protocols during that session and half during this session.

Of the five student examiners from the Spring/Summer 1999 session who were deleted from the study due to missing protocols, three did not complete one WAIS-III administration and two did not complete two WAIS-III administrations. Only one of these students had completed an extra SBIV.

In total, 16 students in the experimental group and 14 students in the control group completed 4 WISC-III and 4 WAIS-III protocols. The final sample of students for the two groups were therefore the same. Thus, there were 240 protocols examined for administration and scoring errors.

A description of the final sample of student examiners is presented in Table 5.

Table 5.
Demographic Information of Student Examiners

Session	**1998-99**		**1999**	
Group	Experimental	Control	Experimental	Control
	9	7	7	7
Gender	1 Male/ 8 Females	0 Male/ 7 Females	1 Male/ 6 Females	1 Male/ 6 Females
Education Level	8 Master's 1 Ph. D.	5 Master's 1 B. A./1 Ph.D.	7 Master's	5 Master's 2 Ph. D.
Position	5 Students 4 School pers.	4 Students 3 School pers.	2 Students 5 School pers.	1 Student 6 School pers.
Prev. Exp with Intelligence testing	8 = None 1 = Undergrad	6 = None 1 = Undergrad	4 = None 3 = Undergrad	5 = None 2 = Undergrad
Prev. Exp with other testing	2 = None 7 = Acad/Voc	3 = None 4= Acad/Voc	7 = Acad/Voc	7 = Acad/Voc
# Prev. Assessment courses	2 = None 7 = 1-2	4 = None 3 = 1-2	1 = None 6 = 1-2	1 = None 6 = 1-2

Winter1998-99 Sample

For the 1998-99 session, there were 15 females and one male. Thirteen were Master's level students, 2 were Ph.D. level students, and one was a Bachelor of Arts level student who had received special permission to take the course. Seven student examiners identified themselves primarily as working within a school system (i.e., teacher, school counsellor, or administrator), while nine identified themselves primarily as graduate students. In terms of pervious experience with testing, 11 students had administered academic or vocational tests previously, and two had some previous knowledge with IQ testing in their undergraduate courses (i.e., reading and learning about intelligence tests). Ten student examiners had taken previous assessment courses, such as introductory level assessment courses that were based on reading and learning about these tests.

Spring Summer 1999 Sample

For the 1999 session, there were 12 females and two males. Twelve of the student examiners were Master's level students and two were Ph.D. level students. Eleven student examiners identified themselves primarily as working within the school system while three identified themselves primarily as graduate students. Nine student examiners had no previous experience while five had some undergraduate knowledge of

intelligence testing. All participates had some experience with academic or vocational testing and 12 student examiners had taken previous assessment courses while two had no such experience.

Teaching Methods

The first objective of this study was to compare two different methods of teaching the intelligence tests and to determine if one method would result in significantly fewer errors. A revised or experimental teaching method was devised for this study based on suggestions from previous research on training students to properly administer and score the Wechsler tests (Blakey, Fantuzzo, & Moon, 1985; Fantuzzo, Sisemore & Spradlin, 1983; Gorsuch & Fantuzzo, 1991; McQueen, Meshino, Pike, & Poelstra, 1994; Moon, Blakey; Slate & Jones, 1989). The control method consisted of the usual teaching method. Table 6 outlines the experimental and control teaching methods.

Table 6.
Experimental and Control Teaching Methods

Experimental Teaching Method	Control Teaching Method
1. Introductory lecture	1. Introductory lecture
2. Practice administration	2. Practice administration
3. Lecture on Administration and scoring	3. Lecture on Administration and scoring
4. * Lecture on specific and frequent administration and scoring errors	4. Practice administration
Experimental Teaching Method	Control Teaching Method
5. * Observation of a Wechsler administration	5. Practice administration
6. Interpretation lecture	6. Interpretation lecture

* denotes the two experimental teaching interventions

The experimental teaching method included two teaching interventions: the first was a specific lecture about the areas of difficulty when administration and scoring the tests with a focus on helping students to become aware of these scoring errors (particularly on the more difficult to score verbal subtests). The second intervention was a demonstration of a WISC-III/WAIS-III administration where students scored their own record form while observing the administration. This second intervention was aimed at helping students to learn about the correct administration and scoring through observation

of a live demonstration. The students recorded their responses on the record form and compared them to the actual record form, which was followed by interactive discussion about any administration or scoring questions.

The control teaching method was the instructional format historically used in this assessment class. This teaching method placed more emphasis on practice administration. There was no specific lecture about administration and scoring errors nor was there a live Wechsler demonstration.

While the experimental group was taking part in the two teaching interventions, the control group practiced administering the tests. All other instruction was the same for both groups. A description of the classes follows.

1. Introductory lecture

Both the experimental and control groups took part in this introductory lecture that included historical information about the development of the Wechsler tests, reasons for renorming the tests, changes in the addition/omission of certain subtests, and information about the standardization sample. An overview was provided about the subtests, the Index, and IQ scores with an emphasis on explaining the different types of scores (i.e.: derived scores, percentile ranks, standard scores) and the reliability and validity of the scores yielded by the tests.

2. Practice administration

The experimental and control groups were separated while practicing the administration of the WISC-III/WAIS-III. The control group spent three class periods practicing administration while the experimental group spent only one class in practice administration.

3. Lecture on administration and scoring

The experimental and control groups were separated during the administration and scoring lecture. During this lecture, information was provided about completing the record form, which tables to consult to obtain scores, a description of the individual subtests, and additional information about administering the tests such as explanations about querying responses and how to address multiple and spoiled responses.

4. Lecture on specific/frequent administration and scoring errors

Only the students in the experimental group received this lecture while the students in the control group were completing their second practice test administration. The experimental lecture consisted of instruction about commonly encountered errors of administration and scoring. For example, errors in establishing correct basal and ceiling levels and errors when cueing for responses.

5. Observation of a Wechsler administration

Only the experimental group observed the instructor and a volunteer demonstrate the administration of a WISC-III/WAIS-III using a script of set responses. This demonstration took place in the classroom and the students observed the administration, recorded their responses on their protocol, and scored the protocol. A discussion about administration and scoring followed the test administration. The corresponding activity for the control group students was completion of the third practice administration.

6. Interpretation lecture

Both the experimental and control groups took part in the interpretation lecture as one group. The topics in this lecture were checking the validity of the results, profile analysis of the pattern of scaled scores and deviation IQ's, and learning about the many different types of comparative procedures used to analyze results (e.g., comparing verbal and performance IQ's, comparing each verbal or performance subtest scaled score with the mean verbal or performance scaled score, and/or comparing pairs of individual subtest scaled scores).

During the three common lectures (i.e., the introductory lecture, the administration/scoring lecture, and the interpretation lecture), the students were directed to chapters, appendices, and tables in Sattler's textbook *Assessment of Children* (1992) for a more detailed account of the information taught in class.

Instruction Time

During the 1998-1999 Winter session, there was one and a half hours of instruction for each of the six class times available to teach the WISC-III (9 hours in total) and five class times to teach the WAIS-III (7.5 hours in total). During the 1999 Spring-Summer session, there was one full day of class time (5.8 hours in total) and one evening of class time (2.75 hours in total) available to teach both the WISC-III and the

WAIS-III. The total amount of instruction time for teaching both tests during the Winter session was 16.5 hours and during the Spring-Summer session was 17.1 hours.

Teaching Agendas

A series of five teaching agendas was provided to the instructors to help ensure that the same course content was being taught to each of the experimental groups and each of the control groups. They provided the instructors in each group with the lecture information and timing of instruction for their group. The first teaching agenda, a general teaching agenda provided instructors with an overview of the teaching instructions for the experimental and control groups. Four more detailed teaching agendas, three of which were used with both the experimental and control groups, were provided to instructors to assist them in teaching the particular lectures. The first teaching agenda was used for the introductory lecture of the test, the second teaching agenda was used for the lecture on administration and scoring, and the fourth teaching agenda was used for the interpretation lecture. All students learned the same information from these three lectures. The third teaching agenda was used only with the experimental group and consisted of a lecture on specific and frequent administration and scoring errors. This teaching agenda included extra information on administration and scoring rules with a focus on specific administration and scoring errors, particularly for the verbal subtests. The experimental teaching agenda also included the Wechsler administration information for the live Wechsler administration.

The Teaching Agendas are provided in Appendix B.

Wechsler Instructors

The two instructors who taught the experimental group were both female Ph.D. candidates registered in the Counselling Psychology program. For the control group, the one instructor was a female Ph.D. candidate in the Counselling Psychology program while the other instructor was a male Professor and Director of the Education Clinic at the University of Alberta. All instructors had experience with the Wechsler tests and all had completed the Individual Assessment course. In addition, these instructors had at least two years of practical experience in the area of assessment (i.e., a range of 3-4 years for the experimental instructors and range of 2 to 25 years for the control instructors). There was continual contact between the researcher and the instructors during the

research period to assist them with instruction and to answer questions about the corresponding teaching agendas.

Student Supervisors

Graduate students enrolled in the Individual Assessment course are assigned to a supervisor, who is a professional psychologist knowledgeable and experienced in intellectual assessment. The supervisor oversees the student's work by grading the student's test protocols and psychological reports and signs the final report prior to its presentation to the client. A supervisory meeting takes place at the beginning of each session with all new and returning supervisors. During this meeting, the Clinical Director of the Education Clinic informs supervisors about the expectations of the supervisor, the expectations for the students in the assessment course, and the procedures to be followed in the Education Clinic. Each supervisor is expected to be a resource person and mentor to the students they supervise. The supervisors orient the students by arranging meeting times to answer students' questions, to discuss the assessment process, including testing and diagnostic considerations, and to monitor the students' progress. The supervisor provides both verbal and written feedback to the student. Supervisors are not deliberately matched to students but randomly assigned. One or two students are assigned to each supervisor depending on the availability of supervisors. During the Winter 1998-99 and the Spring/Summer 1999 sessions, 24 supervisors were available for the students.

Research Design

A non-equivalent control group design (Campbell & Stanley 1966) was the research design used in this study. Student examiners had registered in the Individual Assessment class of either the Winter 1998-99 session or the Spring/Summer 1999 session. The students from both sessions were divided into either an experimental or a control group. The independent variable was the teaching method. Students in the control group were instructed using the current teaching method while students in the experimental group were instructed using the new teaching method.

Threats to Internal Validity

Campbell and Stanley (1966) emphasized the importance of theoretical knowledge and probable explanations to account for results from research using the quasi-experimental designs. Cook and Campbell (1979) concluded that sound findings

could be drawn from non-equivalent control group studies when the groups were different as long as threats to internal validity were carefully evaluated.

The difficulties with non-equivalent groups is the initial concern that the groups may differ on the dependent measure(s) at the beginning of a study and that the groups may be different on variables other than the dependent variable that may influence the dependent variable (Raulin & Graziano, 1995). Demographic information, including gender, student status, previous experience with intelligence testing and other forms of testing, and previous coursework in testing, was collected from the student examiners and examined to provide some indication of group consistency.

A second selection threat is the possibility that the four instructors who taught the Individual Assessment course to the experimental and control groups might have taught the course material in different ways rather than following the prescribed teaching method. To avert this threat to the internal validity of the study, teaching agendas for both groups were developed and provided to the instructors to ensure that the same course content was being taught by the instructors of each of the experimental and control groups.

A third selection threat is the possibility that students from the experimental group might inadvertently divulge information to the students in the control group about the additional information being taught to the experimental group. This diffusion of treatment is a potential confounding variable.

Scoring Errors

The second objective of this study was to determine the number and types of errors committed during the administration of the Wechsler tests and the student and the supervisor had scored the protocols for errors. Determining if the experimental group or the control group committed fewer administration and scoring errors was investigated.

Administration and Scoring Checklists

Scoring checklists were developed for the WISC-III and the WAIS-III. These checklists provided information about the number, type, and frequency of scoring errors that were committed by the student examiners. Previous researchers developed such checklists for their studies. For example, Fantuzzo et al. (1983) developed scoring checklists for the WISC-R and the WAIS-R called the Criteria for Competent WISC-R

Administration and WAIS-R Administration (CCWA). Supervisors used these scoring checklists while observing students administer the tests. The checklists were used to record both examiner errors made during a live administration of the test and scoring errors made on the protocols.

The present study involved scoring more recent versions of the WISC (WISC-III) and the WAIS (WAIS-III) but only the protocols were scored, not the actual live administration of the test. Fantuzzo et al.'s checklists were therefore not used. Instead, Alfonso et al.'s (1998) checklist was adapted for use. They developed a WISC-III administration and scoring checklist for their study that only involved scoring student examiner's WISC-III protocols for administration and scoring errors. Two categories of errors were considered; the first included general errors that could occur on all the subtests and on the front and back pages of the protocol. The second category of errors included specific errors for each subtest such as failure to query, failure to record responses verbatim, and failure to record time.

The checklist modifications for this study included a more detailed analysis of the errors. The Subtest errors were modified to include failing to begin the subtest at the correct starting point, incorrectly establishing the ceiling level, incorrectly discontinuing and exceeding the subtest, and incorrectly adding subtest scores. The "Pages" errors was modified to include error categories like incorrectly using American norms and more detailed errors when transferring and charting subtest, Index, and IQ scores to the graphs.

The checklists were also modified to include information about missing data (e.g., responses not recorded verbatim, failing to record demographic information, tables left blank, or optional procedures not completed). Since the students were learning how to administer and score these intelligence tests, it was assumed that they would complete the protocols in their entirety. Following the approach of Slate et al. (1991), it was decided to separately examine the number of scoring errors and not recorded or missing data errors. As these authors stated, "Although failures to record are not best practice, they do not necessarily affect the accuracy of test scores" (p. 376). Therefore, both checklists contain information about missing data. Both the WISC-III and the WAIS-III checklists incorporate missing data in the "Subtest" and "Pages" error scores. The WAIS-III is the most recent adult Wechsler intelligence test and there was no current research about the

number of scoring errors that students made on this test. A more detailed breakdown of scoring errors and missing data for this test was deemed useful. A copy of each of the checklists is included in Appendix C.

Research Assistants

Because of the large number of protocols to be scored (240), the researcher enlisted four graduate students to help with the scoring. All four of these research assistants were female, second year Master's level graduate students. They were each in the process of completing their coursework and their own research projects. They had all previously completed the Individual Assessment course and were knowledgeable about the administration and scoring of the Wechsler tests.

To ensure accurate and consistent scoring of the protocols four steps were followed. First, an orientation session was organized with all of the research assistants in attendance. During this session, Wechsler administration and scoring manuals were provided to the research assistants to review and use when scoring the protocols. Second, administration and scoring checklists for the WISC-III and the WAIS-III to be used to score the protocols were reviewed. Third, meetings were held with the research assistants two times a month in order to answer questions and review the scoring they had completed. Any discrepancies identified by the research assistants were usually resolved after referring to the administration and scoring manual. Any difficult-to-score responses were discussed in a group format with the researcher and the research assistants and a mutual decision was reached regarding the correct score that should be assigned to a particular item. Other researchers (Alfonso et al., 1998 and LoBello & Holley, 1999) used a similar discussion format to resolve discrepancies or uncertainties in scoring errors. Fourth, random reviews of each research assistant's checklists occurred throughout the study to ensure scoring accuracy and correct calculation of errors.

CHAPTER FOUR

Results

WISC-III Subtest Errors

All examiners made errors when administering and scoring the Wechsler tests. These errors were nonsystematic and occurred across all protocols. Based on their random nature, only the totals of the different types of errors are presented. The errors committed at the subtest level are reported first followed by the errors made at the aggregated total test score level. The results for the WISC-III are reported in Chapters four and five and the results for the WAIS-III are reported in the Chapters six and seven.

Overview of WISC-III Subtest Errors

The WISC-III is composed of thirteen subtests, seven of which are performance subtests and six are verbal subtests. The administration of the WISC-III begins with a performance subtest followed by a verbal subtest and continues in this alternating manner for the administration of all thirteen subtests. Not all examiners completed all of the thirteen subtests. Two supplementary subtests, the Symbol Search and Mazes subtests, were not uniformly administered by all examiners and were therefore not analyzed for errors.

The subtest errors are the sums of errors made on each of the subtests. The subtest errors contain four general error types: administration errors, computation errors, verbal analysis errors, and missing data errors. The administration errors generally include beginning the subtest at the incorrect starting point according to the examinee's age, failing to establish a correct basal level, discontinuing the subtest prematurely, and exceeding the discontinuance criteria. Not all subtests have the same administration errors. For example, since all items for the object assembly subtest are administered there can be no administration errors attributed to incorrectly beginning the subtest at the appropriate starting point or exceeding the discontinuance level. Further, there are different administration rules for different subtests resulting in errors specific to the subtest. For example, the Coding subtest is administered on a response sheet that the examiner removes from the record booklet. Failing to remove the response sheet from the record booklet is an administration error specific to this subtest.

The computation errors generally include incorrectly adding the subtest scores and failing to award bonus points for quick completion of an item. Some computation rules are specific to a subtest. For example, with the Object Assembly subtest, examinees assemble puzzle pieces to form a picture and points are awarded for the number of correct junctures assembled even though examinees may not have completed the picture. If an examinee completes only a portion of the puzzle, points are awarded by multiplying the number of correct completed junctures by a designated number. Errors may occur when examiners complete this calculation.

"Verbal Analysis Errors" are specific to one performance subtest (Picture Completion) and four of the verbal subtests (Information, Similarities, Vocabulary, and Comprehension). These errors include incorrect point assignment, failure to query, and incorrect querying. Verbal responses may be scored a zero, one, or two depending on the quality of the response. Verbal analysis errors can occur if the examiner assigns an incorrect point value for a response. Failing to query a response or querying a response incorrectly occurs when examiners either fail to query or inappropriately query an incomplete response.

Missing data errors consist of missing information such as when examiners do not record the examinees' responses verbatim, fail to record the time to complete tasks, and fail to circle point assignments and bonus points. Missing data was considered an error type, although an error type considered separately from the administration, computation, and verbal analysis errors because the omission of information does not necessarily have an adverse impact on the WISC-III scores.

Performance Subtests

The analysis of subtest error types for the performance subtests is presented first because the rules for administration and computation are generally more similar for these subtests. In order of administration, the performance subtests are Picture Completion, Coding, Picture Arrangement, Block Design, and Object Assembly. The Symbol Search and Mazes subtests are supplementary subtests, which the examiners did not always administer. Due to the lack of administrative consistency of these supplementary subtests, the number of errors made would be an inaccurate representation of the actual errors if all

examiners had administered this subtest in a routine manner. Therefore, there is no analysis of errors for these two subtests.

A more in-depth analysis of examiner subtest errors is presented for the Picture Completion subtest. This is done to demonstrate the level of individual analysis that occurred when examining the subtest errors made by examiners. Summary level results are presented for the remaining subtests. As pointed out before and as seen in the presentation of the subtests errors for the Picture Completion subtest, the errors made by the examiners in both the experimental and control groups were not systematic nor predictable within the four protocols completed by each examiner or across examiners.

Picture Completion Subtest

The Picture Completion subtest requires examinees to identify the missing element in a series of pictures. In terms of administration rules, examiners begin with the sample item for all ages and then start with a designated age-appropriate starting point. An error can occur if a basal level is not correctly established. The subtest ends when the examinee makes five consecutive response failures. Errors can occur if the examiner terminates the subtest prematurely or if the examiner continues the subtest beyond the five consecutive response failures. Failing to query or making an incorrect query are verbal analysis errors that can occur. Examiners are instructed to make three types of queries, only once each as needed, on this particular subtest. In terms of computation rules, the examiners must add the scores correctly. Table 7 contains a summary of the administration and computation errors for the Picture Completion subtest. The first set of numbers is the number of examiners who made errors listed in that column and the second set of numbers is the total number of errors made.

Table 7

Picture Completion Subtest Administration and Computation Errors

Group	Administration Errors	Verbal Analysis Errors	Computation Errors	Total Errors
Exp	10-12	7-12	2-2	26
Cntrl	4-4	6-14	4-4	22

Administration errors. In terms of the administration errors, 10 examiners in the experimental group made 12 errors. There were six errors of beginning the subtest incorrectly according to the child's age, one error of failing to establish a correct basal

level, and five errors of exceeding the discontinuance level. The first examiner started the subtest incorrectly with her first and second examinees. The second, third, and fourth examiners started the subtest incorrectly with their fourth examinees. The fifth examiner failed to begin the subtest correctly for her third examinee. The sixth examiner did not establish a correct basal level of testing for her third examinee. The seventh examiner exceeded the discontinuance level for her third and fourth examinees. The eighth examiner exceeded the discontinuance criteria for the third examinee. The ninth and tenth examiners exceeded the discontinuance criteria for their first and fourth examinees respectively.

For the control group, four examiners committed four administration errors. There were three errors of beginning the subtest incorrectly and one error of discontinuing the subtest incorrectly. Three examiners began the subtest incorrectly for their second examinees while the fourth examiner discontinued the subtest incorrectly for her third examinee.

Verbal analysis errors. Twelve verbal analysis errors, including nine failure to query errors, and three incorrect query errors, were committed by the seven examiners in the experimental group. The first examiner made two errors of failing to query a response - once for her first examinee and once for her fourth examinee. The second examiner failed to query a response for her fourth examinee. The third, fourth, and fifth examiners failed to query responses for their first examinees. The sixth examiner failed to query responses for her first and third examinees and incorrectly queried a response for her third examinee. The seventh examiner failed to query a response for her third examinee and the eighth examiner failed to query a response for his first examinee and queried incorrectly for his third and fourth examinees.

Fourteen verbal analysis errors, of which all were failure to query errors, were committed by six examiners in the control group. The first examiner made this error for her first, second, and third examinees and the second examiner made this error for her second, third, and fourth examinees. The third and fourth examiners failed to query a response for their first and fourth examinees. The fifth examiner made this error for her second and third examinees while the sixth examiner made this error for her second and fourth examinees.

Computation errors. In terms of the computation errors, two examiners from the experimental group made two errors in correctly adding scores for their last examinees. Four examiners from the control group made incorrect score calculations. Two examiners made calculation errors for their first examinees and the next two examiners made calculation errors for their second examinees.

Table 8
Picture Completion Subtest Missing Data Errors

Group	Not recording verbatim responses	Not recording incorrect response	Total Missing Data Errors
Exp	11-22	7-14	36
Cntrl	10-18	5-10	28

Missing data errors. The number of missing data errors committed by the examiners in each group while administering the picture completion subtest are reported in Table 8. Eleven examiners from the experimental group and ten examiners from the control group did not record examinees responses verbatim. Furthermore, 7 of the 11 examiners from the experimental group and five of the 10 from the control group tended to also omit recording the examinees' incorrect responses.

This in-depth analysis of the errors committed by the student examiners in the experimental group and in the control group reveals the haphazard incidence of these errors in both groups. There is no real systematic pattern of errors. The lack of systematic pattern makes it difficult to pinpoint exactly what could be done to reduce errors. Rather a more pervasive treatment would be required. As indicated earlier, the results of the analyses for the remaining subtests in the WISC-III and WAIS-III are presented at the group level.

Coding Subtest

The Coding subtest is a written, timed (two minutes) subtest. On the coding response sheet, a key of numbers with corresponding symbols is listed. The examinee is instructed to copy the symbols from the key into the appropriate space on the response sheet. The administration errors for the coding subtest include choosing the incorrect coding response sheet, failing to remove it from the record book prior to administering the test, failing to correctly record the time taken by the examinee to complete this subtest, and inaccurately allowing more time or failing to allow sufficient time to

complete the subtest. The computation errors include failing to correctly add the scores and award bonus points if the examinee accurately completed the task is less than two minutes. Table 9 contains a summary of the administration and computation errors for the Coding subtest.

Table 9
Coding Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	2-2	2-3	5
Ctr.	3-6	4-5	11

Administration errors. In terms of administration errors, two examiners from the experimental group made one error each. The first failed to remove the form from the response booklet and the other allowed too much time to complete the task. In contrast, three examiners from the control group made six administration errors. The majority of these errors, five, were failing to remove the form and one examiner did not allow enough time for the task to be completed.

Computation errors. Three computation errors were committed by two examiners in the experimental group and five were committed by four examiners in the control group. In each case, all six examiners failed to correctly add the scores.

Table 10
Coding Subtest Missing Data Errors

Group	Failing to record time	Failing to record scores	Total Missing Data Errors
Exp.	3-8	2-2	10
Ctr.	6-11	1-1	12

Missing data errors. A total of 10 missing data errors were committed by the examiners in the experimental group and 12 by the examiners in the control group. As shown in Table 10, most of these missing data errors involved failing to correctly record the time taken by an examinee to complete the coding subtest.

Picture Arrangement Subtest

The Picture Arrangement subtest requires a child to arrange a mixed up set of pictures into the correct order so that together they tell a logical story. First, a sample item is administered to all examinees. Then the examiner must begin the subtest at the correct starting point according to the examinee's age. The subtest is discontinued when the examinee makes three consecutive errors. Failure to begin, establish the correct basal,

and end this subtest correctly are administration errors. As with the previous subtest, the computation errors include incorrectly adding the item scores and failing to award bonus points if the examinee accurately completes the task within the time limit. Table 11 contains a summary of the administration and computation errors for the Picture Arrangement subtest.

Table 11
Picture Arrangement Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	7-10	3-6	16
Ctr.	9-15	3-4	19

Administration errors. Seven examiners in the experimental group committed ten administration errors. These 10 errors occurred at the beginning of the subtest: seven failures to begin the subtest with the correct item for the child's age and three failures to correctly establish the basal level. The majority of the 15 errors committed by the examiners in the control group were committed at the beginning: failing to begin the subtest with the correct item for the child's age occurred 12 times and failing to establish a correct basal level occurred once. The discontinuance criteria was exceeded twice.

Computation errors. Three examiners from each group made computation errors. While five of the six errors from the experimental group were due to addition errors and one error was due to failing to award the correct number of bonus points, the four errors committed by the three examiners in the control group were for failing to award the correct number of bonus points.

Table 12
Picture Arrangement Subtest Missing Data Errors

Group	Failing to record time	Failure to record correct /incorrect sequence	Failure to circle bonus points	Total Missing Data Errors
Exp.	1-1	5-9	2-3	13
Ctr.	0-0	4-7	1-1	8

Missing data errors. Thirteen missing data errors and eight missing data errors were committed, respectively, by six examiners in the experimental group and four examiners in the control group (see Table 12). Two examiners in the experimental group and one in the control group made two errors each. The majority of missing data errors,

nine for the experimental group and seven for the control group, occurred when examiners failed to record the response order for their examinees. Failure to circle bonus points occurred three times in the experimental group and once in the control group.

Block Design Subtest

The Block Design subtest requires the examinee to arrange cubes into a design to match the pattern depicted on a card. The administration rules state that the examiner must begin the subtest at the correct starting point according to the examinee's age. Administration errors can occur if the examiner begins with the incorrect starting point , establishes a basal level incorrectly, discontinues the subtest prematurely and exceeds the discontinuance criteria of two consecutive errors. As with the previous subtest, computation errors can occur when examiners incorrectly add the scores or fail to award bonus points if an examinee accurately completes the task within the time limit. Table 13 contains a summary of the administration and computation errors for the Block Design subtest.

Table 13

Block Design Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	7-9	2-2	11
Ctr.	8-10	3-3	13

Administration errors. Seven examiners from the experimental group made nine administration errors. The majority of these errors, six, were made when four examiners failed to begin the subtest with the correct item according to the child's age. One examiner failed to establish a correct basal level, one discontinued the subtest incorrectly, and the other exceeded the discontinuance criteria. While the total number of administration errors committed by the examiners in the control group was essentially the same as the total number for the experimental group (9 vs. 10), the variety of errors committed by the control group examinees was greater. Control group examiners incorrectly began the subtest four times. Another incorrectly stopped the subtest, while five examiners exceeded the discontinuance criteria five times.

Computation errors. Two addition errors were committed by two examiners in the experimental group. Three addition errors were committed by three examiners in the control group.

Table 14
Block Design Subtest Missing Data Errors

Group	Failing to record time	Failing to circle yes or no	Failure to circle correct point assignment	Failure to fill in incorrect designs	Total Missing Data Errors
Exp.	6-12	9-13	2-2	15-23	50
Ctr.	8-9	4-5	2-2	12-24	40

Missing data errors. The Block Design subtest contains many administration and scoring steps. Examiners from both groups omitted recording information that would indicate that they had followed the correct administration and scoring procedures. However, these errors of omission do not adversely impact the obtained scores.

As shown in table 14, the total number of missing data errors for the Block Design subtest was greater than the total number for the previous subtests. Approximately half, 23 out of 50 and 24 out of 40, of the errors occurred when examiners from both groups failed to record the incorrect designs made by their examinees. On the record form there are square grids beside each design for the purpose of recording the examinees' incorrect design patterns. The examiners are instructed to fill in the grids that correspond to the incorrect design made by their examinees. Many of the examiners did not fill in the grids but instead either indicated that the design was incorrect by circling the "no" and/or by assigning a zero point value. The benefit of recording the incorrect design is to have more in-depth information about potential examinee error patterns (e.g., reversed block patterns). The next most frequently committed errors were failing to record the time taken to complete the Block Design subtest and failing to circle yes or no to indicate if the item was completed correctly. Approximately half of the remaining errors, 25 errors for the experimental group and 14 errors for the control group, were due to these omissions. Two examiners from each group failed to circle the correct point assignment resulting in the final four missing data errors.

Object Assembly

The Object Assembly subtest requires examinees to assemble five different sets of puzzle pieces with each puzzle to be completed within a set time period. All five puzzles were to be administered to all examinees. If an examinee completes only a portion of the puzzle, points are still awarded by multiplying the number of correctly

completed junctures by a designated number. Errors can occur if the subtest is discontinued prematurely and if the time limit to complete a puzzle is exceeded. The computation errors included incorrect addition of the scores, failure to award bonus points for quick completion, incorrect multiplication of scores, and failure to round half scores up before adding item scores. Table 15 contains a summary of the administration and computation errors for the Object Assembly subtest.

Table 15
Object Assembly Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	1-1	6-9	10
Ctr.	2-2	6-10	12

Administration errors. One examiner in the experimental group and two examiners in the control group made the same administration error by exceeding the time limit to complete the puzzles.

Computation errors. In contrast to the previous subtests, the number of computation errors committed by the examiners in both groups was greater than the number of administration errors. Six examiners in each group committed computation errors. Of the nine errors committed by the experimental group, six errors involved incorrect multiplication, two errors involved failing to award the correct number of bonus points for quick puzzle completion, and one involved incorrectly adding scores. For the control group, five of the ten computation errors involved incorrect multiplication, four errors involved failing to award the correct number of bonus points, and one involved a score miscalculation.

Table 16
Object Assembly Subtest Missing Data Errors

Group	Failing to record time	Failure to record the # of correct junctures	Failure to circle correct point assignment	Total Missing Data Errors
Exp.	2-2	5-8	2-3	13
Ctr.	0-0	6-16	3-4	20

Missing data errors. As shown in Table 16, two examiners in the experimental group failed to record the time to complete the puzzles, five examiners failed to write down the number of correct junctures achieved by their examinees eight times, and two

examiners failed to circle the correct point assignment three times. For the control group, six examiners made 16 errors when they failed to write down the number of correct junctures achieved by their examinees and three examiners failed to circle the correct point assignment on four occasions.

Both groups committed the fewest number of administration errors on the Object Assembly subtest in comparison to the other performance subtests. The experimental group made fewer administration and computation errors than the control group, although the differences were small (i.e., one point difference in each instance). The teaching intervention included an instructional emphasis on administration and computation that included accurately recording the number of correct junctures and correctly multiplying the scores. While the experimental group consistently recorded the number of correct junctures, over half of the computation errors for each group were due to incorrectly multiplying the scores. The experimental group also made fewer missing data errors than the control group.

Summary of Performance Subtest Errors

The experimental group examiners committed fewer administration, verbal analysis, and computation errors than the control group examiners on four out of the five performance subtests. The Picture Completion subtest represents the only subtest where the control group examiners committed fewer of these errors. While the total number of subtest errors for both groups of examiners was generally similar: Picture Completion 26 vs. 22, Coding 5 vs. 11, Picture Arrangement 16 vs. 19, Block Design 11 vs. 13, Object Assembly 10 vs. 12 errors respectively between the experimental and control groups, the experimental group committed fewer total subtests errors on four of the five subtests. Conversely, the control group examiners committed fewer missing data errors on three of five performance subtests: Picture Completion 36 vs. 28, Picture Arrangement 13 vs. 8, and Block Design subtest 50 vs. 40 errors respectively. There were fewer missing data errors for the experimental group on the Coding and Object Assembly subtests (10 vs. 12 and 13 vs. 20 errors respectively). Overall the differences of the missing data errors between the two groups of examiners was variable (range = 2-10) relative to the differences of the administration, verbal analysis, and computation errors between the two groups of examiners (range = 2-6 errors).

Verbal Subtests

The verbal subtests are presented in the order in which they were administered: Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span.

Information Subtest

The Information subtest requires examinees to respond to general knowledge-based questions. Examiners begin with the correct starting point depending on an examinee's age and the subtest is discontinued once five consecutive errors are made. Administration errors can occur if the examiner begins at the incorrect starting point, fails to establish a basal level correctly, discontinues the subtest prematurely, and exceeds the discontinuance level. Verbal responses can warrant zero, one or two points depending on the quality of the response. Verbal analysis errors include assigning an incorrect point value for a response, failing to query, and querying a response incorrectly. The computation errors included incorrect addition of the item scores. Table 17 contains a summary of the Information subtest administration, verbal analysis, and computation errors.

Table 17

Information Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	7-10	7-9	1-1	20
Ctr.	6-10	7-11	3-3	24

Administration errors. Of the 16 examiners in the experimental group, seven examiners made 10 administration errors. The basal was incorrectly established four times, there were three errors of beginning the subtest incorrectly, and three errors of exceeding the discontinuance criteria. Of the 14 examiners in the control group, six made 10 administration errors. There were three errors of incorrectly establishing a basal level, three errors of discontinuing the subtest incorrectly, two errors of beginning the subtest incorrectly, and two errors of exceeding the discontinuance criteria.

Verbal analysis errors. Seven examiners from the experimental group made nine verbal analysis errors: seven incorrect point assignments and two errors of failing to query a response. Seven examiners from the control group made 11 verbal analysis errors: eight incorrect point assignments and three errors of failing to query a response.

Computation errors. One examiner from the experimental group made one addition error while three examiners from the control group made three addition errors.

Table 18
Information Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	6-9	9
Ctr.	5-9	9

Missing data errors. Six examiners in the experimental group and five examiners in the control group made the same number of missing data errors (see Table 18). In each case, these examiners failed to record verbatim the responses from examinees.

Similarities Subtest

An examinee is asked to explain how two words share common characteristics or similarities for the Similarities subtest. Examinees begin with the same item, regardless of age group, and the subtest is discontinued after four consecutive failures. Administration errors can occur if the subtest is terminated prematurely or extended past four consecutive failures. The verbal analysis errors consist of incorrect point assignments, failing to query responses, and incorrectly querying responses. Computation errors occur if the examiner fails to add the scores correctly.

Table 19
Similarities Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	7-10	15-77	2-2	89
Ctr.	7-8	14-54	0-0	62

Administration errors. As shown in Table 19, seven examiners in the experimental group made 10 administration errors. There were eight errors of exceeding the discontinuance criteria and two errors of discontinuing the subtest incorrectly. For the control group, seven examiners exceeded the discontinuance criteria on eight occasions.

Verbal analysis errors. Fifteen of the 16 examiners in the experimental group made 77 verbal analysis errors. Incorrectly assigning point values accounted for the majority of these errors (41 errors). The remaining errors were more evenly distributed across making incorrect queries (19 errors) and failing to query (17 errors).

All fourteen examiners in the control group made 54 verbal analysis errors. As with the experimental group, incorrectly assigning point values accounted for the majority of the errors (25 errors) while making incorrect queries (15 errors) and failing to query (14 errors) were again more evenly distributed.

Computation errors. Two examiners in the experimental group made two addition errors. There were no addition errors made by examiners in the control group.

Table 20
Similarities Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	2-2	2
Ctr.	4-4	4

Missing data errors. As shown in table 20, two examiners in the experimental group and four examiners in the control group failed to record the examinee's responses verbatim.

Arithmetic Subtest

The Arithmetic subtest requires an examinee to respond to a series of arithmetic questions. Examiners must begin the subtest at the correct starting point according to the examinee's age. The subtest is discontinued once three consecutive errors are made. Administration errors can therefore occur if the examiner begins with the incorrect starting point, incorrectly establishes the basal level, discontinues the subtest prematurely, or exceeds the discontinuance criteria. Computation errors were incorrect addition of scores and failure to award bonus points for quick response. There are no verbal analysis errors for this subtest as only one response is correct and no additional verbalizations are required. Table 21 contains the summary of errors for the Arithmetic subtest.

Table 21
Arithmetic Subtest Administration and Computation Errors

Group	Administration errors	Computation errors	Total Errors
Exp.	11-16	4-6	22
Ctr.	7-10	2-4	14

Administration errors. Eleven of the 16 examiners from the experimental group made 16 administration errors. There were five errors of beginning the subtest incorrectly, six errors of incorrectly establishing a basal level, and five errors of

exceeding the discontinuance criteria. Seven of the 14 examiners from the control group made 10 administration errors. There were five errors of exceeding the discontinuance level, two errors of beginning the subtest incorrectly, two errors of incorrectly establishing a basal level, and one error of discontinuing the subtest incorrectly.

Computation errors. Four examiners from the experimental group made four calculation errors and two errors of failing to award bonus points. In contrast, two examiners from the control group made four calculation errors.

Table 22
Arithmetic Subtest Missing Data Errors

Group	Failure to record time	Failure to record correct/incorrect responses	Total Missing Data Errors
Exp.	6-10	4-6	16
Ctr.	6-12	3-8	20

Missing data errors. Six examiners from each group failed to record the time examinees took to answer questions resulting in 10 errors for the experimental group and 12 errors for the control group (see Table 22). Four examiners from the experimental group failed to record the sequence of responses resulting in six errors while three examiners from the control group made eight of these omissions.

Vocabulary Subtest

Examinees must define words for the Vocabulary subtest. Examiners should begin with the correct starting point according to the examinee's age and establish the correct basal level. The subtest is discontinued once four consecutive errors are made. Possible administration errors include beginning at the incorrect starting point, failing to establish a correct basal level, discontinuing the subtest prematurely, or exceeding the discontinuance level. The verbal analysis errors consist of incorrect point assignments, failing to query responses, and incorrectly querying responses. Computation errors include failing to add scores correctly. Table 23 contains a summary of errors for the Vocabulary subtest.

Table 23
Vocabulary Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	10-19	16-138	2-2	159
Ctr.	10-19	14-111	0-0	130

Administration errors. Ten examiners from the experimental group made 19 administration errors. Two examiners began the subtest at the wrong starting point for the examinee's age, five failed to establish a correct basal level, nine discontinued the subtest incorrectly, and three exceeded the discontinuance criteria. Likewise, ten examiners from the control group made 19 administration errors. The majority of the errors, 12, occurred when examiners exceeded the discontinuance criteria. There were four errors of incorrectly establishing a basal level, two errors of incorrectly beginning the subtest and one error of exceeding the discontinuance criteria.

Verbal analysis errors. All 16 examiners in the experimental group made verbal analysis errors. Incorrectly assigning point values accounted for the majority of the errors (70 errors), followed by failing to query (55 errors) and making incorrect queries (13 errors). All 14 examiners from the control group made verbal analysis errors. Again, incorrectly assigning point values accounted for the majority of the errors (68 errors, followed by failing to query (33 errors) and making incorrect queries (10 errors).

Computation errors. In comparison to the administration and verbal analysis errors, they were comparatively few computation errors. Two examiners from the experimental group made two addition errors while there were no computation errors made by examiners from the control group.

Table 24
Vocabulary Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	0-0	0
Ctr.	1-2	2

Missing data errors. There were no missing data errors for the experimental group and only one examiner in the control group failed to record examinee responses verbatim on two occasions (see Table 24) .

Comprehension Subtest

Examinees are asked to provide an explanation or solution to everyday situations and display an understanding of life occurrences or events for the Comprehension subtest. All examinees begin the subtest at the same starting point and the subtest is discontinued once three consecutive errors are made. Thus, the possible administration errors are discontinuing the subtest prematurely and exceeding the discontinuance level. The verbal analysis errors consist of incorrect point assignments, failing to query responses, and incorrectly querying responses. As well, examiners may fail to request a second response for the eight questions that require responses that reflect two general concepts. If the response is related to one but not both concepts, examiners are instructed to ask the examinee for a second response that will complete the two concepts and warrant the full two points. Computation errors are failing to add scores correctly.

Table 25

Comprehension Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	8-11	15-134	2-2	145
Ctr.	4-4	14-141	1-1	146

Administration errors. For the experimental group, 8 examiners made 11 administration errors (see Table 25). All but one of these errors occurred when examiners exceeded the discontinuance criteria. One examiner discontinued the subtest incorrectly. In contrast, four examiners from the control group exceeded the discontinuance criteria.

Verbal analysis errors. Of the 16 examiners from the experimental group, 15 made 134 verbal analysis errors. Failing to obtain a second response accounted for the majority of these errors (60 errors). Incorrectly assigning point values and failing to query occurred evenly (42 occasions). Making incorrect queries (11 errors) accounted for the remaining verbal analysis errors. All 14 examiners from the control group committed verbal analysis errors which resulted in a large number of errors (141). Incorrectly assigning point values was the most frequent error (72 errors). However, failing to obtain a second response (43 errors) was the second most frequent error and failing to query and

making incorrect queries were committed approximately the same number of times (15 and 11 respectively).

Computation errors. Two examiners from the experimental group made two addition errors while one examiner in the control group made one addition error. Similar to previous verbal subtests, computation errors accounted for the fewest number of errors.

Table 26
Comprehension Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	2-2	2
Ctr.	1-1	1

Missing data errors. Two examiners in the experimental group and one examiner in the control group failed to record verbatim the responses made by the examinees.

Digit Span Subtest

The Digit Span subtest is a supplemental subtest. However, unlike the Symbol Search and Mazes supplemental subtests, all examiners administered this subtest across all protocols. Therefore, the Digit Span subtest is included in the error analysis.

For the Digit Span subtest, examinees are required to repeat a series of numbers in order as relayed by the examiner. There are two series of numbers or two trials that form an item. All ages begin with the first item and the administration rules state that the subtest is discontinued once the examinee makes an error on both trials of an item. Consequently, administration errors can occur if the examiner discontinues the subtest incorrectly or exceeds the discontinuance criteria. There are no verbal analysis errors for this subtest as only one response is correct and no additional verbalizations are required. Computation errors include failing to add the scores correctly or assigning an incorrect point value for either the trial of the item or the item itself. Table 27 contains a summary of the administration and computation errors for the Digit Span subtest.

Table 27
Digit Span Subtest Administration and Computation Errors

Group	Administration errors	Computation errors	Total Errors
Exp.	4-5	3-3	8
Ctr.	2-4	0-0	4

Administration errors. In comparison to the previous four verbal subtests, only a few errors were made while administering the Digit Span subtest. For the experimental

group, four examiners exceeded the discontinuance criteria for five of their examinees. For the control group, two examiners made four administration errors and, similar to the experimental group, all exceeded the discontinuance criteria.

Computation errors. In terms of computation errors, three examiners from the experimental group made two calculation errors and one incorrect point assignment. There were no computation errors made by the examiners from the control group.

Table 28
Digit Span Subtest Missing Data Errors

Group	Failing to record correct response	Failure to record incorrect response	Failure to record trial/item scores	Total Missing Data Errors
Exp.	15-37	9-23	5-10	70
Ctr.	13-42	7-17	6-14	73

Missing data errors. In contrast to the administration and computational errors, the number of missing data errors committed by the examiners when administering the Digit Span subtest exceeded the number of missing data errors for the other four verbal subtests. As shown in Table 28, 70 missing data errors were committed by the examiners in the experimental group and 73 by the examiners in the control group. More than half of the errors in each group involved failing to record the correct response: 15 experimental examiners on 37 occasions and 13 control examiners on 42 occasions. Further, 9 of the 15 experimental examiners and 7 of the 13 control examiners failed to record the incorrect response 23 times and 17 times respectively. Lastly, five examiners in the experimental group failed to record trial/item scores 10 times; 6 examiners in the control group did likewise 14 times.

Summary of Verbal Subtest Errors

The control group examiners committed fewer administration errors than the experimental group examiners on four of the six verbal subtests (i.e. Similarities, Arithmetic, Comprehension, and Digit Span subtests). Both groups committed the same number of administration errors on the Information and Vocabulary subtests (10 and 19 errors respectively). Overall, there were bigger discrepancies for the verbal analysis errors between the two groups (Information 9 vs.11, Similarities 77 vs.54, Vocabulary 138 vs.111, and Comprehension 134 vs.141 errors respectively between the experimental and control groups) compared to the other types of errors. The control group examiners

committed fewer computation errors on all but one subtest (the Information subtest). Regarding the total number of subtests errors, there were fewer of these errors committed by the control group examiners with a large discrepancy in these errors on the Similarities and Vocabulary subtests (89 vs. 62 and 159 vs. 130 errors respectively) and a smaller discrepancy in these errors on the Arithmetic and Digit Span subtests (16 vs. 20 and 70 vs. 73 errors respectively). The Information and Comprehension subtests were two subtests where the experimental group committed fewer total subtest errors although, the differences were slight (four errors and one error difference respectively between the two groups on these two subtests). There were fewer missing data errors for both groups in comparison to the other error types and the experimental group committed the same or fewer missing data errors on five of the six verbal subtests (range of missing data errors = 0-4); however, there was much less variability between the two groups with respect to the missing data errors.

CHAPTER FIVE

Results

WISC-III Total Errors

The WISC-III total errors are the sum of errors made on the cover and back of the cover page of the record form. Located on the cover page of the WISC-III is space for recording the identifying information about a child and recording the child's scores. The identifying information includes the child's name, chronological age calculation, gender, school, grade, handedness, and the examiner's name. The scores include subtest raw scores, the scaled scores, and index score conversions. The reverse side of the cover page contains behavioural observation headings such as attitude toward testing, attention, affect/mood, and unusual behaviour/verbalizations. As part of best practice in administering the WISC-III, an examiner is to complete both sides of the cover page.

Overview of WISC-III Total Errors

Examiners made errors on the record form that fit into five total error types:

1. Age miscalculation
2. Used U. S. norms rather than Canadian norms
3. Conversion errors (i.e,: incorrect calculation and conversion of scores)
4. Plotting errors (i.e,: errors that are committed when transferring and marking scores onto the graphs)
5. Missing data errors (i.e,: not recording the child's demographic information, not recording behavioural observations on the second page, and failing to fill in the graphs)

Table 29 contains a summary of the frequency of these five error types committed by participants in each group.

Table 29
WISC-III Total errors

	Error Type					
Group	Age Miscalculation	U. S. Norms Used	Conversion Errors	Plotting Errors	Missing Data Errors	Total Errors
Exp.	2-2	1-2	10-32	14-39	13-45	120
Ctr.	1-1	5-8	8-13	8-33	10-49	104

Age Miscalculation

Examiners must correctly calculate the age of their examinee in order to select the correct age norms. This calculation involves computing the chronological age in the form of years, months, and days from the date of birth. The date of birth is provided by the examinee, or if the examinee if not able, a parent/guardian. Once the chronological age is calculated in terms of years, months, and days, the examiner selects the norms to be used to convert the raw scores achieved by the examinee to age equivalent scaled scores that are subsequently interpreted.

Two examiners in the experimental group and one examiner in the control group made errors in computing the ages (see Table 29). One of the examiners in the experimental group computed the correct number of years and months but did not compute the number of days. The second examiner in the experimental group incorrectly computed the number of months. However, in both cases the correct norms table for each of the examinees was selected; although an error was made, the nature of the error was such that the norms table selected corresponded to both the age as computed and the correct age of the examinees. Like the second of these two examiners, the one examiner in the control group also incorrectly computed the month. However, in this case the nature of the error was such that although he selected the norms table that corresponded to the age he computed, it was the incorrect table given the correct age of the examinee.

Incorrect Use of U. S. Norms

The children's version of the Wechsler Intelligence Scale was re-normed in 1991 (WISC-III) and separate Canadian norms were prepared for use in Canada in 1996. The Canadian norms are presented in a separate booklet called the *WISC-III Manual-Canadian Supplement.* In the preface of this supplement, the examiner is instructed to use the Canadian normative data for scoring and interpretation purposes when using the WISC-III with Canadian children. Circumstances may arise where an examiner may opt to use the U.S. norms for a particular examinee, such as if the U. S. norms would provide a better comparison for the examinee. However, generally examiners in Canada must use the WISC-III manual and remember to replace the American norms included in that manual with the Canadian norms from the supplement. The need for Canadian norms and information about the Canadian study that produced the Canadian WISC-III norms are

included in this supplement. Additional tables about significant differences and frequencies between index scores and supplemental scores and a guide for unique Canadian responses are also provided in this supplement.

To remind Canadian examiners to use the correct norms, the WISC-III record form is identified as a "Canadian Record Form" but only recent record forms, since 2000, have a table located on the cover page labeled "Norm Set Used." This table is available for the examiner to indicate with a check mark whether Canadian or U. S. Norms were used.

One examiner in the experimental group incorrectly used the U. S. norms for the first two examinees she assessed. In contrast, five examiners in the control group incorrectly used the U.S. norms resulting in eight errors. Two examiners used the U.S. norms for their first examinees. Another examiner used the U.S. norms for the third examinee she tested while the fourth examinee used the U.S. norms for her last examinee. The fifth examiner used the U.S. norms for all four examinees. No reason was provided in each case in which the U. S. norms were used. Consequently, the use of the U. S. norms was considered an error.

Conversion Errors

After the raw scores are obtained for the individual subtests, they are converted to derived scores. Derived scores such as scaled scores, index scores, and percentiles allow an examinee's converted score to be compared to the normative group in order to determine relative standing. The raw scores from the 13 individual subtests are converted to subtest scaled scores that have a mean of ten and a standard deviation of three. Examiners had to consult the norms tables to convert the subtest raw scores to the corresponding age-appropriate subtest scaled score equivalents. Selected subtest scaled scores are then added to form, respectively, four index scores and three IQ scores that have a mean of 100 and a standard deviation of 15. The four index scores are the Verbal Comprehension (VCI), Perceptual Organization (POI), Freedom from Distractibility (FD), and Processing Speed (PSI). The three IQ scores are the Verbal (VIQ), Performance (PIQ) and Full Scale IQ scores (FSIQ). Examiners had to consult the IQ and Index Scores tables to convert the sums of scaled scores to the appropriate IQ and index

score equivalents and percentiles. As well as reporting each IQ and index score, the examiner also chose whether to use the 90% or the 95% confidence interval.

A total of 32 conversion errors were committed by 10 examiners in the experimental group (see Table 29). The range of errors was from one to four with one examiner committing 11 of the conversion errors. Thirteen conversion errors were committed by eight examiners in the control group. The range of errors was from one to three. The distribution of errors across the points in the conversion activity is shown in Table 30.

Table 30
WISC-III Conversion Errors

	Error Type					
Group	Incorrect conversion of subtest raw scores	Incorrect selection of subtests for Index and IQ scores	Incorrect calculation sum of scaled scores	Incorrect table used for Index/IQ scores	Incorrect % and confidence intervals	Total Conversion Errors
Exp.	1-1	1-1	4-10	6-10	6-10	32
Ctr.	3-3	0-0	0-0	5-6	3-4	13

Incorrect Conversion of Subtest Raw Scores to Subtest Scaled Scores

One examiner from the experimental group made one incorrect conversion of a subtest raw score to a subtest scaled score that produced additional significant errors. The subtest scaled score was incorrect and written on the cover page as 22, which was 14 points higher than the correct subtest scaled score of 8. The impact of this error was an elevation of 21 points for the VCI, 18 points for the VIQ, and 10 points for the FSIQ.

Three examiners from the control group made errors when converting subtest raw scores to subtest scaled scores. The first two examiners made one error each and these errors occurred for their third and fourth examinees, respectively. The impact for both errors was a one-point elevation to the VCI and VIQ index scores. It was unclear which table the third examiner had consulted when converting her subtest raw scores to subtest scaled scores for her third examinee. The impact of her errors resulted in a six-point elevation on all index scores except for the PSI.

Incorrect Selection of Subtests used for Index and IQ scores

One examiner from the experimental group incorrectly included the score of a spoiled subtest when calculating the sums of scaled scores for the VCI for her fourth examinee. A subtest becomes "spoiled" when the examiner incorrectly administers the subtest. Examiners are instructed to not include the spoiled subtest in the computation of subtest scaled scores to index scores because the results would be misleading. In this particular instance, the examiner failed to discontinue the Comprehension subtest once her examinee had reached the discontinuance level, which resulted in a spoiled subtest. She incorrectly included this subtest in the calculation of the VCI. No selection errors were committed by the examiners in the control group.

Incorrect Calculation of Sum of Scaled Scores

Four examiners from the experimental group committed ten errors when determining the index/IQ scores. The first examiner made five such errors: an incorrect addition of subtest scaled scores when calculating the PSI for her first examinee, resulting in a one point elevation of this index; omission of ten points from the calculation of the scaled scores for both the POI and the PIQ resulting in a deficit of 17 IQ points for the POIQ for her second examinee, 15 points for PIQ score and 8 points for the FSIQ; and incorrect addition of the subtest scaled scores for the VIQ and FSIQ for her fourth examinee which resulted in a somewhat smaller two point IQ deficit difference for both the VIQ and the FSIQ. The second examiner made three addition errors when calculating the index scores for her fourth examinee. These errors resulted in one point deficit on the VIQ and FSIQ and a two-point deficit on the VCI. The third and fourth examiners made one error each when determining the FSIQ and POI resulting in a one point IQ deficit of each index score. No calculation errors were committed by the examiners in the control group.

Incorrect Table used for Index/IQ scores

Six examiners in the experimental group committed ten errors when consulting the table to determine Index/IQ scores. Eight of the errors occurred when all five examiners mistakenly used the PSI table rather than the FDI table when determining the Index/IQ scores. One examiner made this error for her first, second and fourth examinees and the second examiner made this error for her second and fourth examinees. One IQ

point deficit occurred for each examinee. The third examiner made this error for her second examinee resulting in a three-point deficit on the PSI. The fourth examiner made this error for her third examinee resulting in a two-point elevation on the PSI and the fifth examiner made this error for her fourth examinee resulting in a one point PSI deficit. One examiner also mistakenly used the FDI table rather than the PSI table for her first and second examinees resulting in a one-point deficit for each FDI score. One examiner also looked at the wrong line in the FDI table and made one error by assigning an FDIQ that was three points lower than it should have been.

Five examiners in the control group made six errors by using the wrong table or looking at the wrong line from a table. Three errors occurred when three examiners mistakenly used the FDI table rather than the PSI table. These errors occurred for their first, second and third examinees respectively and resulted in a one-point elevation for each PSI. One examiner made two errors by using the U. S. norms for her first examinee when converting the VCI and the POI but used the Canadian norms for the remainder of the Index conversions. Both the VCI and the POI were incorrectly elevated by one point resulting in two errors. If U. S. norms were to be used, they should have been used consistently for all Index conversions. One examiner made one error when recording the FSIQ by looking at the wrong line from the IQ table leading to a two point incorrect elevation on the FSIQ.

Incorrect % and Confidence Intervals

Six examiners from the experimental group made ten errors in correctly identifying percentiles and confidence intervals. Three examiners from the control group committed the same error for four examinees. In all cases, the examiners wrote down the wrong confidence interval range from the tables.

Plotting Errors

Plotting errors occurred when student examiners transferred and marked scores onto the graphs. A total of 39 plotting errors were committed by 14 examiners from the experimental group while 33 plotting errors were made by 8 examiners in the control group (see Table 31). The range of errors for the student examiners across the plotting activities was 1 to 11 for the experimental group and 1 to 9 for the control group. Within the experimental group, one examiner committed 11 plotting errors while another

committed 8 errors. One examiner in the control group committed 9 plotting errors. The remaining examiners who committed plotting errors in both groups made between 1 and 5 plotting errors. The distribution of errors across the points in the plotting activity is show in Table 31.

Table 31
WISC-III Plotting Errors

	Error Type					
Group	Incorrectly transferring raw subtest scores to profile page	Incorrectly transferring subtest scaled scores onto the graph	Incorrectly transferring Index/IQ scores	Incorrectly plotting subtest scaled scores onto the graph	Incorrectly plotting Index/IQ scores onto the graph	Total Plotting Errors
Exp.	1-1	9-11	5-10	3-3	8-14	39
Ctr.	2-2	4-7	4-9	2-2	6-13	33

Incorrectly Transferring Raw Subtest Scores to Profile Page

One examiner from the experimental group incorrectly transferred the raw score of one subtest to the table on the cover page. This error did not result in a difference in the subtest scaled score. Two examiners in the control group committed the same error. Neither error resulted in a change to the subtest scaled scores.

Incorrectly Transferring Subtest Scaled Scores onto the Graph

Nine examiners from the experimental group made 11 errors when transferring the subtest scaled scores onto the graph. Two examiners made this error twice. For the control group, four examiners made seven errors with two examiners making this error twice.

Incorrectly Transferring Index/IQ scores

Five examiners from the experimental group made ten errors when transferring the Index and IQ scores onto the graph. The majority of these errors were made by one examiner who committed this error four times-twice when transferring the IQ scores and twice when transferring the Index scores on both the first and fourth protocols. Similar to the experimental group, one examiner in the control group committed the majority of these errors, six of the nine errors, by incorrectly transferring the Index scores and the IQ scores one time on each of the second, third, and fourth protocols.

Incorrectly Plotting Subtest Scores onto the Graph

Three examiners from the experimental group made one error each when plotting the subtest scores onto the graphs. Two examiners from the control group made one error each when plotting their subtest scores onto the graphs.

Incorrectly Plotting Index/IQ Scores onto the Graph

Eight examiners from the experimental group made 14 errors when plotting the Index and IQ scores onto the graph. Three examiners made at least two errors each (four, three, and two errors respectively) with these errors occurring on all but the third protocols. Six examiners from the control group made thirteen errors when plotting the Index and IQ scores onto the graph. Four examiners made at least two errors each (four, three, and two errors respectively) with the errors occurring across all protocols.

Missing Data Errors

Missing data include failing to record the examinee's demographic information, failing to record behavioural observations, and failing to fill in the graphs. A summary of the number of missing data errors committed by the examiners in both groups is presented in Table 32.

Table 32
WISC-III Total Missing Data errors

	Error Type			
Group	Failing to record demographic information	Failing to record behavioural observations	Failing to fill in graphs	Total errors
Exp.	3-3	10-41	1-1	45
Ctr.	5-8	10-35	3-6	49

Failing to Record Demographic Information

Space is provided on the top left corner of the cover page of the WISC-III record form to write the examinee's pertinent information. Examiners are to enter the demographic information including the examinees' name, gender, school, grade, hand dominance, and the name of the examiner administering the test on the space provided.

Three examiners in the experimental group committed three errors and five examiners in the control group made eight errors by failing to completely record this demographic information. For the experimental examiners, two did not record the examinee's school. The third experimental group examiner did not identify the gender of

her fourth examinee. Of the control group examiners, three did not record the hand dominance for their second examinees with one of these examiners also omitting to record the hand dominance for her third examinee. The fourth examiner omitted recording the hand dominance for her third examinee. The fifth examiner failed to list the school for her second, third and fourth examinees.

Failing to Record Behavioural Observations

Examiners are instructed to record behavioural observations of their examinees as part of a comprehensive and thorough assessment on the reverse cover page of the WISC-III record form. The topics identified on this page include the reason for the child's referral for testing, attitude toward testing, physical appearance, attention level, visual/auditory/motor problems, language, affect/mood, unusual behaviours/verbalizations, and other remarks.

Only three of the 16 examiners from the experimental group completely recorded their behavioural observations on the reverse cover page. However, ten of the 13 who did not recorded behavioural observations on the reverse cover page nevertheless jotted down comments and observations throughout the protocol as they proceeded with the testing. Only four of the 14 examiners from the control group completely recorded their behavioural observations on the reverse cover page; however, in a similar manner to the experimental group, five of the ten examiners who did not record behavioural observations in the correct place did record comments throughout their protocols as they assessed their examinees.

Failing to Fill in the Graphs

As part of recording information for an examinee, the examiners plot the examinee's scores onto the two graphs to provide a visual display of this information. One examiner from the experimental group failed to fill in the graphs on her last protocol, resulting in one error of omission. Three examiners from the control group failed to fill in the graphs resulting in six errors of omission. The first examiner did not fill in the graphs on her fourth protocol, which produced one error. The second examiner made one error by failing to fill in the graphs on her first protocol while the third examiner omitted filling in the graphs on all four protocols resulting in four errors.

Summary of WISC-III Total Errors

There were relatively few age miscalculations made by either group compared to all other error types. In regard to incorrectly using US norms, more examiners in the control group made this type of error compared to the examiners in the experimental group although, these errors also accounted for few of the total errors committed by either group of examiners. The conversion, plotting, and missing data errors accounted for most of the total errors committed by both groups of examiners and the experimental group committed more of these errors compared to the examiners in the control group.

In terms of the conversion errors, the majority of these errors committed by both groups were due to examiners consulting the wrong tables to obtain their scores and recording the wrong information from these tables. Fewer of the errors were due to actual calculation errors. The experimental group committed more of these errors (32 vs. 13). With regard to the plotting errors, the main source of plotting error for both groups was due to incorrectly transferring scores from the tables to the graphs and failing to follow through in plotting the corrections made by supervisors in the table areas onto the graphs. Once again, the experimental group committed more of these errors (39 vs. 33).

Students from both groups committed more missing data errors than any of the other error types and the total number of missing data errors was similar for both groups (45 vs. 49). Failing to record behavioural observations on the cover page of the protocol accounted for the majority of the missing data errors. Many of the examiners from both groups did recognize the importance of recording comments and including such observations somewhere on the WISC-III protocol.

All examiners learned about administration and scoring of the WISC-III as part of the introductory lecture. Both groups learned at the same time about general administration information such as filling in the record form, making the score conversions and graphing the results onto the protocol. There was, nonetheless, an expectation that the experimental group would be more accurate in administering and scoring the WISC-III after taking part in the experimental teaching intervention. One part of the teaching intervention involved examiners viewing a live demonstration of a WISC-III administration, they then scored their protocols from this live demonstration and discussed the scoring results. This was a teaching intervention not used with the control

group. Yet, the total number of errors was greater for the experimental group in comparison to the control group.

CHAPTER SIX

Results

WAIS-III Subtest Errors

The WAIS-III is composed of 14 subtests, seven of which are performance subtests and seven of which are verbal subtests. The administration of the WAIS-III begins with a performance subtest followed by a verbal subtest and continues in a generally alternating manner for the administration of all 14 subtests.

Similar to the WISC-III, there are four general error types: administration errors, computation errors, verbal analysis errors, and missing data errors. Also similar to the WISC-III, not all subtests have the same administration, computation or verbal analysis rules but rather some rules are specific to certain subtests. Unlike the WISC-III where examinees of different ages begin at different starting points on several of the subtests, all examinees for the WAIS-III begin at the same designated starting point.

Performance Subtests

The analysis of subtest error types for the performance subtests are presented first because the administration and computation rules are generally more similar for these subtests. In order of administration, the performance subtests are Picture Completion, Digit Symbol-Coding, Block Design, Matrix Reasoning, Picture Arrangement, Symbol Search, and Object Assembly. The Symbol Search, and Object Assembly subtests are supplementary subtests, which the examiners did not always administer. Also, the Digit Symbol coding subtest contains two optional procedures (i.e., Incidental Learning and Digit Symbol-Copy) that were not routinely administered. Due on the lack of administrative consistency of these supplementary subtests and optional procedures, the number of errors made would be an inaccurate representation of the actual errors if all examiners had administered them in a routine manner. Therefore, there is no analysis of errors for these two subtests.

Picture Completion Subtest

The Picture Completion subtest requires examinees to identify the missing element in a series of pictures. In terms of administration rules, examiners begin at the designated starting item and the subtest ends when the examinee makes five consecutive response failures. Consequently, the administration errors in the Picture Completion

subtest include incorrectly establishing the basal level, premature termination of the subtest and continuance of the subtest beyond the five consecutive response failures. Examiners are instructed to make three types of queries, only once each as needed, on this particular subtest. Failing to query or making an incorrect query are verbal analysis errors that can occur. The computation errors included incorrect addition of the item scores. Table 33 contains a summary of the administration, verbal analysis, and computation errors for the Picture Completion subtest.

Table 33
Picture Completion Subtest Administration and Computation Errors

Group	Administration Errors	Verbal Analysis Errors	Computation Errors	Total Errors
Exp	6-16	5-13	1-1	30
Ctr.	7-9	9-20	7-7	36

Administration errors. Six examiners in the experimental group made 16 administration errors. There were 13 errors of failing to establish a correct basal level, 2 errors of exceeding the discontinuance level, and one error of discontinuing the subtest incorrectly. There was less variability of errors for the control group with seven examiners making nine administration errors of failing to establish a correct basal level.

Verbal analysis errors. Five experimental group examiners made 13 verbal analysis errors, all of which were failure to query. Similarly, nine examiners from the control group made 20 verbal analysis errors, of which all were also failure to query.

Computation errors. Only one examiner in the experimental group incorrectly added the item scores. In contrast, seven examiners in the control group each committed this error once.

Table 34
Picture Completion Subtest Missing Data Errors

Group	Not recording verbatim responses	Not recording incorrect response	Total Missing Data Errors
Exp	7-26	6-12	38
Ctr.	6-13	1-1	14

Missing data errors. As shown in the Table 34, while approximately the same number of examinees in both groups failed to record verbatim the responses made by the examinee, the examinees in the experimental group did so twice as often as the

examinees in the control group, 26 versus 13. Further, of the seven experimental examiners that committed this error, 6 did not record the incorrect response on 12 occasions while only one control examiner committed this error.

Digit Symbol-Coding Subtest

The Digit Symbol-Coding subtest has additional optional procedures called Incidental Learning and Digit Symbol-Copy that the examiner can choose to administer. Not all examiners administered the optional procedures and there were therefore not analyzed for errors. All examiners administered the Digit Symbol-Coding subtest. It is a written, timed (two minutes) subtest. On the response sheet, a key of numbers with corresponding symbols are listed and the examinee is instructed to copy the symbols from the key into the appropriate space on the response sheet. The examiner must correctly record the time taken by the examinee to complete this subtest. Timing errors can occur, such as allowing more time or failing to allow sufficient time to complete the subtest. Computation errors include incorrectly counting the correct number of item scores and including the sample items in the total number of correct items. A Digit Symbol Scoring Template is available to help examiners score this subtest. The template is placed over the response sheet to check for errors in the examinee's copied symbols. Table 35 contains a summary of the administration and computation errors for the Digit Symbol-Coding subtest.

Table 35

Digit Symbol-Coding Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	0-0	11-18	18
Ctr.	0-0	9-17	17

Administration errors. There were no administration errors committed by either group of examiners. As such, all examiners correctly administered the sample items and correctly timed their examinees for the Digit Symbol-Coding subtest.

Computation errors. Despite the use of the scoring template, examiners made errors in correctly adding scores and incorrectly including the sample items in the total number of completed items. Eleven experimental group examiners made 18 computation errors while nine control group examiners made 17 computation errors.

Table 36
Digit Symbol-Coding Subtest Missing Data Errors

Group	Failing to record time	Failing to record score	Total Missing Data Errors
Exp.	3-3	5-6	9
Ctr.	1-4	2-2	6

Missing data errors. There were few missing data errors committed by either of the groups with the majority of examiners remembering to record the time to complete the subtest and the recording the examinee's score (see Table 36).

Block Design Subtest

The Block Design subtest requires the examinee to arrange cubes into a design to match the pattern depicted on a card. The administration rules state that the examiner must begin the subtest at the designated starting point. Administration errors can occur if the examiner begins at the incorrect starting point, if the basal level is established incorrectly, the subtest is discontinued prematurely, or the discontinuance criteria is exceeded. The computation errors include incorrect calculation of the items and failure to award bonus points if the examinee accurately completed the task within the time limit. Table 37 contains a summary of the administration and computation errors for the Block Design subtest.

Table 37
Block Design Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	9-15	2-4	19
Ctr.	4-6	2-3	9

Administration errors. Nine examiners from the experimental group committed 15 administration errors. In contrast, four examiners from the control group made six administration errors. In all cases, the examiners failed to achieve a correct basal level.

Computation errors. An equal number of examiners from each group, two, made respectively, four and three computation errors.

Table 38
Block Design Subtest Missing Data Errors

Group	Failing to record time	Failing to circle yes or no	Failure to circle correct point assignment	Failure to fill in incorrect designs	Total Missing Data Errors
Exp.	7-10	11-14	4-4	8-10	38
Ctr.	6-9	7-11	4-6	9-15	41

Missing data errors. There were numerous omissions of information that would indicate that examiners from both groups had followed the correct administration and scoring procedures (See Table 38). These errors of omission did not necessary have an adverse impact on the obtained scores but indicated that examiners had not filled in all required information on the record form for this subtest.

Matrix Reasoning Subtest

The Matrix Reasoning subtest is a nonverbal reasoning task where examinees look at a patterned picture or matrix with a missing section and choose the missing section from a selection of five choices. Three sample items are administered to all examinees prior to the beginning of the subtest. Administration errors can occur by failing to administer the sample items, incorrectly establishing a basal level of testing, discontinuing the subtest prematurely, and exceeding the discontinuance criteria. Computation errors include incorrectly adding the item scores. Table 39 contains a summary of the administration and computation errors for the Matrix Reasoning subtest.

Table 39
Matrix Reasoning Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	14-43	3-3	46
Ctr.	12-42	3-3	45

Administration errors. Fourteen of the 16 examiners from the experimental group made 43 administration errors. These errors included 16 errors of failing to administer the sample items, 16 errors of failing to establish a correct basal level, 9 errors of exceeding the discontinuance criteria, and 2 errors of incorrectly discontinuing the subtest. Twelve of the 14 examiners from the control group made 42 administration errors. These errors included 18 errors of failing to administer the sample items, 12 errors

of exceeding the discontinuance criteria, 11 errors of failing to establish a correct basal level, and one error of incorrectly discontinuing the subtest.

Computation errors. The same number of examiners from each group, three, each made one computation error.

Table 40
Matrix Reasoning Subtest Missing Data Errors

Group	Failing to circle response option	Total Missing Data Errors
Exp.	1-2	2
Ctr.	0-0	0

Missing data errors. Only one examiner from the experimental group failed to circle the response option made by an examinee on the record form. This examiner made this error of omission on two occasions.

Picture Arrangement Subtest

Examinees need to arrange a mixed up set of pictures into the correct order so that they tell a logical story in the Picture Arrangement subtest. The administration rules state that the sample item is administered to all examinees followed by the administration of remaining items. Administration errors can occur if the basal level is established incorrectly, if the subtest is discontinued prematurely, and if the discontinuance criteria is exceeded by continuing to test after four consecutive errors are made. The computation errors included the incorrect addition of item scores. Table 41 contains a summary of the administration and computation errors for the Picture Arrangement subtest.

Table 41
Picture Arrangement Subtest Administration and Computation Errors

Group	Administration Errors	Computation Errors	Total Errors
Exp.	1-1	5-5	6
Ctr.	0-0	4-4	4

Administration errors. Only one administration error was committed across both groups. One experimental examiner exceeded the time limit for one item.

Computation errors. A similar number of computation errors, five and four, were made, respectively, by five examiners in the experimental group and four examiners in the control group.

Table 42
Picture Arrangement Subtest Missing Data Errors

Group	Failing to record time	Failure to record correct /incorrect sequence	Failure to circle scores	Total Missing Data Errors
Exp.	4-7	3-7	1-4	18
Ctr.	5-5	3-5	2-2	12

Missing data errors. There was a variety of missing data errors for this subtest (See Table 42). For the experimental group, four examiners failed to record the time, three examiners failed to record the correct and incorrect sequence of responses, and one examiner failed to circle the scores. For the control group, five examiners failed to record the time, three examiners failed to circle record the response order, and two examiners failed to circle the scores.

Summary of Performance Subtest Errors

The control group examiners committed fewer administration and computation errors than the experimental group on four out of the five performance subtests. The Picture Completion subtest represents the only subtest where the experimental group committed fewer of these errors. While the total number of subtest errors for both groups of examiners on the subtests were generally similar: Picture Completion 30 vs. 36, Digit Symbol-Coding 18 vs. 17, Block Design 19 vs. 9, Matrix Reasoning 46 vs. 45, and Picture Arrangement 6 vs. 4 (comparing the experimental versus control groups), the control group committed fewer total errors on four of the five subtests. The control group also committed fewer missing data errors than the experimental group for four out of five performance subtests: Picture Completion 38 vs. 14, Digit Symbol-Coding 9 vs. 6, Matrix Reasoning 2 vs. 0, and Picture Arrangement 18 vs. 12. Only the Block Design subtest contained fewer missing data errors for the experimental group (38 vs. 41). Overall, the differences between the missing data errors between the two groups of examiners was larger (range = 2-24) relative to the differences between the administration, verbal analysis, and computation errors (range = 1-10).

Verbal Subtests

The verbal subtests are presented in the order in which they were administered: Vocabulary, Similarities, Arithmetic, Digit Span, Information, and Comprehension. The Letter-Number Sequencing subtest is a supplemental subtest which examiners did not

always administer. As with previously mentioned performance supplemental and optional subtests, there is no analysis of errors for this verbal supplemental subtest because the errors made would be an inaccurate representation of the actual errors made if all examiners had administered it in a uniform manner.

Vocabulary Subtest

Examinees must define words in the Vocabulary subtest. Possible administration errors include beginning the subtest at the incorrect starting point, failing to establish a correct basal level, discontinuing the subtest prematurely, or exceeding the discontinuance level of six consecutive errors. Verbal responses can warrant zero, one or two points depending on the quality of the response. Verbal analysis errors can include assigning an incorrect point value for a response, failing to query, and querying a response incorrectly. Computation errors include failing to add scores correctly. Table 43 contains a summary of errors for the Vocabulary subtest.

Table 43

Vocabulary Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	12-20	16-169	2-2	191
Ctr.	10-13	14-153	3-5	171

Administration errors. Twelve examiners from the experimental group made 20 administration errors. There were 11 errors of failing to establish a correct basal level, six errors of exceeding the discontinuance criteria, and three errors of discontinuing the subtest incorrectly. Ten examiners from the control group made 13 administration errors. There were eight errors of failing to establish a correct basal level and five errors of discontinuing the subtest incorrectly.

Verbal analysis errors. All of the examiners committed at least one verbal analysis error. The 16 examiners from the experimental group made 169 verbal analysis errors. Incorrectly assigning point values accounted for the majority of these errors (112 errors) followed, in turn, by failing to query (47 errors), and making incorrect queries (10 errors). Likewise, the 14 examiners from the control group made 153 verbal analysis errors, the majority of which (111 errors) were incorrectly assigning point values. The

next most frequent error was failing to query (31 errors) followed by making incorrect queries (11 errors).

Computation errors. Two experimental group examiners and three control group examiners made, respectively, two and five addition errors.

Table 44
Vocabulary Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	3-7	7
Ctr.	0-0	0

Missing data errors. Three examiners in the experimental group failed to record the responses made on seven occasions while no missing data errors were made by the control group examiners (See Table 44).

Similarities Subtest

For the Similarities subtest, the examinee is asked to explain how two words share common characteristics or meanings. As with the previous subtest, the administration errors include beginning at the incorrect starting point, failing to establish a correct basal level, discontinuing the subtest before the four consecutive errors are made and exceeding the discontinuance level. Verbal analysis errors can include assigning an incorrect point value for a response, failing to query, and querying a response incorrectly. Computation errors were failure to add the scores correctly. Table 45 contains a summary of the administration, verbal analysis, and computation errors for the Similarities subtest.

Table 45
Similarities Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	8-16	16-58	3-4	78
Ctr.	4-4	14-51	2-2	57

Administration errors. Twice as many examiners in the experimental group than in the control group committed administration errors: eight and four. The eight experimental examiners made 16 administration of failing to establish a correct basal level 13 times, exceeding the discontinuance criteria twice, and discontinuing the subtest

incorrectly once. Of the four examiners in the control group, two failed to establish a correct basal level and two exceeded the discontinuance criteria.

Verbal analysis errors. All examiners in both groups made verbal analysis errors. The numbers committed were approximately equal, given the different sample sizes. The error most frequently made was incorrectly assigning point values: 38 times for the experimental group and 42 times for the control group. There were 11 incorrect query errors made by examiners in the experimental group; the examiners in the control group committed this error once. There were similar numbers of failing to query errors: nine for the experimental group and eight for the control group.

Computation errors. Three experimental group examiners made four addition errors while two control group examiners each made an addition error.

Table 46
Similarities Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	3-6	6
Ctr.	1-1	1

Missing data errors. Three examiners in the experimental group failed to record their examinees' responses verbatim. Only one examiner from the control group failed to record responses verbatim for her examinees (see Table 46).

Arithmetic Subtest

The Arithmetic subtest contains a series of arithmetic questions. Examiners are to begin the subtest at the correct item and stop after making four consecutive errors. Consequently, the administration errors are incorrectly beginning the subtest, failing to establish a correct basal level, discontinuing the subtest prematurely, and exceeding the discontinuance criteria. Computation errors include adding scores incorrectly and failing to award bonus points for quick responses. Table 47 contains the summary of errors for the Arithmetic subtest.

Table 47
Arithmetic Subtest Administration and Computation Errors

Group	Administration errors	Computation errors	Total Errors
Exp.	4-7	4-5	12
Ctr.	2-2	4-5	7

Administration errors. Four examiners from the experimental group made seven administration errors. There were four errors of incorrectly establishing a basal level and three errors of exceeding the discontinuance criteria. Two examiners from the control group made two administration errors. There was one error of discontinuing the subtest incorrectly and one error of exceeding the discontinuance level.

Computation errors. In both groups, the same number of examiners, four, made the same number of errors, five. All errors committed by the experimental group examiners were addition errors while there were three addition errors and two errors of failing to award bonus points made by the control group examiners.

Table 48
Arithmetic Subtest Missing Data Errors

Group	Failure to record time	Failure to record correct/incorrect responses	Total Missing Data Errors
Exp.	7-15	7-10	25
Ctr.	9-14	6-6	20

Missing data errors. The missing data errors are shown in Table 48. Failing to record the time examinees took to answer questions accounted for the majority of the missing data errors: 7 examiners in the experimental group made 15 errors while 9 examiners in the control group made 14 errors. Failing to record the responses made by examinees accounted for ten errors made by seven examiners in the experimental group and six errors made by six examiners in the control group.

Digit Span Subtest

For the Digit Span subtest, examinees are required to repeat a series of numbers in order as relayed by the examiner. There are two series of numbers or two trials that form an item. All examinees begin with the first item and the administration rules state that the subtest is discontinued once the examinee makes errors on both trials of an item. Administration errors can occur if the examiner discontinues the subtest incorrectly or exceeds the discontinuance criteria. Computation errors include failing to add the scores correctly or assigning an incorrect point value for either the trial of the item or the item itself. Table 49 contains a summary of errors for the Digit Span subtest.

Table 49
Digit Span Subtest Administration and Computation Errors

Group	Administration errors	Computation errors	Total Errors
Exp.	1-2	0-0	2
Ctr.	2-2	0-0	2

Administration errors. For the experimental group, one examiner exceeded the discontinuance criteria twice. For the control group, one examiner exceeded the discontinuance criteria and another discontinued the subtest incorrectly.

Computation errors. There were no computation errors for either group.

Table 50
Digit Span Subtest Missing Data Errors

Group	Failing to record correct response	Failure to record incorrect response	Failure to record trial/item scores	Total Missing Data Errors
Exp.	10-29	5-10	7-12	51
Ctr.	7-23	8-15	5-11	49

Missing data errors. As shown in Table 50, 10 examiners in the experimental group failed to record the correct response for 29 of their examinees. Seven of the examiners in the control group, committed 23 of the same errors. Five experimental group examiners failed to record the incorrect responses made by 10 of their examinees while 8 examiners in the control group committed this error 15 times. Seven experimental group examiners and five control group examiners failed to record the trial/item scores for their examinees 12 and 11 times respectively.

Information Subtest

The Information subtest requires the examinee to respond to general knowledge-based questions. Administration errors can occur if the examiner begins at the incorrect starting point, fails to establish a basal level correctly, discontinues the subtest prematurely, and exceeds the discontinuance level of six consecutive errors. The verbal analysis errors consisted of incorrect point assignments, failing to query responses, and incorrectly querying responses. Computation errors include incorrectly adding the scores. Table 51 contains a summary of the Information subtest administration, computation, and verbal analysis errors.

Table 51
Information Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	7-12	8-23	0-0	35
Ctr.	6-6	6-13	0-0	19

Administration errors. Seven examiners in the experimental group made 12 administration errors. There were seven errors of incorrectly establishing a basal level, three errors of discontinuing the subtest incorrectly, and two errors of exceeding the discontinuance criteria. For the control group, six examiners made six administration errors. There were four errors of incorrectly establishing a basal level and two errors of discontinuing the subtest incorrectly.

Verbal analysis errors. Eight experimental group examiners made 23 errors while six control group examiners made 13 errors. The majority of these errors, 17 for the experimental group and 9 for the control group, were incorrect point assignments, followed by three and four errors, respectively, of failing to query a response. Three incorrect query errors were made by the experimental group examiners while none were made by the control group examiners.

Computation errors. No computation errors were made by the examiners of either group.

Table 52
Information Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	6-11	11
Ctr.	0-0	0

Missing data errors. Six examiners in the experimental group failed to record examinees responses verbatim on 11 occasions (See Table 52). There were no missing data errors for the control group.

Comprehension Subtest

Examinees are asked to provide an explanation or solution to everyday situations and an understanding of life occurrences or events for the Comprehension subtest. There are five questions requiring responses that reflect two general concepts. If not provided

by the examinee, examiners are instructed to ask for a second response that will complete the two concepts and warrant the full two points. Administration errors include beginning with the incorrect item, failing to establish a basal level correctly, discontinuing the subtest once the four consecutive errors are made, and exceeding the discontinuance level. The verbal analysis errors consist of incorrect point assignments, failing to query responses, incorrectly querying responses, and failing to request a second response. Computation errors include failing to add scores correctly. Table 53 contains a summary of errors for the Comprehension subtest.

Table 53
Comprehension Subtest Administration, Computation and Verbal Analysis Errors

Group	Administration errors	Verbal analysis errors	Computation errors	Total Errors
Exp.	8-15	16-125	3-3	143
Ctr.	9-14	14-95	0-0	109

Administration errors. Half of the examiners in the experimental group, 8, made 15 administration errors. The majority of these errors, 11, occurred when examiners failed to establish a correct basal level. Three errors were due to discontinuing the subtest incorrectly and one error was due to exceeding the discontinuance criteria. For the control group, 9 of the 14 examiners made 14 administration errors. These errors were more evenly distributed: eight errors of failing to establish a correct basal level and six errors of exceeding the discontinuance criteria.

Verbal analysis errors. The 16 examiners in the experimental group made 125 verbal analysis errors. Incorrectly assigning point values (77 errors) accounted for the majority of the errors with failing to obtain a second response (28 errors), failing to query (17 errors), and making incorrect queries (3 errors) accounting for the remaining verbal analysis errors. Similarly, all 14 examiners from the control group made verbal analysis errors. Of the 95 errors made, incorrectly assigning point values accounted for the majority of the errors (61 errors), followed by failing to obtain a second response (18 errors), failing to query (13 errors), and making incorrect queries (3 errors).

Computation errors. Three experimental group examiners committed three computation errors. None of the control group examiners made computation errors. Table 54 outlines the missing data errors.

Table 54
Comprehension Subtest Missing Data Errors

Group	Not Recording Responses Verbatim	Total Missing Data Errors
Exp.	4	4
Ctr.	0	0

Missing data errors. Two of the examiners in the experimental group failed to record their examinees' responses verbatim on four occasions. There was no missing data for the control group examiners.

Summary of Verbal Subtest Errors

The control group examiners committed fewer administration errors than the experimental group examiners on five of the six verbal subtests (i.e., Vocabulary, Similarities, Arithmetic, Information, and Comprehension). The Digit Span subtest represents the one subtest where both examiners made the same number of errors: two administration errors and no computation errors for either group. Both groups made the same number of computation errors on three subtests (Arithmetic, Digit Span, and Information) and the control group examiners committed fewer computation errors on two of the three remaining subtests (Similarities and Comprehension). There was a large discrepancy for the verbal analysis errors between the groups compared to the other types of errors: Vocabulary 169 vs.153, Similarities 58 vs.51, Information 23 vs.13, and Comprehension 125 vs. 95 (errors respectively between the experimental and control groups) and the control group committed fewer verbal analysis errors. There was a large discrepancy between the total number of subtest errors for both groups of examiners on four of the subtests in favour of the control group who committed fewer errors: (Vocabulary 191 vs. 171, Similarities 78 vs. 57, Information 35 vs. 19, Comprehension 143 vs. 109). There was less of a discrepancy between the scores for the Arithmetic subtest (i.e., 12 vs. 7) in favour of the control group. There were fewer missing data errors for the control group on all the verbal subtests; however, there was much less variability between the two groups with respect to the missing data errors (range of errors = 2-11).

CHAPTER SEVEN

Results

WAIS-III Total Errors

The WAIS-III total errors are the sum of errors made on the four cover pages of the fold-out WAIS-III record form. Each page is labeled as follows: Demographics Page, Score Conversion Page, Profile Page, and Discrepancy Analysis Page.

The Demographics Page contains space for recording the identifying information about the examinee including name, chronological age calculation, gender, address, highest level of education achieved, and the examiner's name. Behavioral observations are also recorded on this page with headings such as referral source, language, attitude toward testing, physical appearance, attention and concentration, and visual/auditory/motor problems.

The Score Conversion Page is used to make the initial subtest score conversions and contains four tables. In the first table, subtest raw scores, subtest scaled score conversions, reference group scaled scores from the standardized sample of same age adults, and the sum of subtest scaled scores are recorded. In the second table, the average verbal and performance subtest scores are recorded. The third table is the Optional Procedures table where raw scores and cumulative percentile conversions from the two optional procedures (Incidental Learning and Copy) of the Digit Symbol-Coding subtest are recorded. The fourth table is used for determining strengths and weaknesses for an examinee as determined by a qualitative analysis of the different subtest scores. The individual subtest scaled scores, mean scores, difference from the mean scores, statistical significance level, and determination of the subtest as a relative strength or weakness for a particular examiner are all recorded on this table. A column for listing the frequency of differences as obtained by the individual examinee compared to the same frequency of difference as obtained by the standardized sample is also provided.

The Profile Page contains one table to record the IQ and Index scores and two graphs to visually display the subtest, IQ, and Index scores. In the table, the sums of scaled scores, the IQ/Index scores, percentiles, and confidence intervals are recorded. The IQ and Index scores are plotted onto the first graph and the subtest scaled scores are plotted onto the second graph.

The Discrepancy Analysis Page contains two tables for showing discrepancies between obtained scores. In the first table, the differences between IQ and Index scores are recorded as well as the statistical significance level required for a discrepancy and the frequency of such differences as obtained from the standardized sample. In the second table, the differences between the digits forward and digits backward spans of the Digit Span subtest are recorded as well as the frequency of such differences as obtained from the standardized sample.

Examiners made errors on the record form that fit into three total error types:

1- Conversion errors (i.e., incorrect calculations and incorrect conversion of scores).
2- Plotting errors (i.e., errors that are committed when transferring scores onto the tables and plotting scores onto the graphs).
3- Missing data errors (i.e., not recording the examinee's demographic information and behavioral observations, not recording information in the tables, and failing to fill in the graphs).

Table 55 contains a summary of the frequency of these three error types committed by participants in each group according to the page on the record form.

Table 55
WAIS-III Total Errors by Group, Error Type and Page

	Error Type										
Group	Conversion Errors				Plotting Errors			Missing Data Errors			
	Demo	SCP	Prof	DAP	SCP	Prof	DAP	Demo	SCP	Prof	DAP
Exp.	2	71	6	23	10	40	30	64	27	1	27
Cntrl.	0	57	6	23	13	22	21	43	34	0	20

[a] Demo = Demographic Page
[b] SCP = Score Conversion Page
[c] Prof = Profile Page
[d] DAP = Discrepancy Analysis Page

Table 56 provides a summary of the total number of errors by group according to error type.

Table 56
WAIS-III Total Errors by Group and Error Type

	Error Type			
Group	Conversion Errors	Plotting Errors	Missing Data Errors	Total Errors
Exp.	102	80	119	301
Cntl.	86	56	97	239

Conversion Errors

All error types are presented according to the cover page in which the errors occur. For example, the Demographics Page contains one potential conversion error: age miscalculation which is discussed first followed by the conversion errors from the Score Conversion Page, Profile Page, and Discrepancy Analysis Page.

Demographics Page

Two examiners from the experimental group made two age calculation errors while there were no examiners from the control group who made age calculation errors. The first experimental examiner made an incorrect calculation of her examinee's age and the wrong norm table was used. This initial error produced eight scaled score conversions and five sum of scaled scores errors resulting in a one point POIQ and PSI deficit, a four point PIQ deficit, and a three point FSIQ deficit. The second examiner made an error calculating the month of her examinee's chronological age but her error did not affect the norm table used or subsequent IQ conversions.

Score Conversion Page

Only three of the four tables on the Score Conversion page were analyzed for errors. The Optional Procedures table for the Digit Symbol-Coding subtest was not included in the analysis of errors because not all examiners administered these optional subtests. A summary of the conversion errors for the first two tables of the Score Conversion Page is presented in Table 57. A summary of the conversion errors for the third table, Determining Strengths and Weaknesses table, which is also located on the Scores Conversion Page, is presented in Table 58.

Table 57
Score Conversion Errors: First Two Tables of the Score Conversion Page

	Error Type					
Group	Errors converting subtest raw scores to subtest scaled scores	Incorrect addition of subtest scaled scores	Incorrect selection of subtests used for Index and IQ scores	Using wrong ref. group scaled scores	Incorrect conversion of mean scores	Total Errors
Exp.	4-5	3-4	0-0	0-0	7-17	26
Cntrl	4-5	2-2	0-0	1-1	9-16	24

Incorrect conversion of subtest raw scores to subtest scaled scores. The first number in the table represents the number of examiners who made errors listed in that column and the second number is the total number of errors made. For both groups, four examiners committed five errors when converting subtest raw scores to subtest scaled scores. For the experimental group, the first examiner incorrectly converted the Letter-Number Sequencing subtest raw score to a scaled score that was two points higher than it should have been. This subtest is a supplemental subtest and not used in the calculation of the VIQ or FSIQ although, the subtest scaled score figures for the WMI was, as a result, four points incorrectly elevated. The second examiner incorrectly converted the raw subtest scores to subtest scaled scores for the Vocabulary, Arithmetic, and Comprehension subtests resulting in a one point elevation for all three scaled scores. The VIQ was subsequently three points higher than it should have been and the FSIQ two points higher than it should have been. The third examiner made two such errors, one for her second examinee and one for her fourth examinee. For her second examinee, she incorrectly converted the Block Design raw score to a scaled score that was one point higher than it should have been. The result was a two-point elevation on the PIQ and a one-point elevation on the FSIQ. For her fourth examinee, she incorrectly converted the raw Vocabulary subtest score as ten points lower than it should have been resulting in a two point subtest scaled score deficit. The impact was a two-point deficit for the VIQ and a one-point deficit for the FSIQ. The fourth examiner made one error for her first examinee by incorrectly converting the vocabulary raw score to a scaled score that was one point higher than it should have been resulting in a one-point elevation on the VIQ and a one-point elevation on the FSIQ.

For the control group, the first examiner made an error when converting subtest raw scores to subtest scaled scores with her second and fourth examinees. For her second examinee, she incorrectly converted the Vocabulary, Arithmetic, and Comprehension raw subtest scores to scaled scores resulting in an erroneous three-point elevation for the VIQ and two-point elevation for the FSIQ. For her fourth examinee, she incorrectly converted the Comprehension raw subtest score to a scaled score that was one point too high resulting in one-point elevations for both the VIQ and the FSIQ. The second examiner incorrectly converted the raw subtest score for the Letter-Number Sequencing subtest to a scaled score that was one point too low resulting in a two-point deficit on the WMI. The third examiner made an error in converting the raw subtest score for the Matrix Reasoning subtest resulting in a two-point elevation for the subtest scaled score and a one-point elevation on the PIQ. The fourth examiner made an incorrect conversion of the raw subtest score for the Symbol Search subtest resulting in a one point elevated subtest scaled score and a subsequent erroneous one point elevation on the VIQ and FSIQ.

Incorrect addition of subtest scaled scores. Three examiners from each group made addition errors when adding the subtest scaled scores to arrive at the sums of scaled scores for the IQ and Index scores. For the experimental group there were four addition errors committed by three examiners. The first examiner incorrectly added the performance subtest scaled scores which were three points too high resulting in a four-point elevation on the PIQ. The first examiner also incorrectly added the performance subtest scaled scores to form the PIQ for her third examiner. The sums of scaled scores was one point too low resulting in a two-point deficit on the PIQ. The second examiner incorrectly added the subtest scaled scores to form the WMI for her fourth examinee with the sums of scaled scores one point too high resulting in a two point elevation on the WMI. The third examiner made an error for her third examinee when adding the sums of scaled scores for the WMI. She incorrectly calculated the sums of scaled scores with a one-point deficit resulting in a two point WMI deficit.

For the control group, there were two addition errors committed by two examiners. The first examiner made an addition error with her first examinee by incorrectly adding the subtest scaled scores. She had a seven point deficit for the sums of scaled scores for the VCI resulting in a twelve-point deficit for the VIC. The second

examiner made an incorrect calculation with a ten-point excess on the sums of scaled scores resulting in a POI that was 24 points higher than it should have been.

Incorrect selection of subtests used for Index and IQ scores. No examiners from either group made errors in selecting the correct subtests to calculate the Index and IQ scores.

Using wrong reference group scaled scores. Only one examiner from the control group made an error by recording the incorrect reference group scaled scores for her third examinee. There were no errors of using the wrong reference group scaled scores for the experimental group.

Incorrect conversion of mean scores. The first step to determine examinee individual strengths and weaknesses at the subtest level was to calculate the verbal, performance, and overall mean scores for an examinee. All verbal subtest scaled scores were summed and divided by the number of total verbal subtests administered in order to obtain the mean verbal score. The same procedure occurred for the performance and the combination of verbal and performance scores to obtain the overall subtest mean score.

Seven experimental group examiners and nine control group examiners made close to the same number of errors (17 for the experimental group and 18 for the control group) when converting mean scores for the verbal and performance subtests and the mean overall score for the combination of verbal and performance subtests.

Of the 17 errors made by examiners from the experimental group, 10 were incorrect conversion of verbal scores, 4 were incorrect conversions of performance scores, and 3 were incorrect overall mean conversions. The differences between the incorrect and correct mean scores were not large (e.g., obtained performance mean of 6.2 that should have been 6.4) with seven errors reaching a one point difference. No error exceeded a one-point difference.

The pattern of errors was different for the control group with 4 of the 18 errors incorrect conversion of verbal scores, 9 incorrect conversions of the performance scores, and 5 incorrect overall mean conversions. Similar to the experimental group, the differences between the incorrect and correct mean scores were not large and no error difference exceeded a one-point value.

For both groups, miscalculation was the source of the majority of the errors although two errors from each group were attributed to examiners failing to include all verbal or performance supplemental subtests in the mean score calculations.

The third table located on the Score Conversion Page was used to determine subtest strengths and weakness for examinees. Table 58 summarizes the errors made when examiners entered information into this table.

Table 58
Score Conversion Errors: Score Conversion Page: Determining Strengths and Weaknesses

	Error Type				
Group	Incorrect difference from mean	Incorrect Freq. of Diff. in Std. Sample	Incorrect determination of Strength	Incorrect determination of Weakness	Total Errors
Exp.	7-13	11-21	3-4	3-7	45
Cntrl	10-15	7-13	3-3	2-2	33

Incorrect difference from the mean. After the subtest scores for the mean verbal and performance scores and overall means are calculated, the next step in determining individual strengths and weaknesses was to subtract the mean score from the individual subtest scaled score. This calculation was made to determine, relative to the standardized sample, which subtests could be determined as strengths and which subtests could be determined as weaknesses for an examinee. As shown from Table 58, 7 experimental examiners and 10 control examiners made 13 and 15 calculation errors, respectively, when calculating the difference between the scaled score and the mean score.

Incorrect frequency of difference in the standardized sample. Examiners consulted Table B.3 in Appendix B of the *WAIS-III Administration and Scoring Manual* to locate the frequency of differences obtained by the standardized sample at the significance level of 0.15 or 0.05. More examiners from the experimental group, 11, made 21 errors in reading the table to determine the frequency of a difference in the standardization sample for a particular strength or weakness. By comparison, fewer examiners in the control group, 7, made fewer errors, 13, when determining this frequency.

Incorrect determination of strength. Three examiners from each group made similar numbers of errors (four for the experimental group and three for the control group) in determining which subtest was considered a strength after the calculation of the scaled and mean score was completed. A subtest is considered a strength when the examinee's difference from the mean is equal to or greater than the value listed in the table. In total, six examiners from both groups failed to correctly make the determination of a subtest strength.

Incorrect determination of weakness. Similar to the previous error of incorrectly determining subtest strengths, three examiners from the experimental group made seven errors and two examiners from the control group made two errors when attempting to determine which subtest was considered a weakness. A subtest is considered a weakness when the examinee's difference from the mean is less than the value listed in the table. In total, five examiners from both groups failed to correctly make the determination of a subtest weakness.

The majority of errors for the Score Conversion Page were calculation errors. These errors occurred most frequently when examiners incorrectly determined mean score values for the verbal, performance, and overall means and when examiners calculated the difference between the scaled score and mean scores to determine strengths and weaknesses.

Profile Page

The Profile Page contains one table to record the sums of scaled scores, the IQ and Index scores, and their percentiles and confidence intervals. The two graphs provide a visual display of the IQ and Index scores and the subtest scaled scores. The score conversion errors for the Profile Page are summarized in Table 59.

Table 59

Score Conversion Errors: Profile Page Errors

	Error Type		
Group	Incorrect calculation of sums of scaled scores to IQ/Index scores	Incorrect conversion of IQ/Index scores to Confidence Intervals and %	Total Errors
Exp.	3-4	2-2	6
Cntrl	1-1	2-5	6

Incorrect calculation of sums of scaled scores to IQ/Index scores. Three examiners from the experimental group committed four errors by calculating the incorrect sums of scaled scores to IQ/Index scores. The first examiner recorded the VIQ for her second examinee as 101 when it should have been 102; the second recorded her third examinee's VCI as 82 when it should have been 84; the third recorded her third examinee's PIQ as 105 when it should have been 106 and the FSIQ as 111 when it should have been 112. One examiner from the control group erroneously recorded the POI confidence interval for her third examinee as 110 when it should have been recorded as 111. All errors were due to the examiners reading the wrong line from Tables A.3 to A.8 in Appendix A of the *WAIS-III Administration and Scoring Manual.*

Incorrect conversion of IQ/Index scores to confidence intervals and percentiles. Examiners from both groups committed errors when converting the IQ/Index scores to confidence intervals and percentiles. For the experimental group, two examiners committed two errors. The first examiner recorded the VCI confidence interval for her second examinee as 110-126 when it should have been 110-121. The second examiner recorded the VIQ confidence interval for her third examiner as 92-103 when it should have been recorded as 92-102. For the control group, two examiners committed five errors when converting the IQ/Index scores to confidence intervals and percentiles. The first examiner made four errors for her first, second, and third examinees. For her first examinee, she recorded the POI confidence interval as 107-119 when is should have been recorded as 106-120. For her second examinee, the FSIQ confidence interval was recorded as 95-105 when it should have been recorded as 96-104. For her third examinee, the PIQ confidence interval was recorded as 94-110 when it should have been recorded as 95-109 and the FSIQ confidence interval was recorded as 99-109 when it should have been recorded as 100-108. The second examiner erred by recording the VIQ confidence interval for her second examinee as 89-97 when it should have been recorded as 88-99. Similar to the previous error type, all errors were due to the examiners reading the wrong line from Tables A.3 to A.8 in Appendix A of the *WAIS-III Administration and Scoring Manual.*

All conversion errors made by both groups of examiners on the Profile Page were due to incorrect recording of information from the tables in the manual.

Discrepancy Analysis Page

There are two tables located on the Discrepancy Analysis Page. The first table is used to make discrepancy comparisons between the IQ and Index scores. Once the examiners made the calculation of differences between these scores, the statistical significance level and frequency of difference in the standardized sample were both obtained to determine significant and non-significant discrepancies between the IQ and Index scores. The second table is provided to allow comparisons for the Digit Span subtest between the longest digits forward and backward and the frequency of obtaining such differences in the standardized sample. The score conversion errors for the Discrepancy Analysis Page are summarized in Table 60.

Table 60
Score Conversion Errors: Discrepancy Analysis Page

	Error Type					
Group	Incorrectly calculating difference	Incorrectly recording stat. Signif.	Incorrectly recording Freq. of Diff.in stnd. Sample (table 1)	Incorrectly calculating difference in digits	Incorrectly recording Cumulative %ages in stnd. Sample (table 2)	Total errors
Exp.	7-9	2-3	2-3	3-4	4-4	23
Cntrl.	4-8	2-2	3-4	3-7	2-2	23

Incorrectly calculating the difference between the IQ and Index scores. The first step in making the discrepancy comparisons is to calculate the differences between the selected IQ and Index scores. Seven examiners from the experimental group and four examiners from the control group made nine and eight errors, respectively, when calculating the difference between the selected IQ and Index scores.

Incorrectly recording the statistical significance level for the Discrepancy Comparison Table. Examiners consulted Table B.1 in Appendix B of the *WAIS-III Administration and Scoring Manual*, identified their examinee's age, chose a 0.05 or a 0.15 significance level of significance and copied the differences between the IQ and Index scores required for statistical significance onto the Discrepancy Analysis Page Table. Two examiners from each group made three and two errors respectively when recording the statistical significance level on the Discrepancy Comparison Table. For

example, one examiner from the control group chose a statistical significance level of .05 and erroneously recorded the information for the .15 statistical significance level.

Incorrectly recording the frequency of difference in the standardized sample for the Discrepancy Comparison Table. Examiners consulted Table B.2 in Appendix B of the *WAIS-III Administration and Scoring Manual* to locate the frequency of the discrepancies for each individual comparison obtained by the standardization sample and compare them to the amount of discrepancy found for their examinee's individual comparisons. The examiners recorded the frequency of difference in the standardization sample in the column provided. Two examiners in the experimental group made three errors when recording this information for their examinees. For the control group, three examiners made four errors when recording this information for their examinees.

Incorrectly calculating the difference in digits for the Digit Span subtest. The second table located on the Discrepancy Analysis Page is used for recording the longest Digit Span forward and the longest Digit Span backward that was recalled by the examinee. The examiners are to write down this information and subtract the longest digits forward by the longest digit backward. Three examiners from each group made this subtraction error. There were four of such errors for the experimental group and seven for the control group.

Incorrectly recording cumulative percentages in the standardized sample for the Digit Span subtest. Table B. 6 in Appendix B of the WAIS-III Administration and Scoring Manual contains cumulative percentages of the longest digits forward and digits backward spans. Examiners were to locate the longest digit forward and backward made by their examinees and record the corresponding cumulative percentage achieved from the standardized sample. Table B.7 provided the cumulative percentages of difference between the longest digits forward and digits backward spans. Twice as many experimental examiners than control examiners (four vs. two) made twice as many errors (four vs. two) when recording this information.

Both groups of examiners made the same number of total score conversion errors for the Discrepancy Analysis Page when performing calculations to arrive at differences between scores and when consulting the tables to correctly record information.

Plotting Errors

Plotting errors were errors committed when transferring scores onto the tables and plotting scores onto the graphs. Plotting errors were found on the Score Conversion Page, the Profile Page, and the Discrepancy Analysis Page.

Score Conversion Page

The plotting errors for this page included incorrectly transferring raw subtest scores to the first table and incorrectly transferring scaled scores to the Determining Strengths and Weaknesses Table. A summary of these errors is presented in Table 61.

Table 61

Plotting Errors: Score Conversion Page

	Error Type		
Group	Incorrectly transferring raw subtest scores	Incorrectly transferring scaled scores	Total errors
Exp.	3-3	5-7	10
Cntrl.	3-3	7-10	13

Incorrectly transferring raw subtest scores. The same number of examiners from both groups, three, committed the same number of errors, three, when transferring the raw subtest scores from the pages of the record form onto the first table of the Score Conversion Page.

Incorrectly transferring scaled scores. Five examiners in the experimental group made seven errors when transferring the obtained scaled scores from the first table to the Determining Strengths and Weaknesses Table. For the control group, seven examiners committed ten of these transferring errors.

Both groups committed a similar number of errors when transferring scores from the pages of the record form to the table and from one table to another. These appeared to be all errors of inattention.

Profile Page

The plotting errors for this page included incorrectly transferring sums of scaled scores to the IQ/Index Scores Table and incorrectly plotting IQ/Index scores onto the first graph and incorrectly plotting subtest scaled scores onto the second graph. A summary of the plotting errors made on the Profile Page is provided in Table 62.

Table 62
Plotting Errors: Profile Page

	Error Type			
Group	Incorrectly transferring sums of scaled scores	Incorrectly plotting IQ/Index scores	Incorrectly plotting subtest scaled scores	Total errors
Exp.	3-3	12-17	11-20	40
Cntrl.	3-3	5-8	8-11	22

Incorrectly transferring sums of scaled scores. The same number of examiners from both groups, three, committed the same number of errors, three, when transferring the sums of scaled scores from the first table of the Score Conversion Page to the IQ/Index Scores Table of the Profile Page.

Incorrectly plotting IQ/Index scores. There were twice as many errors made by the experimental group, 17, in comparison to the control group, 8, when 12 experimental examiners and 5 control examiners committed errors in plotting the IQ and Index scores onto the graph.

Incorrectly plotting subtest scaled scores. Both groups committed more errors when plotting the subtest scaled scores in comparison to the previous error type of plotting the IQ and Index scores. Eleven experimental group examiners and eight control group examiners committed these plotting errors, 20 and 11 times respectively.

The plotting errors for the Profile Page appear attributed to errors of inattention and carelessness when examiners erroneously plotted scores onto the wrong lines of the graphs.

Discrepancy Analysis Page

The plotting errors for this page included incorrectly transferring IQ and Index scores onto the Discrepancy Comparisons table and incorrectly transferring Digit Span scores onto the Digit Span table. Table 63 contains a summary of the plotting errors for the Discrepancy Analysis Page.

Table 63
Plotting Errors: Discrepancy Analysis Page

	Error Type		
Group	Incorrectly transferring IQ and Index scores	Incorrectly transferring Digit Span scores	Total errors
Exp.	10-21	5-9	30
Cntrl.	5-5	9-16	21

Incorrectly transferring IQ and Index scores. When transferring the IQ and Index scores from the IQ/Index Scores Table of the Profile Page to the Discrepancy Comparisons Table of the Discrepancy Analysis Page, twice as many experimental examiners, 10, made 21 errors in comparison to five control examiners who committed five such errors.

Incorrectly transferring Digit Span scores. Five experimental group examiners committed nine errors and nine control group examiners committed 16 errors when transferring the longest digit span forward and longest digit span backward from the Digit Span page in the record form to the Digit Span Table on the Discrepancy Analysis Page. One examiner in the experimental group erroneously used the total scores for the Digit Span forward and backward rather than recording the longest digit span forward and backward. She made this error on all four protocols. Two examiners from the control group also made this error seven times.

Similar to the total number of plotting errors on the Profile Page, the experimental group made more plotting errors than the control group on the Discrepancy Analysis Page.

Missing Data Errors

Missing data errors were committed on all four pages of the WAIS-III record form. Like the previous error types, the missing data errors are discussed according to the order of the pages in which they occur.

Demographics Page

Examiners from both groups omitted including data on this page, particularly in terms of not recording demographic information and not recording behavioural observations. The Missing Data errors for the Demographics Page are summarized in Table 64.

Table 64
Missing Data Errors: Demographics Page

	Error Type		
Group	Not recording demographic information	Not recording behavioural observations	Total errors
Exp.	14-32	10-32	64
Cntrl.	10-31	5-12	43

Not recording demographic information. Fourteen of the 16 examiners in the experimental group omitted some or all of their examinee's demographic information. Most of these examiners, 10, omitted their examinees' addresses more frequently than the examinees' education level, and examiner's name. One examiner omitted all demographic information for all four of her examinees. Likewise, 10 of the 14 examiners in the control group omitted some or all of their examinee's demographic information. Seven omitted their examinee's address more frequently than the examinee's education level, gender, or the examiner's name. Similar to the experimental group, one examiner omitted all demographic information for all four of her examinees. The control group omitted recording the examinees' gender four separate times, but no such omission was made with the experimental group.

Not recording behavioural observations. Twice as many experimental examiners than control examiners omitted recording behavioural observations for their examinees. Ten experimental group examiners made this error of omission 32 times while five control examiners made this error of omission 12 times. While these examiners may not have recorded behavioural observations in the space provided on the Demographics Page, 11 of the examiners from the experimental group who omitted recording behavioural observations in the space provided nonetheless recorded behavioural observations throughout 17 of the protocols while 8 of the control examiners recorded some behavioural observations throughout 15 protocols.

Both groups of examiners omitted information for their examinees on the Demographics Page. The experimental examiners made the same number of omissions for demographic information and behavioural observations while the control group made fewer omissions for the behavioural observations.

Score Conversion Page

Examiners made omissions when completing the mean scores for the verbal, performance, and overall mean scores and in completing the Determining Strengths and Weaknesses Table. The Missing Data errors for the Score Conversion Page are summarized in Table 65.

Table 65
Missing Data Errors: Score Conversion Page

	Error Type		
Group	Failing to complete mean scores	Failing to complete strengths and weaknesses	Total errors
Exp.	3-3	9-24	27
Cntrl.	4-6	9-28	34

Failing to complete mean scores. Three experimental group examiners omitted including the mean score information three times each. In contrast, four control group examiners omitted this information six times in total.

Failing to complete strengths and weaknesses. The same number of examiners from each group, nine, failed to completely fill in the Determining Strengths and Weaknesses Table. For the experimental group, the frequency of difference in the standardization sample column was not filled in on 16 occasions, the statistical significance level column was not filled in on 4 occasions, and the entire table was left blank on 4 occasions. For the control group, the frequency of difference in the standardization sample column was not filled in on 17 occasions, the statistical significance level column was not filled in on 8 occasions, and the entire table was left blank on 3 occasions.

The control group examiners omitted more information than the experimental group examiners on the Score Conversion Page.

Profile Page

The Profile Page contained only one missing data error type, which was failing to fill in the graphs. Only one examiner from the experimental group failed to fill in the graphs for her first examinee. None of the control group examinees failed to fill in the graphs.

Discrepancy Analysis Page

Examiners made omissions when completing the mean scores for the verbal, performance, and overall mean scores and in completing the Determining Strengths and Weaknesses Table. Table 66 contains a summary of the Missing Data errors for the Discrepancy Analysis Page.

Table 66
Missing Data Errors: Discrepancy Analysis Page

	Error Type		
Group	Failing to complete Discrepancy Comparisons table	Failing to complete Digit Span table	Total errors
Exp.	8-17	6-10	27
Cntrl.	6-14	4-6	20

Failing to complete the Discrepancy Table. Examiners from both groups failed to entirely complete the Discrepancy Comparisons table on the Discrepancy Analysis Page. Eight experimental group examiners failed to complete some portion of this table on 17 occasions while six control group examiners made this same type of omission on 14 occasions.

Failing to complete the Digit Span Table. Six experimental group examiners omitted completing filling out the Digits Span Table on 10 occasions, and four control group examiners also made this omission on six occasions.

While both groups made omissions on the Discrepancy Analysis Page, the experimental group examiners made more of these omissions than the control group examiners.

Summary of WAIS-III Total Errors

In terms of conversion errors, the majority of these errors occurred on the Score Conversion Page for both groups of examiners and calculation errors accounted for most of these errors. In contrast, errors in converting scores such as recording information from tables accounted for far fewer of the total conversion errors. The experimental group examiners committed more conversion errors on the Score Conversion Page than the control group examiners. A similar pattern of larger number of calculation errors compared to recording errors was observed for the Discrepancy Analysis Page with both

groups of examiners committing the same number of errors. Overall, the experimental group examiners committed more conversion errors.

In regard to the plotting errors, the majority of these errors occurred on the Profile Page, particularly when examiners when plotting scores onto the graphs. There were also a large number of plotting errors committed on the Discrepancy Analysis Page when examiners transferred IQ and Index scores onto the table although the experimental group were able to transfer Digit Span scores more accurately than the control group. Overall, the experimental group made more plotting errors than the control group.

Students from both groups committed more missing data errors than any of the other error types although the experimental group committed more of these errors of omission compared to the control group. The Demographic Page contained the highest number of missing data errors with both groups committing a similar number of demographic information omissions. While the experimental examiners tended to omit recording behavioural observations in the space provided on the Demographics Page, they instead recorded comments more frequently throughout the protocol than members of the control group suggesting that they did recognize the importance of including such observations somewhere on the WAIS-III protocol. On the Score Conversion Page, many examiners from both groups failed to complete the strength and weaknesses table while a number of them failed to complete the Discrepancy Comparisons Table on the Discrepancy Analysis Page.

The control group made fewer total errors according to all error types. When differences in group size was accounted for (i.e., 16 examiners in the experimental group and 14 examiners in the control group) the average number of errors committed by examiners in the experimental group was 18.8 and 17.0 for the control group examiners. Therefore, the experimental group committed more total errors than the control group. Both groups learned at the same time about the general administration information and the experimental group was provided with some additional time to learn more specific information about accurately completing the four fold out pages of the record form. There was an expectation that the experimental examiners would have committed fewer total errors than the control group, however, this was not the result.

CHAPTER EIGHT

Summary, Conclusions, and Recommendations

This chapter first contains a summary of the purpose, procedures, and findings of the study. This summary is then followed, in turn, by the limitations of the study and the conclusions drawn in light of the limitations. The chapter concludes with implications for practice and recommendations for future research.

Summary of Purpose and Procedures

The purpose of this study was to determine if the number of errors committed when administering and scoring the WISC-III and WAIS-III would be reduced through the application of a specific teaching method designed to train psychology students who are learning to administer and score the tests. First, a new method of teaching the administration and scoring of the Wechsler Intelligence tests was developed. Second, the number and types of administration and scoring errors on the WISC-III and the WAIS-III committed by students who were instructed using the new teaching method (the experimental group) were compared to the student administration and scoring errors committed by students who received the unaltered instruction method (the control group). A non-equivalent group design was used in which the students chose their groups based on the partner with whom they were sharing a testing kit.

The participants were graduate level students enrolled in the Individual Assessment Course (Educational Psychology 545) offered by the Department of Educational Psychology, University of Alberta. The students, from two teaching terms, 1998-99 Winter session and the 1999 Spring/Summer session, were included. Comparison of the demographic characteristics, as assessed through information provided by the students, did not reveal any between session differences nor differences between the experimental and control groups. The two groups were the same at the beginning of the research project.

The experimental group received the experimental teaching method that contained two teaching interventions: a) a lecture on specific and frequently made administration and scoring errors and b) an observation of a Wechsler administration. These experimental teaching interventions were reported as effective in reducing the number of administration and scoring errors in previous research findings (Blakey et al., 1987,

McQueen et al, 1994, Moon et al, 1986, Slate et al, 1991, & Slate et al, 1992). The control group received the regular teaching method commonly used to teach administration and scoring of the Wechsler tests. The control groups were engaged in practice administrations of the Wechsler tests during the two experimental interventions. In other research (Slate et al., 1991, 1992) practice administrations were not found to be effective in reducing scoring and administration errors. Two instructors taught the experimental groups and two different instructors taught the control groups. Teaching agendas were developed for the study and supplied to the instructors to ensure that the same information was being taught to each of the groups.

As part of the course requirements, students were required to administer, score, and interpret in report format two Wechsler tests (i.e., WISC-III and the WAIS-III) four times each. The four WISC-III and four WAIS-III protocols were each initially reviewed for administration and scoring errors by the student's supervisor. The supervisor's task was to oversee the student examiner's work by grading the student's test protocols and psychological reports and signing the final report prior to the client receiving it.

Two checklists, one for the WISC-III and one for the WAIS-III, were developed as tools to identify and calculate the numbers of student examiner errors committed on the protocols. Research assistants used these checklists to record the types and numbers of errors committed in the final protocols. Subtest errors were the sums of errors made on each of the subtests while total errors were the sums of errors made on the cover page and the reverse cover page of the record form.

Summary of the WISC-III Subtest Errors

The WISC-III subtest errors consisted of administration, computation, and missing data errors for both the verbal and performance subtests. Verbal Analysis Errors, characterized by incorrect questioning and incorrect point assignment, was an additional error category for the verbal subtests and the Picture Completion performance subtest. The subtest errors tended to be haphazard with no systematic errors across examiners. In general, there was little discrepancy between the number of errors made by the experimental and control groups on the performance subtests. In contrast, the control group committed fewer errors than the experimental group on the verbal subtests.

Of the five performance subtests, the Picture Arrangement subtest had the highest number of administration errors for both groups, particularly regarding beginning the subtest incorrectly. The Object Assembly subtest had the fewest number of administration errors for both groups. At the same time, the highest number of computation errors (mainly related to incorrectly multiplying the correct number of junctures completed by the examinee and rounding up half scores) were committed on the Object Assembly subtest. The Block Design contained the fewest number of computation errors but the highest number of missing data errors.

Of the six verbal subtests, the Vocabulary subtest had the highest number of administration errors for both groups and these were mainly related to incorrectly beginning and ending the subtest. The Digit Span subtest had the fewest number of administration errors for both groups. In regard to computation errors, the Arithmetic subtest had the highest number of computation errors due to incorrect calculation of scores while the Comprehension and Vocabulary subtests had the lowest number of computation errors. A large number of missing data errors was reported for the Digit Span subtest compared to missing data errors for the other verbal subtests.

For both groups, there were more verbal analysis errors than any other error category. The Comprehension subtest contained the highest number of verbal analysis errors, followed by the Vocabulary, Similarities, Picture Completion, and Information subtests. Regarding the type of verbal analysis errors, incorrect point assignments occurred most frequently followed by query errors and failing to obtain a second response for the Comprehension subtest. The experimental group committed a greater number of verbal analysis errors on the Similarities and Vocabulary subtests than the control group.

Summary of the WAIS-III Subtest Errors

Similar to the WISC-III, the subtest errors consisted of administration, computation, and missing data errors for both the verbal and performance subtests. The verbal subtests and the Picture Completion performance subtest had an additional error category, Verbal Analysis Errors. The WAIS-III subtest errors tended to be haphazard with no systematic errors across examiners within groups or between groups. In general, there tended to be fewer performance subtest errors compared to the verbal subtest errors

for both groups. Furthermore, the control group committed fewer administration, computation, and missing data errors on both the performance and verbal subtests compared to the experimental group.

Of the five performance subtests, the Matrix Reasoning subtest had the highest number of administration errors for both groups. This is a new subtest for the WAIS-III and it posed administrative challenges for both groups of student examiners who committed a similar number of administration errors on this subtest. These errors included failing to administer the sample items, failing to establish a correct basal level, discontinuing the subtest incorrectly, and exceeding the discontinuance criteria. The Digit Symbol-Coding subtest had no administration errors for either group. Despite the lack of administration errors, the Digit Symbol-Coding subtest had the highest number of computation errors for both groups while the Matrix Reasoning subtest had the fewest. In terms of missing data errors, the Block Design subtest contained the highest number of missing data errors.

Of the six verbal subtests, the Vocabulary subtest had the highest number of administration errors for both groups with the majority of errors due to failing to begin the subtest correctly. There were very few computation errors for any of the verbal subtests: neither group committed any computation errors on the Digit Span and Information subtests and the control group committed no errors for the Comprehension subtest. In terms of missing data errors, the Digit Span subtest contained the highest number of missing data errors.

The majority of errors for the verbal subtests were the verbal analysis errors. For both groups, the Vocabulary subtest contained the highest number of verbal analysis errors followed by the Comprehension, Similarities, and Information subtests. The majority of the verbal analysis errors were incorrect point assignments followed by questioning errors. For all the verbal subtests, the experimental group committed more verbal analysis errors than the control group.

Comparisons between the WISC-III and WAIS-III Subtest Errors

For both the WISC-III and the WAIS-III, the verbal analysis errors accounted for the highest number of administration errors committed by both groups of examiners on both tests. There was no reduction in these errors from the WISC-III to the WAIS-III for

either group, however, the experimental group committed more verbal analysis errors than the control group for both tests. The practice in learning how to score the verbal subtests of the WISC-III did not assist in decreasing scoring errors on the verbal subtests of the WAIS-III and difficulties with scoring the verbal subtests persisted for both groups, but more so for the experimental group. Results from this study are similar to Slate et al.'s (1991) finding that practice on the WISC-R did not assist in reducing scoring errors on the WAIS-R. There were also a high number of Verbal Analysis errors, particularly for the Comprehension, Vocabulary, and Similarities subtests that seemed to garner most of the incorrect point assignment and questioning errors. These results are consistent with previous results from studies assessing the type of scoring errors on the Wechsler tests (Alfonso, Johnson, Patinella, and Rader, 1998; Belk et al., 2002; Slate and Chick, 1989, Slate & Jones, 1990a, Slate et al., 1992).

Computation errors for both tests accounted for the fewest number of errors committed by both groups of examiners and the differences between the experimental and control groups were slight. However, for both tests, the experimental group committed fewer computation errors on the performance subtests while the control group committed fewer computation errors on the verbal subtests. In contrast, the number of missing data errors for both tests accounted for the highest number of errors committed by both groups of examiners. The Block Design and Digit Span subtests accounted for the highest number of missing data errors for the performance and verbal subtests respectively. The control group committed more missing data errors on the verbal subtests of the WISC-III but fewer missing data errors on the performance subtests of the WISC-III and on all subtests of the WAIS-III.

Summary of the WISC-III Total Errors

The WISC-III total errors consisted of conversion, plotting, and missing data errors with the experimental group committing the majority of these errors. Most of the conversion errors for both groups occurred when consulting and recording information from the tables. The main source of plotting error occurred when examiners from both groups incorrectly plotted Index and IQ scores onto the graphs. Failing to record behavioural observations on the cover page of the record form accounted for the majority of the missing data errors committed by both groups of examiners.

Summary of the WAIS-III Total Errors

The WAIS-III total errors consisted of conversion, plotting, and missing data errors with the experimental group committing the majority of these errors. Most of the conversion errors for both groups were due to calculation errors compared to errors of consulting and recording information from the tables which accounted for fewer of the errors. The main source of plotting error occurred when examiners from both groups incorrectly transferred scores onto the tables. Failing to record demographic information on the cover page of the record form accounted for the majority of the missing data errors committed by both groups of examiners.

Comparisons between the WISC-III and WAIS-III Total Errors

For both the WISC-III and the WAIS-III, the experimental group committed more total errors than the control group. There were more total errors committed by both groups of examiners on the WAIS-III compared to the total number of errors committed on the WISC-III. This result is likely due to the more extensive four page record form that provides space for additional calculations and comparisons of the obtained results to be made. Nonetheless, both groups of examiners followed the same pattern of errors for both tests. For example, missing data errors accounted for the majority of the total errors committed by both groups and the types of conversion and plotting errors were similar between the groups.

Limitations of the Study

The internal validity of the quasi-experimental design used is a concern. This is concern attributable to the lack of random assignment of the students to groups. The lack of test kits required that students share kits. Consequently it was not possible to employ randomization. The groups did not differ in terms of demographic information such as gender composition of the groups, education level, participant-identified position as student or school personnel, previous experience with intelligence or other forms of testing, or enrollment in previous assessment courses. Nonetheless, pre-treatment differences on another unmeasured variable between the two groups of students may have confounded the results that were obtained. It is possible, for example, that the student examiners who identified themselves as school personnel may have received extra knowledge and, perhaps, training about intellectual testing, administration, and scoring

from presenters at a school-based professional development day. Such pre-existing differences between the two groups of student examiners may have existed.

The small number of students who took part in the study was another limitation prevented statistical tests of the differences between groups. Further, the students who participated were from one university, which limits the generalizability of the results.

Two instructors taught the experimental groups and two instructors taught the control groups. These instructors had previous experience with assessment and the Wechsler tests and used the teaching agendas provided to them to ensure consistency and treatment integrity. However, it is likely that the instructors varied in their presentation style, anecdotal discussions, and emphasized in their own ways the material that was taught. Therefore, differential teaching effects may have confounded the results.

A "diffusion" of the treatment may also have occurred. The students knew that they had been divided into two groups. Students from the experimental group may have spoken to students in the control group about the additional information being taught in their group resulting in a lack of treatment integrity. The students in the control group, who committed fewer errors than the students in the experimental group, may have had the benefit of the two teaching interventions without the loss of two practice sessions.

It is also possible that the type and timing of feedback provided to students were variables that could have had an impact on reducing the occurrence of errors on future administrations. The student examiners in the Education Clinic taking the Individual Assessment course are provided written feedback in terms of corrections and edits to the reports and comments on a graded feedback sheet once the entire assessment process is completed and the report is written and submitted for grading. The extent of the oral feedback between supervisor and student is not formalized but left up to the supervisor and student examiner. This written and oral feedback format was not changed for either group during this study. However, some researchers described a formalized procedure of immediate and consistent written and oral feedback as helpful for student examiners to immediately catch errors and correct them rather than having them practice errors (Fantuzzo & Moon, 1985; Slate et al. 1992). The combination of practice administrations and formalized feedback to students may have resulted in fewer scoring and administration errors.

Conclusions

The conclusions drawn from this study are considered with the limitations previously stated. Based on the findings of this study, a decrease in administration and scoring errors on the WISC-III for the experimental group did not occur. Furthermore, there was an increase in these errors on the WAIS-III for both groups with the experimental group committing more of all forms of errors than the control group. The experimental teaching method, as implemented, cannot be viewed as effective in helping students to reduce their administration and scoring errors on either of the tests in light of these findings.

The experimental teaching method developed for this study was based on previous research that suggested that administration and scoring errors would be reduced if 1) a lecture of specific administration and scoring errors frequently identified by researchers as negatively impacting accurate administration and scoring of the Wechsler tests was provided (Fantuzzo et al., 1983: Fantuzzo & Moon, 1984; McQueen et al., 1994; Moon et al., 1986, Slate & Jones, 1989) and 2) the observation of a Wechsler administration was provided (Blakey et al., 1987; Fantuzzo, et al., 1983; Fantuzzo & Moon, 1984; Moon et al., 1986; Slate & Jones, 1989; Slate et al., 1991). While these interventions were described as helpful teaching interventions for reducing administration and scoring errors in other research, they were not effective in reducing the errors in this study. There are a number of possible explanations for this lack of effectiveness. As stated in the limitations, the presence of a differential teaching effect could account for the lack of reduction of errors for the experimental group and collaboration between students in both groups may have confounded the results.

There were also teaching intervention differences between this study and other studies that revealed positive results. Only two teaching interventions were used in the present study, while other studies included additional teaching interventions such as, a) a structured amount of time to study the Wechsler manual prior to other teaching interventions being implemented (Fantuzzo & Moon, 1984), b) formalized feedback (i.e. the CCWA rating scale) that was provided to each student about administration and scoring errors (Moon et al., 1986), c) weekly quizzes (Blakey et al., 1985), d) completion of a written test with a pre-set accuracy level achieved before students could proceed with

the training (Fantuzzo et al., 1983, e) inclusion of peer training (Blakey et al., 1987), and f) discussion about the WISC-III as an instrument of intellectual assessment (Alfonso et al., 1998, & McQueen et al., 1994). A combination of teaching strategies was described as most effective in teaching cognitive assessment and reducing errors (Moon et al., 1986; Slate & Jones, 1989). It could be that the combination of many of these interventions is required for a significant instructional impact and subsequent reduction in administration and scoring errors.

A substantial difference between the experimental and control groups was the inclusion of three practice administrations for the control group and only one practice administration for the experimental group. It was initially thought that replacement of two practice administrations with the two teaching interventions would result in a reduction in administration and scoring errors for student examiners in the experimental group. However, the results of this study reveal that the control group committed fewer overall subtest and total errors, suggesting strongly that the inclusion of practice administrations helped to reduce the number of administration and scoring errors. Therefore, as suggested in the limitations, had the experimental group received more practice in administering these tests rather than less, the results may very likely have been a reduction in the number of administration and scoring errors.

Implications for Practice

This research study was unable to identify the experimental teaching method as more effective in reducing administration and scoring errors for student examiners. Nonetheless, the value of this study can be seen in terms of the in-depth investigation of the type and frequency of administration and scoring errors for the WISC-III and the WAIS-III.

Value of the WISC-III and WAIS-III Checklists

The two checklists developed for this study categorized subtest errors (i.e., administration, computation, verbal analysis, and missing data errors) and total errors (i.e., conversion, plotting, and missing data errors). While other research studies provided information about the types and frequency of such errors, there were oftentimes overlap or generalization of the error types which, for this study, were separated and investigated independently. For example, earlier studies of scoring errors (Miller & Chansky, 1972 ;

Sherrets, Gard, & Langner, 1979) assigned a category of "clerical errors" that included computation, administration, and total errors (as determined for this study). Slate and Chick (1989) described mechanical errors, scoring errors, errors in questioning, errors in determining the basal and /or ceiling, and errors in converting raw scores to standard scores. Conner and Woodall (1983) defined four types of errors as follows: response scoring errors, IQ errors, administrative errors, and mathematical errors. Slate and Jones (1990b), and Slate et al. (1992) investigated errors in terms of scoring errors, errors in questioning, mechanical errors, errors in determining basal and ceilings, and errors in converting raw scores to standard scores. While these studies were informative in providing early indications that such errors did exist and made initial attempts at categorizing them, the method of calculating the errors left the possibility that many errors would be left undetected. Alfonso et al. (1998) used a checklist where the errors were investigated at a more specific level although, their checklist was limited regarding the frequency of error occurrence. For example, the error "Incorrect Point Assignment" was listed as one possible error type for the Information subtest; however, for the checklist used in this study, this error type was included for every individual item of the Information subtest to determine how frequently this error occurred. It could be said that the progression of error analysis from previous studies was extended into the present study with a more detailed accounting of such errors.

Blakey et al. (1987) more recently developed their *Criteria for Competent WAIS-R Scoring (CCWS)* used to rated students' adherence to administration and scoring guidelines as outlined by the WAIS-R manual (1981). While this checklist would have been useful, the student examiners in this study were not observed as they administered the tests; rather, the student protocols were checked for errors. Therefore, some of the same administration errors may have been addressed with both checklists, although some of the administration requirements would have required direct observation (e.g., the development of rapport between examinee and examiner and the correct placement of stimulus materials). The checklists used in this study are unique in terms of identifying more specifically the type and frequency of administration and scoring errors according to the error types.

Post-Supervisory Errors

An unexpected finding of this research study was the number of student errors that remained after the students' supervisors had ostensibly reviewed and corrected their reports. Other researchers (Franklin et al., 1982; Miller & Chansky, 1972; Moon et al., 1991; Ryan et al., 1983; Sherrets et al. 1979; Slate & Jones, 1993) observed that professional psychologists who have completed training in how to administer and score the Wechsler tests nonetheless continue to make these errors. In each study, administration and scoring errors, including calculation and incorrect scoring of verbal and performance subtests, were noted. Student supervisors could use the administration and scoring checklists from this study as they check the student examiner's record forms for errors. The checklists might be a way to ensure that both the supervisor and the student are aware of the scoring errors and that they are checked in a formalized manner using the checklist.

Recommendations for Future Research

Based on the results of this study, and in consideration of the results from other studies on reducing administration and scoring errors and increasing effective teaching of intelligence testing, the following teaching components are suggested for investigation in future research studies:

- Pre-study of the Wechsler Administration and Scoring Manual and Technical Manual prior to any administration of the tests and/or a 3 credit course devoted to learning about intellectual assessment
- Practice administrations
- Formalized written and oral feedback from supervisors to student examiners and 40% of the mark devoted to correct administration and scoring of the Wechsler tests
- Peer collaboration and mentoring
- Specific lecture focused on areas of administration and scoring weakness and the more troublesome subtests to score
- Observation of a live Wechsler administration with follow up of administration and scoring errors
- Quizzes used to assess student understanding of the information being taught

- Final exam with a specified accuracy rate percentage needed to pass the course

The pre-study of Wechsler information appears to have made some contribution to increasing student learning. This pre-study could allow student examiners the opportunity to consider the theoretical bases and knowledge about the tests prior to beginning to learn how to administer and score them. Student examiners may need more than one or two hours of pre-study regarding the Wechsler intelligence tests and instead require a more extensive learning experience about intelligence testing in general. A three credit course that addresses the many issues related to intelligence testing may be warranted. There may be a need for students to have the opportunity to learn more about intelligence testing, the rationale behind such testing, the negative and positive contributions of intelligence testing, in addition to a forum where they can voice their opinions and offer new ideas for intelligence testing. The focus of the Individual Assessment course could then be more specifically devoted to learning the correct way to administer and score such tests.

Inclusion of practice administrations has already been addressed and determined as a necessary teaching component to help students learn to accurately administer the Wechsler tests. Formalized written and oral feedback from supervisors could be implemented through the use of the scoring and administration checklist used in this study, in addition to the feedback already in place (i.e., the graded feedback sheet that students currently receive). It is also suggested that 40% of the final mark for each assessment report be devoted to correct administration and scoring of the Wechsler test. The peer collaboration and mentoring idea was deemed a useful process to decrease errors as was the specific lecture focused on areas of administration and scoring weakness and the more troublesome subtests to score. Observation of a live Wechsler administration with follow up regarding the administration and scoring errors, was also listed as an effective teaching intervention. The quizzes and final exam are two additional ways to establish an educational outcome that can be measured in terms of an accuracy level that would need to be obtained by the student examiners.

Characteristics of the Student Examiner

Certain traits and characteristics of the student examiner may be associated with a better ability to learn to administer, score, and complete assessment reports using the WISC-III and/or the WAIS-III. For example, the importance of intelligence and common sense appear to play a role in learning the substantial amount of information involved in the assessment of intelligence. From the gifted literature, McCluskey, Treffinger, and Baker (1998) proposed a number of useful elements in creating a classroom environment where all students are able to develop their strengths and potential to the highest level. One of their criteria was for students to develop metacognition or "thinking about thinking" (Armbruster & Brown, 1984). By virtue of their status as graduate students, the student examiners in this study have presumably achieved a level of higher order thinking and have developed their metacognitive abilities. However, this is perhaps not the case and some sort of entrance criteria for taking part in the Individual Assessment Course should be required. For example, achieving a passing mark on the proposed three credit pre-requisite course on theories and issues of intelligence testing could be one method of ensuring a set level of competence.

The ability to think in a creative manner was described as a unique trait of students in the area of assessment and field of diagnostic work (Woody & Robertson, 1988). McCluskey et al (1998) also described the importance of productive and creative thinking and the ability of the students to become aware of their own learning style, strengths and weaknesses. Nietzel et al. (1991) had described a need for student examiners in the area of assessment to be capable of undergoing a process of self-evaluation and be open to critical self-examination. This ability to self-evaluate, if addressed in future research, could help student examiners to learn more about their own thinking processes and biases that might interfere with their ability to perform accurate and reliable assessment work.

Future studies investigating the number of errors on the new editions of the WISC or WAIS should, based on the results of this study, take into account the need for a more structured teaching approach with tighter control of possible extraneous variables, such as the random assignment of students to experimental and control groups and the separate teaching of the WISC-III and the WAIS-III to each group. In this way, the groups are not aware of each other and the treatment integrity can be enhanced. Associated extraneous

variables that could be addressed are the use of only two instructors, one for the experimental group and one for the control group without any change in instructor for the duration of the study, and a formalized written and oral feedback format between supervisor and student examiner. The possible importance of immediate and consistent written and oral feedback regarding student WISC-III and WAIS-III administrations could be investigated as part of future research on reducing student examiner administration and scoring errors.

An additional suggestion for future research is the inclusion of student examiner pre-test and post-test data. This data could take the form of pre and post test administrations as a way to quantify the level of student examiner progress, or lack of progress, in administrating and scoring accuracy rather than relying solely on the number of administration and scoring errors. This quantification of student proficiency in administration and scoring accuracy could help to determine student progress and assess the effectiveness, or lack thereof, from a teaching perspective.

In regard to the high number of errors on both tests, future research could address if calculation errors are reduced through computer assisted scoring and if verbal analysis errors are reduced with the increased clarification of scoring procedures on the part of the test developers. The Matrix Reasoning subtest, a new performance subtest in the WAIS-III, was reported to cause student examiner's difficulty in terms of administration errors and should be the focus of future research in order to address the need of decreasing the numbers of these errors.

The high numbers of scoring and administration errors that were found in this study are intolerable. These errors need to be addressed both at the student training level and at the professional level to ensure that student examiners do not continue to practice errors when they achieve professional status. Supervisors were discovered to make errors, both of omission and commission, when correcting student protocols. The commission of student examiner and supervisor errors pose considerable ethical concerns. Examinee's could be given false information based on these errors with potentially far reaching consequences. Such consequences might include, for example, erroneous labeling of a student who has undergone an intellectual assessment or assigning an incorrect diagnostic classification that could result in inappropriate patient treatment and care. During this

study, two protocols were discovered where the scoring errors were such that the two examinees was erroneously described as intellectually lower than they actually should have been if the scoring errors had been rectified. Both students and professionals conducting intellectual assessments must adhere to the ethical principles and code of conduct as outlined by the American Psychological Association (APA, 1992) and from the College of Alberta Psychologists' Code of Conduct (1997). These principles state that students and professionals must assume responsibility for maintaining high standards of conduct and competency. Lack of adherence to these codes and principles results in a lack of competency leading to unfair assessment practices. Legal implications could follow including prosecution against individual professionals who make these errors or condemnation of the profession as a whole based on the unacceptable number of errors. One conclusion of this finding is that the supervisors who assume much of the responsibility for student examiner's correct intellectual assessment administration and scoring, be given standard training in the administration and correction of the protocols. The use of the administration and scoring checklists developed for this study could help to identify these errors and be a tool used to help train the supervisors. Future research could address additional content and format of this standard supervisor training.

References

Aiken, L. R. (1996). Assessment of intellectual functioning (2nd Ed.) New York: Plenum.

Alfonso, V. C., Johnson, A., & Patinella, L., and Rader, D. E. (1998). Common WISC-III examiner errors: Evidence from graduate students in training. Psychology in the Schools, 35 (2), 119-125.

Alfonso, V. C., & Pratt, S. I. (1997). Issues and suggestions for training professionals in assessing intelligence. In Flanagan, D. P., Genshaft, J. L. & Harrison, P. L. (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues.* (pp. 326-347). New York: Guilford Press.

American Educational Research Association (1999). Standard for Educational and Psychological Testing.

American Psychological Association (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.

Armbruster, B. B., & Brown, Al. L. (1984). Learning from reading: The role of metcognition. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), Learning to read in American schools (p 273-281). Hillsdale, NJ: Erlbaum.

Armour-Thomas, E. & Gopaul-McNicol, S. (1998). Assessing intelligence: Applying a bio-cultural model. Thousand Oaks: SAGE Publications.

Back, R. , & Dana, R. H. (1977). Examiner sex bias and Wechsler intelligence scale for children scores. Journal of Consulting and Clinical Psychology, 45, (3), 500.

Barolow, D. H., Hayes, S. C., & Nelson, R. O. (1984). The Scientist Practitioner: Research and Accoutability in Clinical and Educational Settings. Pergamon Press: New York.

Barron, J. W. (Ed.) (1998). Making diagnosis meaningful: Enhancing evaluation and treatment of psychological disorders. American Psychological Assocation: Washington, D. C.

Beasley, M. G., Lobasher, M., Henley, S., & Smith, I. (1988). Errors in computation of WISC and WISC-R intelligence quotients from raw scores. Journal of Child Psychology, Psychiatry and allied Disciplines, 29 (1), 101-104.

Belk, M. S., LoBello, S. G., Ray, G. E., & Zachar, P. (2002). WISC-III administration, clerical, and scoring errors made by student examiners. Journal of Psychoeducational Assessment, 20, 290-300.

Binet, A., & Simon, T. (1916).The development of intelligence in children (E. S. Kit, trans.). Baltimore: Williams & Wilkins.

Blakey, W. A., Fantuzzo, J. W., & Gorsuch, R. L. (1986). Teaching WAIS-R administration skills: Comparison of the Mastery model to other existing clinical training modalities. Professional Psychology: Research and Practice 17, (1), 31-35.

Blakey, W. A., Fantuzzo, J. W., Gorsuch, R. L., & Moon, G. W. (1987). A peer-mediated, competency-based training package for administering and scoring the WAIS-R. Professional Psychology: Research and Practice, 18, 17-20.

Blakey, W. A., Fantuzzo, J. W., Moon, G. W. (1985). An automated competency-based model for teaching skills in the administration of the WAIS-R. Professional Psychology: Research and Practice 16, (5), 641-647.

Brabender, V. (1991, November). Training competent assessors: Admnistrative issues. SPA Exchange, 1 (2), 1-2.

Bradbury, P. J., Wright, S. D., Walker, C. E., & Ross, J. M. (1975). Performance on the WISC as a function of sex of E, sex of S, and age of S. Journal of Psychology, 90, 51-55.

Bradley, F., Hanna, G., & Lucas, M. (1980). The reliability of scoring the WISC-R. Journal of Consulting and Clinical Psychology, 48, 530-531.

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

Camara, W. J., Nathan, J. S., Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. Professional Psychology: Research and practice, 31, 141-154.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin.

Clark, L. A., & Watson, D. (1998). Constructing Validity: Basic Issues in objective scale development. In Kazdin, A. E. (Ed). Methodological Issues and Strategies in Clinical Research (2nd Ed.). American Psychological Association: Washington, D. C.

College of Alberta Psychologists code of conduct (1997).

Cormay, R. B. (1973). Faculty attitudes toward standardized testing in the public schools of Pennsylvania. Unpublished doctoral dissertation, Pennsylvania State University, Pennsylvania.

Cormay, R. B. (1974). Faculty attitudes toward standardized testing. Measurement and Evaluation in Guidance, 7 (3), 188-194.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach, L. J. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Cullari, S. (Ed.) (1998). Foundations of Clinical Psychology. Allyn and Bacon: Boston

Dana, R. H., & May, W. T. (1987). Internship Training in Professional Psychology. Hemisphere Publishing Corp.: Cambridge.

Davis, R. J. The effects of racial apperception, race of examiner, race of subject, and stress-inducing instructions of IQ score and reported state anxiety. (Doctoral Dissertation, 1987). Dissertation Abstracts International, 48, (9), 2819-B.

Drabick, D. A. G, Y Goldried, M. R. (2000). Training the scientist-practitioner for the 21st century: Putting the bloom back on the rose. Journal of Clinical psychology, 56 (3), 327-340.

Elbert, J. C. (1984). Training in child diagnostic assessment: A survey of clinical psychology graduate programs, Journal of Clinical Child Psychology, 13, 122-133.

Fantuzzo, J. W., Sisemore, T. A.,& Spradlin, W. H. (1983). A competency-based model for teaching skills in the administration of intelligence tests. Professional Psychology: Research and Practice, 14, 224-231.

Fantuzzo, J. W., & Moon, G. W. (1984). Competency mandate: A model for teaching skills in the administration of the WAIS-R. Journal of Clinical Psychology, 40, 1053-1059.

Feldman, S. E., & Sullivan, D. S. (1971). Factors mediating the effects of enhanced rapport on children's performance. Journal of Consulting and Clinical Psychology, 36, 302.

Flanagan, D. P., Genshaft, J. L., & Harrison, P. L (Eds.), Contemporary intellectual assessment. New York : Guidford Press.

Franklin, M. R., Stillman, P. L., Young Burpeau, M., & Sabers, D. L. (1982). Examiner error in intelligence testing: Are you a source? Psychology in the Schools, 19, 563-569.

Garb, H. n. (2000). Studying the Clinician: Judgment research and psychological assessment. Journal of Clinical Psychology, 56 349-365.

Giordano, F. G., Schwiebert, V. L., & Brotherton, W. D. (1997). School counselors' perceptions of the usefulnes of standardized tests, frequency of their use, and assessment training needs. The School Counselor, 44, 198-205.

Goodenough, F. L. (1940). Some special problems of nature-nurture research. Yearbook of the National Society for the Study of Education, 39 (1), 367-384.

Green, F. The examiner as a possible source of constant error in intelligence testing. In R. L., Cromwell (Ed). (1960-62). Abstracts of Peabody Studies in Mental Retardation (2). Nashville: George Peabody College for Teachers.

Gregory, R. J. (1999). Foundations of Intellectual Assessment: The WAIS-III and Other Tests in Clinical Practice. Needham Heights, MA: Allyn and Bacon.

Gutkin, T. B., & Reynolds, C. R. (Eds.) (1990). The Handbook of School Psychology. (2nd Ed). John Wiley & Sons: New York.

Harrison, P.L., Kaufman, A.S., Hickman, J.A., & Kaufman, N. L. (1988). A survey of tests used for adult assessment. Journal of Psychoeducational Assessment, 6, 188-198.

Irons, D. (1981). The effect of familiarity with the examiner on WISC-R verbal, performance, and full scale scores. Psychology in the Schools, 18, 496-499.

Jaffee, L. S. (1983). A study of various factors influencing examiner scoring reliability on WAIS-R vocabulary, comprehension, and similarities subtests. Unpublished doctoral dissertation, California School of Professional Psychology, San Diego.

Jastak, J. F., & Jastak, S. R. (1964). Short forms of the WAIS and WISC vocabulary subtests. Journal of Clinical Psychology, 20, 167-199.

Kamphaus, R. W. (2001). Clinical Assessment of Child and Adolescent Intelligence. (2nd Ed). Allyn & Bacon: Boston.

Kaufman, A. S. (1983). Test review: WAIS-R. Journal of Psychoeducational Assessment, 1,309-319.

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

Kaufman, A. S., Kamphaus, R. W., & Kaufman, N. L. (1985). The Kaufman Assessment Battery for Children (K-BIT). In C. S. Newmark (Ed). Major Psychological Assessment Instruments. Boston: Allyn and Bacon.

Kaufman, A. S., & Lichtenberger (Eds.) (2002). Assessing Adolescent and Adult Intelligence. (2nd Ed). Allyn and Bacon: Boston

Kaufman, A. S. & Wang, J. (1990). Kaufman Brief Inteligence Test Manual. Circle Pines: MN: American Guidance Service.

Kazdin, A. E. (Ed). (1998). Methodological Issues and Strategies in Clinical Research (2nd Ed). American Psychological Association: Washington, D. C.

Kazdin, A. E. (1998). Research Design in clinical psychology. (3rd. Ed.). Allyn and Bacon: Boston.

Kinnie, E. J., & Sternlof, R. E. (1971). The influence of nonintellective factors on the IQ scores of middle and lower class children. Child Development, 42, 89-95.

Klassen, R. M., & Kishor, N. (1996). A comparative analysis of practitioners' errors on WISC-R and WISC-III. Canadian Journal of School Psychology, 12, (1), 35-43.

Kolbe, K., Shemberg, K., & Leventhal, D. (1985). University training in psychodiagnoistics and psychotherapy. The Clinical Psychologist, 38, 59-16.

Levenson, R. L., Golden-Scaduto, C. J., Aiosa-Karpas, C. J., & Ward, A. W. (1988). Effect of examiners' education and sex on presence and type of clerical errors made on WISC-R protocols. Psychological Reports, 62, 659-664.

Levitt, E. (1973). Internship versus campus: 1964 and 1971. Professional Psychology, 4, 129-132.

Littell, W. M. (1960). The Wechsler Intelligence Scale for Children: Review of a decade of research. Psychological Bulletin, 57, 1332-156.

LoBello, S. G., & Holley, G. (1999). WPPSI-R Administration, clerical, and scoring errors by student examiners. Journal of Psychoeducational Assessment, 17, 15-23.

Masling, J. (1957). The effects of warm and cold interaction of the interpretation of a projective protocol. Journal of Projective Technique, 21, 377-383.

Massey, J. O. (1965). WISC Scoring Criteria. Palo Alto: Consulting Psychologists Press.

Massey, J. O., Sattler, J. M., & Andres, J. (1978). WISC-R Scoring Criteria. Palo Alto: Consulting Psychologists Press.

Masling, J. (1959). The effects of warm and cold interaction on the administration and scoring of an intelligence test. Journal of Consulting Psychology, 23 (4), 336-341.

McCluskey, E. W., Treffinger, D. J, & Baker, Philip. The Amphitheater model: An approach to talent recognition and development (p. 7-17), chapter from: Teaching for talent development: Current and expanding persepectives(1998). Sarasota, FL: Center for Creative learning.

McQueen, W. Meschino, R., Pike, P, & Poelstra, P. (1994). Improving graduate student performance in cognitive assessment: The saga continues Professional Psychology: Research and Practice, 25, 283-287.

Meyer, G. J., Finn, S. E., Eyde, L.D., Kay, G. G. Kubiszyn, T. W., Moreland, K. L., Eisman, E. J., & Dies, R. R. (1998). Benefits and costs of psychological assessment in healthcare deliver: Report of the Board of Professional Affairs Psycholgical Assessment Work Group, Part I. Washington, D. C.: American Psychological Association.

Moon, G. W., Fantuzzo, J. W., & Gorsuch, R. L. (1986). Teaching WAIS-R administration skills: Comparison of the MASTERY model to other existing clinical training modalities. Professional Psychology Research and Practice, 17 (1), 31-35.

Moon, G. W., Blakey, W. A., Gorsuch, R. L, Fantuzzo, J. W. (1991). Frequent WAIS-R administration errors: An ignored source of inaccurate measurement. Professional Psychology: Research and Practice, 22 (3), 256-258.

Nietzel, M. T., Bernstein, D. A., & Milich, R. (1991). Introduction to Clinical Psychology (3rd Ed). Prentice Hall.

O,Donohue, W., Plaud, J. J., Mowatt, A. M., & Fearon, J. R. (1989). Current status of curricula of doctoral training programs in clinical psychology. Professional Psychology: Research and Practice, 20, 196-197.

Olin, J. T., & Keatinge, C. (1998). Rapid Psychological Assessment. John Wiley & Sons: New York.

Peterson, R. L., McHolland, J. D., Bent, R. J., Davis-Russell, E., Edwall, G. E., Polite, K., Singer, D. L., & Stricker, G. (1991). The Core Curriculum in Professional Psychology. Washington, DC: American Psychological Association.

Pfeiffer, S. I., Burd, S., & Wright, A. (1992). Clinicians and research: Recurring obstacles and some possible solutions. Journal of Clinical Psychology, 48, 140-145.

Piotrowski, C., & Keller, J. W. (1984). Psychodiagnostic testing in APA-approved clinical psychology programs. Professional Psychology: Research and Practice, 15, 450-456.

Piotrowski, C., & Zalewski, C. (1989). Psychological testing in outpatient mental health facilities: A national survey of course instructors. Professional School Psychology, 1, 51-59.

Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved PsyD and PhD clinical psychology programs. Journal of Personality Assessment, 61 (2), 394-405.

Prifitera, A., & Saklofske, D. (Eds.). (1998). WISC-III Clinical Use and Interpretation: Scientist-Practitioner Perspectives. Academic Press: San Diego.

Principles for Fair Student Assessment Practice for Education in Canada (1993).

The Psychological Coropration. (1992). Wechsler Individual Achievement Test. San Antonio, TX: Author.

Rapee, R. M., & Barlow, D. H. (Eds.) (1991). Chronic anxiety: Generalized anxiety disorder and mixed-anxiety depression. New York: Guilford.

Raulin, M, L., & Graziano, A. M. (1995). Quasi-experiments and correlational studies. In, Colman, A., M. (Ed). Psychological Methods and Statistics. London: Longman.

Reynolds, C. R., & Kaufman, A. S. (1990). Assessment of Children's intelligence with the Wechsler Intelligence Scale for Children-Revised (WISC-R). In C. R. Reynolds, & R. W. Kamphaus (Eds.) Handbook of Psychological and Educational Assessment of Children: Intelligence and Achievement. New York: Guilford.

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

Reynolds, C. R., & Kamphaus, R. W. (Eds.). (1990). *Handbook of Psychological & Educational Assessment of Children: Personality, Behavioral, & Context.* New York: Guilford.

Sattler, J. M. (2001). Assessment of Children. (4th Ed.) San Diego: Jerome M. Sattler.

Sattler, J. M. (1992). Assessment of Children. (Revised and updated 3rd Ed.). San Diego: Jerome M. Sattler, Publisher.

Sattler, J. M. (1973). Examiners' scoring style, accuracy, ability and personality scores. Journal of Clinical Psychology, 29, 38-39.

Sattler, J. M., & Theye, F. (1967). Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 68 (5), 347-360.

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

Shaffer, D., Lucas, C. P., & Richters, J. E. (Eds.) (1999). Diagnostic Assessment in Child and Adolescent Psychopathology. The Guilford Press: New York.

Shakow, D. (1947). Recommended graduate training program in clinical psychology. American Psychologist, 2 539-558.

Shannon, R. L. (1984). Impact of special education labels: Implications for psychological re-evaluations utilizing the WISC-R. [CD-ROM]. Abstract from: ProQuest File: Dissertation Abstracts Item: 8426339.

Simeonsson, R. J. (1986). Psychological and Developmental Assessment of Special Children. Allyn & Bacon: Boston.

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

Slate, J. R., & Jones, C. H. (1989). Can teaching of the WISC-R be improved? Quasi-experimental exploration. Professional Psychology: Research and Practice: 20 (6), 408-410.

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

Slate, J. R., & Jones, C. H. (1990b). Identifying students' errors in administering the WAIS-R. Psychology in the Schools, 27, 83-87.

Slate, J. R., & Jones, C. H. (1990c). Examiner errors on the WAIS-R: A source of concern. The Journal of Psychology, 124 (3), 343-345.

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

Slate, J. R., Jones, C. H., Covert, T. L. (1992). Rethinking instructional design for teaching the WISC-R: The effects of practice administrations. College Student Journal, 26 (6), 285-289.

Slate, J. R., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale-Revised: An empirical evaluation of practice administrations. Professional Psychology: Research and Practice, 22 (5), 375-379.

Slate, J. R., Jones, C. H., Murray, R. A., Coulter, C. (1993). Evidence that practitioners err in administering and scoring the WAIS-R. Measurement and Evaluation in Counseling and Development, 25, 156-161.
Slater, B. R., & Thomas, J. M. (1983). Psychodiagnostic Evaluation of Children: A Casebook Approach. Teachers College Press: New York.

Smith, D, V. C. The effects of race of subject, race of examiner, and gender of examiner on anxiety and intellectual test performance. (Doctoral Dissertation, 1992). Dissertation Abstracts International, 53, (12), 6610-B.

Spielberger, C. D. (1983a). State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (1983b). State-trait anxiety inventory: A comprehensive bibliography. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (1992). New horizons for personality assessment. SPA Exchange, 2 (1), 6-7.

Terrell, R., & Terrell, S. L. (1983). The relationship between race or examiner, cultural mistrust, and the intelligence test performance of black children. Psychology in the Schools, 20, 367-369.

Thorndike, R. L., Hagen, R. P., & Sattler, J. M. (1986). Guide for administering and scoring , the Standford-Binet Intelligence Scale: Fourth Edition. Chicago: Riverside Publishing.

Thorndike, R. L. (1997). The early history of intelligence testing. In Flanagan, Genshaft & Harrison (Eds.) Contemporary intellectual assessment. New York: Guilford Press.

Tulsky, D., Zhu, J., & Ledbetter, M. (1997). WAIS-III and WMS-III technical manual. San Antonio, TX: The Psychological Corporation.

Wagoner, R. (1988). Socring errors made by practicing psycholgists on the WISC-R. Unpublised master's theseis, Western Carolina University.

Walker, R. E., Hunt, W. A., & Schwartz, M. L. (1965). The difficulty of WAIS comprehension scoring. Journal of Clinical Psychology, 21, 427-429.

Watkins, C. E. (1990). What have surveys taught us about the teaching and practice of psychological assessment? Journal of Personality Assessment, 56, 426-437.

Warren, S. A., & Brown, W. J. Jr. (1973). Examiner scoring errors on individual intelligence tests. Psychology in the Schools, 9 118-122.

Wechsler, D. (1939). The measurement of adult intelligence. Baltimore: Willams & Wilkins.

Wechsler, D. (1939). Wechsler-Bellevue Intelligence Scale. New York: The Psychological Corporation.

Wechsler, D. (1949) Wechsler intelligence scale for children. San Antonia, TX: The Psychological Corporation.

Wechsler, D. (1955) Wechsler adult intelligence scale: Manual. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1958). The Mesaurement and Appraisal of Adult Intelligence (4th Edition). Baltimore: Williams & Wilkins.

Wechsler, D. (1967).Manual for the Wechsler Preschool and Primary Scale of Intelligence. San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (1987). Wechsler Memory Scale-Revised. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1967).Manual for the Wechsler Preschool and Primary Scale of Intelligence. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1989).Manual for the Wechsler Preschool and Primary Scale of Intelligence- Revised. San Antonio, TX: The Psychological Corporation.

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

Wechsler, D. (1997). WAIS-III Administration and Scoring Manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). WAIS-III, WMS-III Technical Manual. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). The Wechsler Adult Intelligence Scale-Third edition. San Antonio, TX: Psychological Corporation.

Wheeler, P. T. A study of the effect of a child's physical attractiveness upon verbal scoring of the Wechsler Intelligence Scale for Children (Revised) and upon personality attributions. (Doctoral Dissertation, Utah State University, 1985). Dissertation Abstracts International, 47, (8), 3550-B.

Woody, R. H., & Robertson, M. (1988). Becoming a Clinical Psychologist. International Universities Press: Madison, CN.

Yerkes, R. M. (1917). The Binet versus the point scale method of measuring intelligence. Journal of Applied Psychology, 1, 111-122.

Yerkes, R. M., & Dobson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 18, 459-482.

Zavin, J. D. The effects of prior suggestions of pathology on the scoring and interpreting of psychological tests. (Doctoral Dissertation, New School for Social Research, 1988) Dissertation Abstracts International, 49, (10), 4566-B.

Zimbardo, P. G. (1985). Psychology and Life- Golden Edition (12th Ed). Scott Foresman and Company: Glenview: Ill.

Appendix A
Ethics Sheet

Ethics Sheet

An information sheet is being handed out to you. On this sheet is a list of six questions about you and your experience with testing and assessment. These questions include your education level (i.e., Bachelor of Arts, Master's level or Ph. D. level student), position (i.e., student, teacher, school administrator), previous experience with intelligence testing, previous experience with other forms of testing/assessment, and previous testing/assessment courses. I have gone through the necessary ethical steps to allow me to ask for this information from you and there are no known risks in requesting this information although, this information will be used to better understand the background of the students taking this course. All information collected will be kept confidential.

You do not have to participate in providing this information. You have a choice of whether you want to participate or not. If you choose to not participate, please leave the sheet blank. Your participation or lack of participation in providing this information will not affect your grade in this course.

If you have any questions about the information being requested today, please ask questions now or feel free to contact Yvonne Legris at the Education Clinic or you can reach me at home at 432-0808 (write name and number of the board).

Thank you for your time.

January 21, 1999

From: Department of Educational Psychology
Research and Ethics Committee

The Research and Ethics Committee of the Department of Educational Psychology has reviewed the attached proposal and finds it acceptable with respect to ethical matters.

Applicant: Dr. H. Janzen on behalf of Yvonne Legris (graduate student)

Title: Scoring errors in intelligence tests: The attitudes of student examiners toward intelligence testing, the examiner's anxiety levels, and the effects of teaching methods.

Participating Agency(ies):

Approval Conditions:

Teachers of the strategies will remain unaware of which students opt in or out as participants.

Chairman or Designate, Research and Ethics Committee

21 Jan 99
Date

Appendix B
Teaching Agendas

Winter Session 1998-99 General Teaching Agenda- WISC-III

The time frame to teach the WISC-III during the Winter Session is six days from October 26 to November 16. Total instructional hours are nine hours.

1- Introductory Lecture- October 26 - All students attend this class
All students take part in this class. The introductory lecture for the WISC-III includes an overview of the WISC-III with overheads. See attached handouts and overheads.

2- Break into experimental and control groups – October 28
Control Group: Practice administration
Experimental Group: Practice administration
Both groups practice the administration of the WISC-III.

3- Break into experimental and control groups– November 2-
Control Group: General lecture on administration and scoring of the WISC-III.
Experimental Group: General lecture on administration and scoring of the WISC-III.

4- Break into experimental and control groups– November 4
Control Group: Students practice administering the test to each other.
Experimental Group: Specific Lecture of administration and scoring errors: Specific lecture on information about administration and scoring errors. See teacher notes and overheads.

5- Break into experimental and control groups – November 9
Control Group: Students practice administering the test to each other.
Experimental Group: WISC-III Administration A standardized administration of the WISC-III takes place. Discussion follows the WISC-III administration. Students' results are compared to the standardized protocol.

6- Interpretation Lecture – November 16 - All students attend this class
An interpretation lecture for the WISC-III class involves all students. General interpretation methods are explained and students are asked to bring in their protocols to be interpreted and discussed. See Interpretation Lecture notes and overheads.

Winter Session 1998-99 General Teaching Agenda- WAIS-III

The time frame to teach the WAIS-III during the Winter Session is six days from January 6 to January 25. Total instructional hours are nine hours.

1- Introductory Lecture- January 6 - All students attend this class
All students take part in this class. The introductory lecture for the WAIS-III includes an overview of the WAIS-III with overheads. See attached handouts and overheads.

2- Break into experimental and control groups –January 11
Control Group: Practice administration
Experimental Group: Practice administration
Both groups practice the administration of the WAIS-III.

3- Break into experimental and control groups–January 13-
Control Group: General lecture on administration and scoring of the WAIS-III.
Experimental Group: General lecture on administration and scoring of the WAIS-III.

4- Break into experimental and control groups–January 18
Control Group: Students practice administering the test to each other.
Experimental Group: Specific Lecture of administration and scoring errors: Specific lecture on information about administration and scoring errors. See teacher notes and overheads.

5- Break into experimental and control groups – January 20
Control Group: Students practice administering the test to each other.
Experimental Group: WAIS-III Administration A standardized administration of the WAIS-III takes place. Discussion follows the WAIS-III administration. Students' results are compared to the standardized protocol.

6- Interpretation Lecture –January 25- All students attend this class
An interpretation lecture for the WAIS-III class involves all students. General interpretation methods are explained and students are asked to bring in their protocols to be interpreted and discussed. See Interpretation Lecture notes and overheads.

Spring/Summer 1999 General Teaching Agenda for the WAIS-III

The time frame to teach the WAIS-III during Spring/Summer Session is 1 full day on Saturday May 29 and one evening on June 2. Total instructional hours are 8.5 hours.

SATURDAY MAY 29

9:00 a.m.- 10:20 a.m. **Introduction to the WAIS-III** (both groups involved)
10:20 a.m.- 10:30 a.m BREAK
10:45 a.m.- 12:00 p.m **Break into Experimental and Control Groups**
Control Group: Practice administration of the WAIS-III
Experimental Group: Practice administration of the WAIS-III
12:00 p.m.- 12:30 p.m. LUNCH
12:30 p.m.- 2:15 p.m. **Break into Experimental and Control Groups**
Control Group: General lecture on scoring and administration of the WAIS-III
Experimental Group: General lecture on scoring and administration of the WAIS-III
2:15 p.m. - 2:30 p.m. BREAK
2:30 p.m.- 4:00 p.m. **Break into Experimental and Control Groups**
Control Group: Practice administration
Experimental Group: Specific lecture of administration and scoring errors

WEDNESDAY JUNE 2

6:00 p.m. - 7:30 p.m. **Break into Experimental and Control Groups**
Control Group: Practice administration
Experimental Group: WAIS-III administration: students observe and score a WAIS-III administration followed by discussion.
7:30 p.m.- 7:45 p.m. BREAK
7:45 p.m.- 9:00 p.m. **Interpretation of the WAIS-III** (Both groups involved)

Spring/Summer 1999 General Teaching Agenda for the WISC-III

The time frame to teach the WISC-III during Spring/Summer Session is 1 full day on Saturday May 15 and one evening on May 19. Total instructional hours are 8.5 hours.

SATURDAY MAY 15

9:00 a.m.- 10:20 a.m. **Introduction to the WISC-III** (both groups involved)

10:20 a.m.- 10:30 a.m BREAK

10:45 a.m.- 12:00 p.m **Break into Experimental and Control Groups**

Control Group: Practice administration of the WISC-III

Experimental Group: Practice administration of the WISC-III

12:00 p.m.- 12:30 p.m. LUNCH

12:30 p.m.- 2:15 p.m. **Break into Experimental and Control Groups**

Control Group: General lecture on scoring and administration of the WISC-III

Experimental Group: General lecture on scoring and administration of the WISC-III

2:15 p.m. - 2:30 p.m. BREAK

2:30 p.m.- 4:00 p.m. **Break into Experimental and Control Groups**

Control Group: Practice administration

Experimental Group: Specific lecture of administration and scoring errors

WEDNESDAY MAY 19

6:00 p.m. - 7:30 p.m. **Break into Experimental and Control Groups**

Control Group: Practice administration

Experimental Group: WISC-III administration: students observe and score a WISC-III administration followed by discussion.

7:30 p.m.- 7:45 p.m. BREAK

7:45 p.m.- 9:00 p.m. **Interpretation of the WISC-III** (Both groups involved)

Introductory Lecture: WISC-III
Brief History of the WISC-III

Development of the WISC

- The *Wechsler Bellvue* was the first Wechsler IQ test published in 1946
- The *WISC-R* was published in 1974 – a downward extension of the WAIS
- The *WISC-III* was published in 1991, 17 years after the WISC-R (1974). "The most widely used assessment of intellectual functioning of children".

Why revise the test?

NORMS

The main reason for revising the test was to update the norms.
A **norm** provides an indication of average or typical performance of the specified group. Norms are needed because a raw test score in itself is not very meaningful. A comparison is made between an individual child's raw score to another child's raw score by converting the scores into some relative measure.

In **norm-referenced testing**, an examinee's performance is compared to a specific group of subjects.
Developing current norms gives more precise scores for individuals and ensures that the **"Flynn Effect"** (where scores become inflated over time and norms need to be reestablished) does not occur.

Changes to the Test

- Changes in test items and materials make them more contemporary and attractive to examinees.
- Test items are reviewed for bias and replaced or modified to make the test fairer
- Exploration of the factor structure of the test

Standardization

- The WISC-III was standardized on 2,200 American children: 100 boys and girls in each of 11 age groups from 6 though 16 years.
- The standardization process was superior to that of the WISC-R because of its extensiveness. The sample, based on 1988 census data was stratified according to :
- age
- race/ethnicity (white, black, Hispanic, Indian, Asian, Eskimo, other)
- geographic location
- parent education

Other Versions of the WISC-III

There are 11 translations of the WISC-III ranging from Japanese to Greek and French.

UK version- 1992
Based on a sample of 824 children: comments regarding these other European versions, "The majority of items in the final US selection work throughout Europe". Some artwork changes and minor scoring changes made to reflect the specific UK setting.

Australian version- 1995
Based on a sample of 468 children : comments regarding this version suggested that the presentation order of some of the items should be modified, but that there was insufficient evidence to suggest the need to develop a full set of Australian norms.

Canadian version- 1996
Based on a representative sample of English-speaking Canadian children. When differences that were larger than could be accounted for by measurement error were found, a comprehensive standardization study was initiated that resulted in the publication of Canadian Norms. The results showed that Canadian children scored: 3.34 IQ points above the US normative sample differences ranged from 1.03 points for the FD Index to 4.96 points for the PIQ in favor of the Canadian children. Higher scores were earned on all subtests except for the Information and Arithmetic subtests. No changes were reportedly needed to the test items.

WISC-III Description

Subtests
13 in total
6 are Verbal subtests – they make up the Verbal IQ
7 are Performance subtests – they make up the Performance IQ

Verbal Subtests	**Performance Subtests**
Vocabulary	Block Design
Similarities	Object Assembly
Information	Picture Completion
Comprehension	Picture Arrangement
Arithmetic	Coding
Digit Span	Symbol Search
	Mazes

Subtests
10 subtests are *standard subtests* that are used to calculate the IQ (Intelligence Quotient).
3 are *supplementary subtests* (Symbol Search, Digit Span, Mazes- not used in the calculation of the Full Scale IQ

IQ Scores

Full Scale IQ = The combination of Verbal and Performance scale scores
Verbal IQ = The total scale scores from the Verbal subtests
Performance IQ = The total scale scores from the Performance subtests

Overall, the WISC-III is a fair measure of "G" with .43 of its variance attributed to g. Subtests from the Verbal scales have higher "g" loading than subtests from the Performance scales.

Highest loadings of "G"
Verbal Scale: Vocabulary, Information, Similarities
Performance Scale: Block Design, Object Assembly, Picture Completion

Basic Principles of Measurement

Derived Scores

- Raw scores are of limited usefulness, but they can be converted into derived scores or scores converted into a relative measure for comparative purposes).
- Derived scores let you compare a child's score to the normative group to determine relative standing. Derived scores allow for comparisons.
- Percentile Scores, Standard Scores, and IQ scores are all derived scores.

Percentile Rank

Based on the normal curve. A percentile rank is a point in a distribution at or below which the scores of a given percentage of individuals fall. Interpretation of a percentile is as follows:
If a child obtains a percentile rank of 60% on the WISC-III, the child has scored as well as or better than 60 percent of the children in the norm sample. The difficulty in explaining this is that all points along the percentile distribution do not represent equal units.

Standard Scores

Standard Scores are raw scores that have been transformed to have a given mean (arithmetic average) and a standard deviation (the extent to which scores deviate from the mean). Examples:

- T-Scores: Mean of 50, SD of 10
- Scaled Scores: Mean of 10, SD of 3 (WISC-III Subtests)
- Deviation IQ: Mean of 100, SD of 16

Psychometric Theory
True Score = Observed Score + Error Score
True score is a hypothetic construct (i.e.: cannot be observed).
The obtained score is a composite of the amount of the trait the child actually possesses (the true score) and the error of measurement (the error score).

The **reliability coefficient** represents the ratio of the true score variance to the observed score variance.
.0 reliability = absence of reliability : 1.00 = perfect reliability

Reliability of the WISC-III

- Stability of the test over time is excellent: a very robust test "outstanding" (Sattler, 1992): "exemplary" (Kamphaus, 1993).
- Several factors affect the reliability of the test:
- Test length (more items, more homogeneity = greater reliability), test-re-test interval (less amount of time between administration = less chance of change), guessing, test situation variables (illness, misunderstanding instructions...)
- Test-re-test **stability coefficients** = Index of stability over time.
- **Internal consistency reliability coefficient**: based on scores obtained during on e test administration and based on the intercorrelations among all comparable parts of the same test. The coefficient reflects the extent to which items measure the same characteristic.
- .89 over the entire age range covered in the standardization (.80 is generally considered to be acceptable).
- .96 for Full Scale IQ
- .95 for Verbal Scale IO
- .91 for Performance Scale IQ

Standard Error of Measurement
Standard Error of Measurement (SEM) is an estimate of the amount of error usually attached to an examinees obtained score. It is directly related to the reliability of the test.

Confidence Interval for Obtained Scores
Because of the uncertainty of exactly where the true score lies, statements are made about the probability that an examinee's obtained score reflects his or her true score and are described in terms of "confidence intervals".

The SEM provides the basis for forming the confidence interval.
- Observed score +/- the error will give you a "confidence band" or a range of scores indicating the probability of the true score within the range of scores.
- You choose the desired level of confidence Interval (C.I.):

68%, 90%, 95% C.I. For example, a 95% C.I. can be thought of as the range in which a person's true score will be found 95 percent of the time. Chances are only 5 in 100 that a person's true score lies outside this confidence interval. It is, however, not possible to construct a confidence interval within which a person's true score is absolutely certain to lie.

Validity of the WISC-III

Validity: the extent to which a test measures what it is suppose to measure. The WISC-III has adequate concurrent and construct validity.

Construct Validity: The extent to which a test measures a psychological construct or trait. Statistical analyses indicates that the WISC-III provides a fair measure of general intelligence and factor analysis shows that the test adequately measures 2 factors corresponding to Verbal and Performance scales on the test.

Content Validity: Whether items on the test are representative of the domain that the test purports to measure.

Criterion-Related Validity: The relationship between test scores and some type of criterion or outcome (e.g.: ratings, classification).

Concurrent Validity: Shows if the test scores relate to some currently available criterion measure (e.g.: WISC-III compared to the WISC-R).

Predictive Validity: The correlation between test scores and performance on a relevant criterion where there is a time interval between the test administration and the performance on the criterion. How accurate is the obtained score as a predictor of furtur performance on the criterion? For example, scores on the WISC-R will be lower than scores on the WISC-III (due to differences in the norm groups).

Factors (that make up the Index Scores)

4 Factors make up the WISC-III:

- Freedom from Distractibility Factor
- Verbal Comprehension Factor
- Perceptual Organization Factor
- Processing Speed Factor (new to the WISC-III)

NOTE: Factor scores aid with hypothesis formulation and interpretation and should not be included in the report.

Freedom from Distractibility Factor

The WISC-R factor structure differs from the WISC-III factor structure. For the WISC-R, Freedom from Distractibility (FD) was named the 3rd factor (included the Arithmetic and Digit Span subtests with the focus on the ability to concentrate or remain attentive with minor loadings on the Information and Coding subtests.

Factor analysis for the WISC-III does not support this factor as it did with the WISC-R and should be disregarded until there is further evidence to support its use. Sattler recommends using a 3 factor model and Saklofske and Prifitera suggest using the FD factor as a Working Memory Index.

On the WISC-III, the FD Index does not really measure distractibility but has continued to be named so for "historical continuity". The Arithmetic and Digit Span FD Index is better interpreted as a "Working Memory Index".

Verbal Comprehension Factor

Describes the hypothetical ability underlying the factor for both:

- Item Content (Verbal) and
- Mental Processes (Comprehension)
- VC factor seems to measure a variable common to most of the verbal scale subtests. In order of highest loading:

Vocabulary, Information, Comprehension, Similarities, Arithmetic (moderate), Digit Span (minimal). Also, Picture Completion, Picture Arrangement, and Block Design (minimal).

- May relate to verbal processing involved on these tasks and high "g" loading on the Block Design subtest.

Perceptual Organization Factor

PC factor describes the hypothetical ability underlying the factor for both:

- Item Content (Perceptual)
- Mental Processes (Organization)
- PO factor seems to measure a variable common to most of the performance subtests. - In order of highest loading:
- Block Design and object Assembly, Picture Completion (moderate), Mazes, Picture Arrangement, and Symbol Search (minimal).

Processing Speed Factor

PS factor describes the hypothetical ability underlying the factor for both:

- Item Content (Perceptual Processing)

Mental Processes (Speed)

- PS factor seems to measure the ability to employ a high degree of concentration and attention in processing information rapidly by scanning an array. In order of highest loading: Coding and Symbol Search

Lecture on Administration and Scoring: WISC-III

See Attached Overheads and handouts for students

Instructional Agenda

1- Use overheads and go over the WISC-III Overview Appendices and Tables in Sattler. Students will be provided with handouts of the overheads so that they have a copy and can follow along making notes as you teach. Present the overhead with the additional information about the rationale of what each subtest measures as you explain the subtests. Go over the starting and discontinuance information with the students. Tell students that the information from the overhead can be found in Appendix J: WISC-III Subtests.
2- Explain Queries, Probes, Multiple responses and Spoiled responses
3- Explain General Administration Information (Filling in the Record Form)

Subtests

Verbal Subtests: 6 in total
Performance Subtests: 7 in total

- **Picture Completion**

The child is asked to identify the single most important missing detail in 30 drawings of common objects, animals, or people. The child must name or point to the essential missing portion f the incomplete picture within the 20 second time limit. All the items are scored 1 or 0 (pass or fail). Discontinue after 5 consecutive failures.

- **Information**

The child responds to a broad range of questions dealing with factual information. There are 30 questions scored 1 or 0 (pass or fail). Subtest not timed. Discontinue after 5 consecutive failures.

- **Coding**

Requires a child to copy symbols paired with numbers. Two parts: Part A for children under age 8 and Part B for children 8 and older. 120 second time limit. One point given for each correct response with additional time-bonus points for a perfect score on Part A.

- **Similarities**

Questions about how objects or concepts are alike. There are 19 pairs of words where the child must state the similarity between the two items in each pair. The first 5 items are scored 1 or 0 (pass or fail) and items 6-19 are scored 2, 1, or 0. Subtest not timed. Discontinue after 4 consecutive failures.

- **Picture Arrangement**

Requires a child to place a series of picture cards in logical order. Each item is timed. All children start with the sample item. Then children age 6 to 8 are given item #1 and children over 8 are given item #3. A number of different trials can be given depending on how successful the child is in correctly completing the trial items. Time-bonus points are awarded. Discontinue after 3 consecutive failures.

- **Arithmetic**

Questions about simple to complex problems involving arithmetical concepts and numerical reasoning. 24 problems altogether with 5 presented on picture cards, 13 presented orally, and 6 presented in written form. Items 1-18 are scored 1 or 0 and items 19 to 24 are scored 2, 1, or 0. All items are timed with different time limits with time bonuses awarded. Discontinue after 3 consecutive failures.

- **Block Design**

Requires a child to reproduce designs using 3-dimensional blocks. There are 12 items. Children 6-8 start with item #1 and children 8 and older start with item #3. All items are timed and time bonus points are awarded. Discontinue after 2 consecutive failures.

- **Vocabulary**

The child is asked to define orally presented words. All items are scored 2, 1, or 0. The subtest is not timed. Discontinue after 4 consecutive failures.

- **Object Assembly**

Requires a child to put jigsaw pieces together to form common objects. Every item is administered to all children. All items are timed and there is no discontinuance (i.e.: all 5 puzzles are administered).

- **Comprehension**

The child is asked to explain situations, actions, or activities that relate to events familiar to most children. All items are scored 2, 1, or 0. The subtest is not timed. Discontinue after 3 consecutive failures.

- **Symbol Search - Supplementary Subtest – Can use Symbol Search to substitute for Coding Subtest ONLY: is not used in the computation of the IQ when the 5 standard performance scales are given.**

Requires a child to look at a symbol (or symbols) and then decide whether the symbol(s) is (are) present in an array of symbols.

- **Digit Span – Supplementary Subtest – Can use Digit Span to substitute for any Verbal subtest: is not used in the computation of the IQ when 5 standard verbal scales are given.**

Has 2 parts: Digits Forward and Digits Backward. The subtest is untimed. Discontinue after failure on both trials of any one item.

- **Mazes – Supplementary Subtest – Can use Mazes to substitute for any Performance subtest: is not used in the computation of the IQ when 5 standard performance scales are given.**

Requires a child to solve paper-and-pencil mazes that differ in level of complexity. Children 6 to 8 start with the sample maze and item #1 whereas children 8 and older start with item #4.

Repetition, Probes, Queries, Spoiled and Multiple Responses

Repetition: Can repeat items if needed except where prohibited (e.g.: Arithmetic = one repetition only; Digit Span = no repetitions).

Probes: If a child refuses to respond or says "I don't know" to an early items but then responds correctly to more difficult items on the same subtest, re-administer the item later and give credit if the child responds correctly. Cannot probe for timed items or Digit Span (see page 46 in the Manual for more information).

Queries: Used for incomplete and ambiguous responses. Used in the Similarities, Vocabulary, and Comprehension subtests when responses are followed by a "Q"which indicates that the response or any equivalent must be queried: "tell me more about it" or "what do you mean".

Spoiled Responses: In a spoiled response, a child's elaboration of an item reveals a fundamental misconception about the item (see page 50 in the Manual for more information).

Multiple Responses: If more than one response is given:

- if the 2nd or 3rd response is meant to replace the 1st response, score only the last response.
- For timed items, score the last response given within the time limit
- If both a correct and incorrect response is given, ask the child which one is intended and score the intended response.

Special Considerations: Prorating
Prorating procedures are used when fewer than 10 subtests are administered. See Appendix A pp.258 in the Manual.

General Administration Information
Show the overhead of the WISC-III front page sheet and use to demonstrate how to complete the Record Form:

- Fill in identifying information
- Show students how to correctly calculate chronological age
- Explain how to transfer raw scores to front page and then convert to scaled scores
- Explain how to sum scaled scores to arrive at IQ's
- Show how to fill in the remainder of the front page (e.g.: IQ scores, %ile, confidence intervals, subtest scores, IQ scores, Index scores).
- Show how to fill in the Behavioral Observations sheet.

Lecture on Specific and frequent administration and scoring errors: WISC-III Experimental Teaching Method

Please instruct student regarding specific and frequent administration and scoring errors on the WISC-III. For each subtest, remind students of the correct starting and discontinuance criteria, establishing the basal and ceiling levels, and scoring the subtests.

1. Picture Completion

Sample Item: All children are given the sample item

3 Queries:

1- If a child only names the object in the picture without saying what part is missing, the query is

Yes, but what's missing?

2- If a child mentions an unessential part that is off the card, note the response on the record form and the query is

A part is missing IN the picture. What is it that is missing?

3- If the child mentions an unessential missing part, note the response on the record form and the query is

Yes, but what is the MOST IMPORTANT part that is missing?

2. Information

Remind students to attend to the specific questions in the manual that must be asked for particular examinee responses. Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

3. Coding and Symbol Search

Remove the Reponse Sheet from the record book. Be sure to administer the correct part: Part A is for ages 6-7 and Part B is for ages 8-16. Be sure to administer the sample items.

Timing: Ensure students know what 120 seconds is on their stopwatches.

Scoring: Remind students that the sample items are not included in the final score. Remind students how to use the scoring template for accurate scoring.

4. Similarities

Sample Item: All children are given the sample item. Remind students to attend to the specific questions in the manual that must be asked for particular examinee responses.

Teaching: Children are given assistance on items #1, #2, and #6.

Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

5. Picture Arrangement

Sample: All children get the sample first.

Procedure: Normal sequence of administration for Item #1 and #2 if a child age 9-16 fails trial #1 of Item #3.

6. Arithmetic

Repetition of items: Any item may be repeated ONCE only. Timing continues while repeating the ENTIRE question, not just portions of it. For example, if a child asks "How many newspapers?" for question #14, the examiner should repeat the entire question. This subtest assesses mental flexibility and concentration.

7. Block Design

Procedure: * Ages 8-16: Pass Trial #1 of Design 3 - go on to Design 4 and give credit for Designs 1 and 2.
* Ages 8-16: Fail Trial #1 of Design 3 - give Trial 1 and 2 in normal sequence (regardless if the child got trial #2 of Design 3 correct)

Time Recording: Ensure accurate time recording for this and other performance subtests (e.g.: PA, OA). Stop timing when the child is clearly finished the task.

8. Vocabulary

Teaching: Give help on item #1 only.

See specific directions for item #9 LEAVE not to be confused with "LEAF", item #16 PRECISE not CONCISE, item #27 AFFLICTION not INFLICTION, item #28, IMMINENT not IMMANENT or EMINENT. Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

9. Object Assembly

Teaching: Give help on item #1 only

Procedure: Administer all items. Ensure that students understand how a child can obtain some points for correct junctures even though the puzzle is not completely put together correctly.

Scoring: Review with students the multiplication procedure to use when examinees complete only a portion of a puzzle correctly.

10. Comprehension

Teaching: On item # 1 only

Procedure: Items with * asterix (#7, 11, 12, 15, 17, 18) require 2 general responses for a 2 point credit and 1 general response earns only 1 point. Review this scoring procedure with students. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

11. Symbol Search

Teaching: Be sure to administer the correct part: Part A is for ages 6-7 and Part B is for ages 8-16. Be sure to administer the sample items.

Scoring: Remind students how to use the scoring template for accurate scoring.

12. Digit Span
Sample Item: All children are given the sample item for Digits Backward.
Procedure: Discontinue after failure on BOTH trials of any item.
Remind students how to say the digits one at a time with a one second interval between digits.

13. Mazes
Procedure: Ensure that the child listens to all directions. Give the sample item for children age 6-7 years old. Be sure to begin at the correct starting point for the child's age. Review how to determine errors such as, entering blind alleys and overshoots.
Scoring: Remind students about how to assign points for this subtest.

Introductory Lecture: WAIS-III
Brief History of the WAIS-III
See attached overheads and handouts

Development of the WAIS

- The *Wechsler Bellvue Intelligence Scale*-Form 1 (1939), Form 2 (1946)
- The *WAIS* was published in 1955
- The *WAIS-R* was published in 1981
- The *WAIS-III* was published in 1997

Usual Reasons for Renorming an IQ test

1. IQ gains over time
2. Average IQ increases 1/3 to ½ a scaled score per year.
3. Development of new norms
4. Re-assessing the factor structure

Reasons for renorming the WAIS-III

1. **Normative range includes older adults**
   -Modified artwork-larger stimuli
   -reduced emphasis on speed and bonus points
2. **Improvement of the floor**
   -from as low as FSIQ 45, VIQ 48, PIQ 47
3. **Non-biased items/Contemporary items**
   -external review by bias experts/analysis of items
4. **New supplementary/Optional subtest**
   -Letter-Number Sequencing (Verbal scale) & Symbol Search (Performance Scale). They contribute only to the Index scores.
5. **New record form**
6. **Measurement of factor domains**
7. **Extensive validation research**

WAIS-III Sampling
Demographic Stratification Variables
WAIS-III standardized on 2, 450 individuals.
Age: extended age range 16-89
Gender: more women in the age group 65-89
Education level: 5 educational categories
Ethnicity: Caucasian, African American, Hispanic, and Other
Region or Country: Northeast, North Central, South, and West
WAIS-III Canadian Norming: results yet to be determined.

Reminder: A **norm** provides an indication of average or typical performance of the specified group. Norms are needed because a raw test score in itself is not very meaningful. A comparison is made between an individual's raw score to another individual's raw score by converting the scores into some relative measure.

In **norm-referenced testing**, an examinee's performance is compared to a specific group of subjects.
Developing current norms gives more precise scores for individuals and ensures that the **"Flynn Effect"** (where scores become inflated over time and norms need to be reestablished) does not occur.

WAIS-III Subtests
Updated from the WAIS-R

Scale	New Subtests
Verbal Scale	
-Information	-Letter-Number Sequencing
-Vocabulary	
-Similarities	
-Comprehension	
-Digit Span	
-Arithmetic	
Performance Scale	
Picture Completion	
Block Design	Matrix Reasoning
Picture Arrangement	Symbol Search
Digit Symbol	
Object Assembly	
Digit Span	

WAIS-III New Subtests

Matrix Reasoning
- Measures abstract fluid reasoning
- Enhances Performance IQ
- Untimed and without manipulative
- Excellent subtest for older adults
- Relatively culture-fair and language-free

Symbol Search
- Loads highest on the Speed of Information Factor
- Correlates highest with Digit Symbol

Letter-Number Sequencing
- Strengthens the Attention/Working Memory Factor
- Correlates with Arithmetic and Digit Span
- Easily administered, five minute subtest

Reliability of the WAIS-III
The WAIS-III has excellent reliability.

WAIS-III Reliability Coefficients

IQ or Index	Reliability
VIQ	0.97
PIQ	0.94
FSIQ	0.98
VCI	0.96
POI	0.93
WMI	0.94
PSI	0.88

Factor Based Composite Scores
4-Factor Model

Verbal	Perceptual Organization	Attention/Working Memory	Speed of Information Processing
Vocabulary	Block Design	Arithmetic	Digit Symbol
Information	Matrix Reasoning	Digit Span	Symbol Search
Similarities	Picture Completion	Letter-Number Sequencing	

Standard Error of Measurement
Standard Error of Measurement (SEM) is an estimate of the amount of error usually attached to an examinees obtained score. It is directly related to the reliability of the test. In IQ points for the WAIS-III are:
FSIQ -2.30
VIQ -2.55
PIQ -3.67
More confidence can be placed on the FSIQ than the VIQ or PIQ and more confidence can be placed in the VIQ than the PIQ.

Verbal Scale: Vocabulary has the smallest average SEM (.79) and Letter-Number Sequencing has the largest average SEM (1.30).

Performance Scale: Matrix Reasoning has the smallest average SEM (.97) and Object Assembly has the largest average SEM (1.66).

Confidence Interval for Obtained Scores
Because of the uncertainty of exactly where the true score lies, statements are made about the probability that an examinee's obtained score reflects his or her true score and are described in terms of "confidence intervals".

Validity of the WAIS-III

Validity: the extent to which a test measures what it is suppose to measure. The WAIS-III has adequate concurrent and construct validity. Approximately 70% of the items on the WAIS-III are from the WAIS-R and approximately 30% are new items. Due to the similarities, it is generally assumed that research concerning the validity of the WAIS-R applies to the WAIS-III.

Construct Validity: The extent to which a test measures a psychological construct or trait. The studies in the *WAIS-III – WMS-III Technical Manual* indicate good construct and concurrent validity, although more research is needed to evaluate the different form s of validity, especially for the Index scores.

Content Validity: Whether items on the test are representative of the domain that the test purports to measure.

Index/ Factor Scores

Index scores* is another term for *factor scores

Verbal IQ

Measures acquired knowledge, verbal reasoning, and attention to verbal materials.

Performance IQ

Measure fluid reasoning, spatial processing, attentiveness to detail, and visual-motor integration.

Full Scale IQ

The overall summary score that estimates an individual's general level of intellectual functioning.

IQ scores versus Index Scores

Not all subtests need be administered to obtain the IQ scores or Index scores (see pages 1-5 in the WAIS-III Administration and Scoring Manual).

1. If all 13 subtests area administered, IQ scores and Index scores will result. For the purposes of this course, students should administer all 14 subtests.
2. If Object Assembly, Letter-Number Sequencing, and Symbol Search are omitted, will have IQ scores, VCI and POI scores, but not WMI or PSI scores.
3. The Picture Arrangement and Comprehension subtests are needed to compute the IQ scores but are note included in the computation of the Index scores.

Four Factors make up the WAIS-III:

- Verbal Comprehension Index
- Perceptual Organization Index
- Working Memory Index
- Processing Speed Index

NOTE: Factor scores aid with hypothesis formulation and interpretation and should not be included in the report.

Verbal Comprehension Factor

Measures verbally-acquired knowledge and verbal reasoning. The difference between the VCI and the VIQ is the a VIC does not include the Comprehension subtest or the Digit Span and Arithmetic subtests which make up the Working Memory Index. This index may be conceptualized as a more refined, "purer' measure of verbal comprehension.

Perceptual Organization Factor

Measures nonverbal, fluid reasoning, attentiveness to detail, and visual-motor integration. The difference between the PI and the PIQ is that the POI does not include the Digit Symbol-Coding subtests which makes up the Processing Speed Index. The composition of the POI makes it a more refined measure of fluid reasoning and visual-spatial problems solving than the PIQ.

Working Memory Index

Measure attention to information, and ability to hold briefly and process information in memory, then formulate a response. Includes only verbally presented items.

Processing Speed Factor

Measures the ability to process visual information quickly. Comparisons between the PSI and POI scores can reveal possible effects of time demands on visual-spatial reasoning and problem solving.

Substitutions

Can substitute subtests if a subtest is spoiled or cnanot be administered. Only one subtsitution per scale should be made. No substitutionons can be made to obtain the Index scores.

- Letter-Number Sequencing can substitute for Digit Span only
- Symbol Search can substitute for Digit Symbol-Coding only
- Object Assembly can Substitute for any other Performance subtest for examinees 74 and younger.

Additional Information

- To compute the IQ scores, must have 3 Verbal and 3 Performance subtests with raw scores greater than 0.
- The range of IQs is 45-155 for all age groups.
- The record form contains a profile page, score conversion page, discrepancy analysis page, and demographics page. These additional tables of information allow for the determination of strengths and weaknesses, and provide confidence limits.

Lecture on Administration and Scoring: WAIS-III
See attached overheads and handouts for students

Instructional Agenda

1- Use overheads and go over the WAIS-III Overview Appendices and Tables in Sattler. Students will be provided with handouts of the overheads so that they have a copy and can follow along making notes as you teach. Present the overhead with the additional information about the rationale of what each subtest measures as you explain the subtests. Go over the starting and discontinuance information with the students. Tell students that the information from the overhead can be found in Appendix J: WAIS-III Subtests.
2- Explain Queries, Probes, Multiple responses and Spoiled responses
3- Explain General Administration Information (Filling in the Record Form)

Subtests

Verbal Subtests: 7 in total
Performance Subtests: 7 in total

- **Picture Completion**

The examinee is asked to identify the single most important missing detail in 25 drawings of common objects, animals, or people. The examinee must name or point to the essential missing portion of the incomplete picture within the 20-second time limit. All items are scored pass or fail (0 or 1). Discontinue after 5 consecutive failures.

- **Vocabulary**

The examinee is asked to define words. Each word is presented orally and in print and the examinee is asked to explain its meaning aloud. All responses are scored 0, 1, or 2. Discontinue after 4 failures.

- **Digit Symbol- Coding**

Is similar to the Coding B on the WAIS-III. The subtest requires the examinee to copy symbols that are paired with numbers.

There are two optional procedures: Digit Symbol **Incidental Learning** and **Copy**: they are intended to help you determine what skills may be deficient if the examinee performs poorly on the Digit Symbol-Coding subtest: they are not used to compute the IQs.

- **Digit Symbol- Incidental Learning**

(Optional Procedure)

Incidental Learning is a measure of the examinee's ability to recall the associated number-symbol pairs and the individual symbols, independent of numbers. This procedure is administered immediately after the Digit Symbol- Coding subtest and is untimed.

- **Digit Symbol- Copy**
(Optional Procedure)

Copy is a measure of the examinee's ability to use graphomotor speed. The examinee is asked to copy into the blank box the symbol in the box above it. The same 9 stimuli used in Digit Symbol- Coding are used here. Time limit is 90 seconds with one point allotted for each correct item.

- **Similarities**

Contains 19 pairs of words: the examinee is asked to explain the similarity between the two words in each pair. The first 5 items are scored 0 or 1; items 6 through 19 are scored 2, 1, or 0, depending on the conceptual level of the response. Discontinue after 4 consecutive failures.

- **Block Design**

Requires an examinee to reproduce designs using 3-dimensional blocks. There are 14 items. The patterns are arranged in order of increasing difficulty. 2 blocks are used for items 1 and 2, 4 blocks are needed to reproduce items 3 through 9, and 9 blocks are required for items 10 through 14. All items are timed and time bonus points are awarded on items 7 through 14. Discontinue after 3 consecutive failures.

- **Arithmetic**

Contains 20 items: 17 are given orally and the other 3 use blocks along with oral directions. All problems are timed, with items 1 though 6 having a time limit of 15 seconds, items 7 though 11, 30 seconds, items 12 through 19, 60 seconds, and item 20; 120 seconds. All items are scored 1 or 0, with one additional time-bonus point possible on items 19 and 20. Discontinue after 4 consecutive failures.

- **Matrix Reasoning**

26 non-verbal reasoning tasks. The items consist of individually presented colored matrixes, each of which is missing a part. The examinee is directed to look at all aspects of each matrix carefully and select the missing part from an array of 5 choices at the bottom of the page. The subtest is untimed. Discontinue after 4 consecutive failures or 4 failures in five consecutive items.

- **Digit Span**

This is a regular subtest used in the calculation of the IQ scores, whereas on the WAIS-III it is a supplementary subtest. Has 2 parts: Digits Forward (which contains series of numbers ranging from 2 to 9 digits in length) and Digits Backward (which contains series of numbers ranging from 2 to 8 digits in length). All series are scored 2, 1, or 0.

- **Information**

28 questions that sample a broad range of general knowledge about common events, objects, and places. Items dealing with historical and geographic facts are also included. Each item is scored 1 or 0 (pass/fail). Discontinue after 6 consecutive failures.

- **Picture Arrangement**

Requires the examinee to arrange a series of pictures in a logical sequence. Each of the 11 series of pictures is presented in a specified disarranged order, and the examinee is asked to rearrange the pictures in the "right" order to tell a story. All examinees begin with item #1. All items are timed. Discontinue after 4 consecutive scores of 0.

- **Comprehension**

Contains 18 questions covering a wide range of situations and proverbs. Questions deal with such issues as government operations and laws, health standards, and social mores. Items 1 through 3 are scored 0 or 1; items 4 though 18 are scored 0, 1, or 2. The subtest is not timed. Discontinue after 4 consecutive failures.

- **Symbol Search - Supplementary Subtest**

Requires an examinee to look at two symbols and decide whether either symbol is present in an array of five symbols. The subtest contains 60 items and has a time limit of 120 seconds. The score is total correct minus total incorrect.

- **Letter-Number Sequencing - Supplementary Subtest**

Contains 7 items, each consisting of 3 trials. Each trial required the examinee to order sequentially a series of numbers and letters that are orally presented in a specified random order. All examinees begin with item #1 and testing discontinues after failure on all 3 trials of an item. Is not used in the computation of the IQ when the 6 standard Verbal Scale subtests are administered. It may give useful information if there are attentional problems.

- **Object Assembly- Optional Subtest**

Requires the examinee to put jigsaw pieces together to form common objects. Examinees are administered all 5 items. All items are timed with bonus points awarded.

Lecture on Specific and frequent administration and scoring errors: WAIS-III Experimental Teaching Method

Please instruct student regarding specific and frequent administration and scoring errors on the WAIS-III. For each subtest, remind students of the correct starting and discontinuance criteria, establishing the basal and ceiling levels, and scoring the subtests.

1. Picture Completion

Sample Item: All examinees are given the sample item

3 Queries:

1- If an examinee only names the object in the picture without saying what part is missing, the query is

Yes, but what's missing?

2- If an examinee mentions an unessential part that is off the card, note the response on the record form and the query is

A part is missing IN the picture. What is it that is missing?

3- If the examinee mentions an unessential missing part, note the response on the record form and the query is

Yes, but what is the MOST IMPORTANT part that is missing?

2. Vocabulary

Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

3. Digit Symbol: Coding

Procedure: Be sure to administer the sample items. If Incidental Learning is to be administered, remind students that the examinee must complete four rows of test items, even if the 120 second time limit has passed (examiners then need to mark the item completed at the 120 second time mark).

Timing: Ensure students know what 120 seconds is on their stopwatches.

Scoring: Remind students that the sample items are not included in the final score. Remind students how to use the scoring template for accurate scoring.

Digit Symbol: Incidental Learning

Procedure: Remind students how to administer and score this optional procedure.

Digit Symbol: Copy

Procedure: Remind students how to administer and score this optional procedure.

4. Similarities

Teaching: Examinees are given assistance on item #6.

Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

5. Block Design

Procedure: Remind students of how to begin this subtest: if the examinee scores a 0 or 1 on either Design 5 or 6, administer Design 1-4 in *reverse* sequence until the examinee obtains perfect scores on 2 consecutive items.

Time Recording: Ensure accurate time recording for this and other performance subtests (e.g.: PA, OA).

6. Arithmetic

Repetition of items: Any item may be repeated ONCE only. Timing continues while repeating the ENTIRE question, not just portions of it. For example, if an examinee asks "How many pieces of chocolate?" for question #12, the examiner should repeat the entire question. This subtest assesses mental flexibility and concentration.

7. Matrix Reasoning

Procedure: Since this is a new subtest, spend time ensuring that the subtest administration and scoring is understood. All examinees must complete the sample items with teaching provided to any incorrect responses. All examinees then begin with #4 and items are administered in reverse sequence until prefect scores are achieved on two consecutive items.

8. Digit Span

Sample Item: All examinees are given the sample item for Digits Backward.

Procedure: Discontinue after failure on BOTH trials of any item.

Remind students how to say the digits one at a time with a one second interval between digits.

9. Information

Remind students to attend to the specific questions (i.e., #6 and # 21) in the manual that must be asked for particular examinee responses. Review scoring for 2, 1 or 0 point responses. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

10. Picture Arrangement

Sample: All examinees get the sample first.

Scoring: Remind students that 5 of the items (i.e., #5-#9) have acceptable arrangement variations that earn 1 point.

11. Comprehension

Procedure: Items with * asterix (5, 6, 7, 10, 13) require 2 general responses for a 2 point credit and 1 general response earns only 1 point. Review this scoring procedure with students. Review how to query appropriately, how to determine if a response is spoiled, and what to do with multiple responses.

12. Symbol Search

Teaching: Be sure to administer the sample items.

Scoring: Remind students how to use the scoring template for accurate scoring.

13. Letter-Number Sequencing:
Procedure: Since this is a new subtest, spend time ensuring that the subtest administration and scoring is understood.
Teaching: Be sure to administer the practice trials.
Scoring: Explain to students that examinees can obtain credit as long as the numbers and letters are recalled in sequence (even if the letters are recited before the numbers).

14. Object Assembly
Procedure: Administer all items. Ensure that students understand how an examinee can obtain some points for correct junctures even though the puzzle is not completely put together correctly.
Scoring: Review with students the multiplication procedure to use when examinees complete only a portion of a puzzle correctly.

Interpretation Lecture: WAIS-III

See attached overheads and handouts

- See Sattler & Ryan's Supplement (1998)
- See Appendix K (Interpreting the WISC-III also applies to the WAIS-III)
- See Appendix O
- See WAIS-III Administration and Scoring Manual and the Technical Manual

Steps in Interpretation

1. Validity of the Profile

Is your profile valid?

- *Status of client:* willingness to complete assessment, level of commitment to the process, physical/emotional state.
- *Test Results:* Any reasons to question the validity of the test results? Reading level too low, comprehension poor...

2. Profile Analysis

Refers to interpreting or analyzing the pattern of scaled scores and deviation IQ's obtained by an individual examinee. Method used to generate hypotheses about the organization of intellectual abilities.

3. Comparing Verbal and Performance IQ's.

- See Table O-2 in Appendix O for critical values.
- See Table O-4 in Appendix O for probabilities associated with the V-P differences.

- **Table B.2 in the WAIS-III Administration and Scoring Manual**: shows cumulative percentages in the standardization sample but are *Absolute Values* (i.e.: they represent both bi-directional differences). No info given regarding differences in either direction alone. To estimate differences had they been directional, divide values by 2.

- **Tables D.1 to D.5 of the Technical Manual** are also absolute values. These values must also be divided by 2 to get an estimate of the base rate (i.e.: frequency of an occurrence) of V/P differences for one direction only.

4. Comparing each Verbal Subtest Scaled score with the mean Verbal Scaled Score

- See Table O-3 provides critical values
- Table B.3 in the WAIS-III Administration and Scoring Manual: gives the cumulative frequencies with which various differences occurred in the standardized sample between an examinee's scaled score on each subtest and the average Verbal scale average

5. Comparing each Performance Subtest Scaled score with the mean Performance Scaled Score

- See Table O-3 provides critical values
- Table B.3 in the WAIS-III Administration and Scoring Manual: gives the cumulative frequencies with which various differences occurred in the standardized sample between an examinee's scaled score on each subtest and the average Performance scale average

6. Comparing each Subtest Scaled Score with on of the following means:
- the standard 11 subtests
- the standard 11 subtests and L-N Subtest
- the standard 11 subtests, L-N Subtest and SS Subtest
- all 14 Subtests
- the standard 11 subtests substituting L-N Subtest with DS
- the standard 11 subtests substituting DS Subtest with SS Subtest

7. Comparing pairs of individual subtest scaled scores.
- See Table O-2 in Appendix O
- for multiple comparisons, ensure that 6 pts spread exists between highest and lowest scaled scores.

8. Comparing Verbal Comprehension, Perceptual Organization, Working Memory, and Processing Speed Index Scores.
- See Appendix O, Table O-5 shows probabilities associated with various differences between WAIS-III Index Scores
- See Table B.2 in the WAIS-III Administration and Scoring Manual for cumulative percentages (absolute values, divide by 2)/
- See Table D.1 and D.5 in the Technical Manual for frequency distributions of pairs of Index score differences in both directions at five ability levels (absolute values).

9. Comparing Subtest Scaled Scores in each Index with their respective Index Mean.
- See Table O-3 in Appendix O for critical values.

10. Comparing the Examinee's subtest scaled score range to the range found in the standardization sample.
- See Table B.5 in the WAIS-III Administration and Scoring Manual. Provides base rate information about what occurred in the standardization sample.

11. Statistically Reliable vs. Empirically Observed Differences
- Table O-2 in Appendix O gives differences required between V/P IQs and Index scores for statistical significance. *Represents statistical difference.*

- Table B.2 in the WAIS-III Administration and Scoring Manual: gives the actual (i.e.: empirically observed) base rates of the frequencies of differences between the IQs and the Index scores found in the standardization sample. *Represents frequency of occurrence.*

- Rule of thumb:
If it occurs in 15 % or less of the population (one direction), it may be considered unusual and rare.

Appendix C
Administration and Scoring Checklists

WAIS-III Checklist

Supervisor's Name: _____

Student's Name: _____

INQ = Incorrect Query FQ = Failure to Query **Protocol Date:**

Front Page	**P1**	**P2**	**P3**	**P4**
Not recording demographic information				
Score Conversion Page				
Incorrectly transferring raw subtest scores (#)				
Making errors in converting raw scores to SS (#)				
Converting Verbal SS to VIQ incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate Verbal IQ				
Converting Performance SS to PIQ incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate Performance IQ				
Converting Full SS to FSIQ incorrectly				
* Adding subtest scaled scores incorrectly				
*Using wrong Verbal and Perf. Subtests				
Converting VCI SS to VCI incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate VCI				
Converting POI SS to POI incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate POI				
Converting WMI SS to WMI incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate FD				
Converting PSI SS to PSI incorrectly				
*Adding subtest scaled scores incorrectly				
*Using wrong subtests to calculate PSI				
Using wrong Reference Group Scaled Scores				
Incorrect Verbal Mean Score				
Incorrect Performance Mean Score				
Incorrect Overall Mean Score				
Optional Procedures:				
Incorrectly transferring raw digit symbol scores				
Using wrong cumulative %				
Determining Strengths & Weaknesses				
Incorrectly transferring scaled score (#)				
Incorrectly identifying mean score(#)				
Incorrect difference from mean(#)				

Failure to complete Reference Group SS column				
Failure to complete Optional Procedures				
Failure to convert scores				
Not identifying statistical significant level (#)				
Incorrectly identifying strengths (#)				
Incorrectly identifying weaknesses (#)				
Incorrectly identifying Freq.of Diff.in Std. Sampl(#)				
Using incorrect table				
Total Score Conversion scoring errors				
Profile Page				
Incorrectly transferring Sums of Scaled Scores (#)				
Converting verbal scaled score to VIQ incorrectly				
Converting Perf. scaled core to PIQ incorrectly				
Converting Full scaled score to FSIQ incorrectly				
Converting VCI scaled score to VCI incorrectly				
Converting POI scaled score to POI incorrectly				
Converting WMI scaled score to WMI incorrectly				
Converting PSI scaled score to PSI incorrectly				
Errors in reporting IQ/Index Percentiles (#)				
Errors in reporting Confidence Intervals(#)				
Moving to the Graphs				
Incorrectly charting IQ scores onto graphs (#)				
Incorrectly charting Index scores onto graphs(#)				
Incorrectly charting subtest scores onto graphs(#)				
Total Profile Page scoring errors				
Discrepancy Analysis Page				
Incorrectly transferring Verbal & Perf. IQ				
Incorrectly transferring VCI, POI, WMI, PSI scores				
Incorrectly calculating Difference				
Incorrectly recording Statistical Significance				
Incorrectly recording Freq. of Diff.in Std. Sample				
Incorrectly transferring Digit Span scores				
Incorrectly calculating difference of Digits				
Incorrectly recording Freq.of Dif. in Std. Sample				
Failure to complete any part of Discrep. Analysis Pg				
Total Discrepancy Analysis Page scoring errors				
Demographics Page				
Not recording demographic information				

Miscalculation of chronological age				
Not recording behavioral observations				
Behavioral observations noted in protocol (Y or N)				
Total Demographics Page scoring errors				
Grand TOTAL "Pages" Errors				
1. Picture Completion Errors	**P1**	**P2**	**P3**	**P4**
Establishing a basal incorrectly				
Not recording responses verbatim				
Not recording incorrect response				
Incorrect point assignment (0 or 1)				
Failing to show OT if response correct				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Picture Completion scoring errors				
2. Vocabulary Errors	**P1**	**P2**	**P3**	**P4**
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-33:				
1. bed (0, 1, or 2)				
2. ship (0, 1, or 2)				
3. penny (0, 1, or 2)				
4. winter (0, 1, or 2)				
5. breakfast (0, 1, or 2)				
6. repair (0, 1, or 2)				
7. assemble (0, 1, or 2)				
8. yesterday (0, 1, or 2)				
9. terminate (0, 1, or 2)				
10. consume (0, 1, or 2)				
11. sentence (0, 1, or 2)				
12. confide (0, 1, or 2)				
13. remorse (0, 1, or 2)				
14. ponder (0, 1, or 2)				
15. compassion (0, 1, or 2)				
16. tranquil (0, 1, or 2)				
17. sanctuary (0, 1, or 2)				
18. designate (0, 1, or 2)				
19. reluctant (0, 1, or 2)				

20. colony (0, 1, or 2)				
21. generate (0, 1, or 2)				
22. ballad (0, 1, or 2)				
23. pout (0, 1, or 2)				
24. plagiarize (0, 1, or 2)				
25. diverse (0, 1, or 2)				
26. evolve (0, 1, or 2)				
27. tangible (0, 1, or 2)				
28. fortitude (0, 1, or 2)				
29. epic (0, 1, or 2)				
30. audacious (0, 1, or 2)				
31. ominous (0, 1, or 2)				
32. encumber (0, 1, or 2)				
33. tirade (0, 1, or 2)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Vocabulary scoring errors				
3. Digit Symbol-- Coding Errors	**P1**	**P2**	**P3**	**P4**
Failure to record completion time				
Failure to add total raw score correctly				
Failure to record total raw score				
Digit Symbol-- Incidental Learning Errors				
Failure to record correct score (Pairing)				
Failure to record correct score (Free Recall)				
Failure to add total raw score correctly				
Failure to complete Incidental Learning				
Digit Symbol-- Copy Errors	**P1**	**P2**	**P3**	**P4**
Failure to record completion time				
Failure to record total raw score				
Failure to add total raw score correctly				
Failure to complete Digit Symbol-Copy				
Total Digit Symbol scoring errors				
4. Similarities Errors	**P1**	**P2**	**P3**	**P4**
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-19:				
1. Fork-Spoon (0, 1, or 2)				

2. Socks-Shoes(0, 1, or 2)				
3. Yellow-Green (0, 1, or 2)				
4. Dog-Lion (0, 1, or 2)				
5. Coat-Suit (0, 1, or 2)				
6. Piano-Drum (0, 1, or 2)				
7. Orange-Banana (0, 1, or 2)				
8. Eye-Ear (0, 1, or 2)				
9. Boat-Automobile (0, 1, or 2)				
10. Table-Chair (0, 1, or 2)				
11. Work-Play (0, 1, or 2)				
12. Steam-Fog (0, 1, or 2)				
13. Egg-Seed (0, 1, or 2)				
14. Democracy-Monarchy (0, 1, or 2)				
15. Poem-Statue (0, 1, or 2)				
16. Praise-Punishment (0, 1, or 2)				
17. Fly-Tree (0, 1, or 2)				
18. Hibernation-Migration (0, 1, or 2)				
19. Enemy-Friend (0, 1, or 2)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Similarities scoring errors				
5. Block Design Errors	**P1**	**P2**	**P3**	**P4**
Failure to record time				
Establishing a basal incorrectly				
Failing to circle Yes or No				
Failure to award correct number of bonus points				
Failure to circle correct point assignment				
Failure to fill in incorrect designs				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Incorrect point assignment				
Total Block Design scoring errors				
6. Arithmetic Errors	**P1**	**P2**	**P3**	**P4**
Failure to record time				
Failing to record correct response				
Failing to record incorrect response				
Failure to award correct score (0 or 1)				
Failure to award correct bonus points				

Failure to circle correct bonus points				
Establishing a basal incorrectly				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Arithmetic scoring errors				
7. Matrix Reasoning Errors	**P1**	**P2**	**P3**	**P4**
Failure to administer sample items (A, B, and C)				
Establishing a basal incorrectly				
Incorrect point assignment (0 or 1)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Matrix Reasoning scoring errors				
8. Digit Span Errors	**P1**	**P2**	**P3**	**P4**
Failure to record correct response				
Failure to record incorrect response				
Failure to record Item Score				
Failure to record Trial Score				
Failure to add scores correctly				
Incorrect point assignment (0, 1, or 2)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Total Digit Span scoring errors				
9. Information Errors	**P1**	**P2**	**P3**	**P4**
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-28:				
1. Saturday (0 or 1)				
2. Age (0 or 1)				
3. Ball (0 or 1)				
4. Months (0 or 1)				
5. Thermometer (0 or 1)				
6. Surmise (0 or 1)				
7. Weeks (0 or 1)				
8. Hamlet (0 or 1)				
9. Brazil (0 or 1)				
10. MLK, Jr. (0 or 1)				
11. Civil War President (0 or 1)				

12. Cleopatra (0 or 1)				
13. Italy (0 or 1)				
14. Relativity (0 or 1)				
15. Olympics (0 or 1)				
16. Sahara Desert (0 or 1)				
17. Genesis (0 or 1)				
18. Sistine Chapel (0 or 1)				
19. Gandhi (0 or 1)				
20. Koran (0 or 1)				
21. Water (0 or 1)				
22. Vessels (0 or 1)				
23. Catherine (0 or 1)				
24. Continents (0 or 1)				
25. Curie (0 or 1)				
26. World Population (0 or 1)				
27. Speed of Light (0 or 1)				
28. Faust (0 or 1)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Information scoring errors				
10. Picture Arrangement	**P1**	**P2**	**P3**	**P4**
Failure to record time				
Failure to record correct sequence				
Failure to record incorrect sequence				
Incorrect point assignment (0, 1, or 2,)				
Establishing a basal incorrectly				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total Picture Arrangement scoring errors				
11. Comprehension Errors	**P1**	**P2**	**P3**	**P4**
Failure to establish correct basal				
Failure to record responses verbatim				
Incorrect point assignment on items 1-18:				
1. Money (0, 1, or 2)				
2. Watches (0, 1, or 2)				
3. Clothes (0, 1, or 2)				
4. Envelope (0, 1, or 2)				
5. Food (0, 1, or 2)				
6. Parole (0, 1, or 2)				

7. Child labor (0, 1, or 2)				
8. Professional service (0, 1, or 2)				
9. Taxes (0, 1, or 2)				
10. History (0, 1, or 2)				
11. Deaf (0, 1, or 2)				
12. Forest (0, 1, or 2)				
13. Jury (0, 1, or 2)				
14. City land (0, 1, or 2)				
15. Marriage license (0, 1, or 2)				
16. Free press (0, 1, or 2)				
17. Swallow (0, 1, or 2)				
18. Shallow brooks (0, 1, or 2)				
Fail to obtain 2nd response				
5. Food				
6. Parole				
7. Child labor				
10. History				
13. Jury				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Failure to record Demographic Information				
Total Comprehension scoring errors				
12. Symbol Search Errors	**P1**	**P2**	**P3**	**P4**
Failure to record the Completion time				
Failure to record correct and incorrect				
Failure to subtract incorrect from correct				
Failure to give the sample item				
Failure to give the practice item				
Incorrect total raw score				
Total Symbol Search scoring errors				
13. Letter-Number Sequencing Errors	**P1**	**P2**	**P3**	**P4**
Failure to record correct response				
Failure to record incorrect response				
Failure to enter correct trial score (0 or 1)				
Failure to enter correct item score (0, 1, 2, or 3)				
Failure to add Total raw score correctly				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				

Did not give credit for reverse sequencing				
Total Letter-Number scoring errors				
14. Object Assembly Errors	**P1**	**P2**	**P3**	**P4**
Failure to record time				
Failure to record the number of correct junctures				
Multiplying score incorrectly				
Failure to award bonus points when appropriate				
Discontinuing subtest incorrectly				
Failure to add total raw score correctly				
Failure to complete Object Assembly				
Total Object Assembly scoring errors				
Grand TOTAL "Pages" Errors				
Grand TOTAL "Pages" Errors ***-with missing data***				
Grand TOTAL All Subtest Scoring				
Grand TOTAL All Subtest Scoring Errors ***- with missing data***				
Grand TOTAL (Subtest & Page Errors)				
Grand TOTAL (Subtest & Page Errors) ***- with missing data***				

Scorer's Initial _____

WISC-III Administration and Scoring checklist

Supervisor's Name : _____

Student's Name: _____ **Session:** _____

INQ = Incorrect Q

FQ = Failure to Q

Date: _____

Front Page	**P1**	**P2**	**P3**	**P4**
Not recording demographic information on protocol				
Miscalculation of chronological age				
Used American Norms				
Incorrectly transferring raw subtest scores to profile page				
Making errors in converting raw scores to scaled scores				

Converting Information				
Converting verbal scaled score to VIQ incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong subtests to calculate Verbal IQ				
Converting performance scaled core to PIQ incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong subtests to calculate Performance IQ				
Converting full scaled score to FSIQ incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong Verbal and Perf. subtests				
Converting VCI scaled score to VCI incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong subtests to calculate VCI				
Converting POI scaled score to POI incorrectly				
* Adding subtest scaled scores incorrectly				
	P1	**P2**	**P3**	**P4**

* Using wrong subtests to calculate POI				
Converting FD scaled score to FD incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong subtests to calculate FD				
Converting PSI scaled score to PSI incorrectly				
* Adding subtest scaled scores incorrectly				
* Using wrong subtests to calculate PSI				
Incorrectly converting IQ scores to % and Conf. Intervals				
Failing to fill in graphs on the profile page				

Moving to the Graphs				
Incorrectly transferring IQ scores				
Incorrectly transferring Index scores				
Incorrectly transferring Subtest scores				
Incorrectly charting IQ scores onto the graphs				
Incorrectly charting Index scores onto the graphs				
Incorrectly charting subtest scores onto the graphs				

Behavioural Observations				
Not recording behavioural observations on page 2				
Behavioural observations noted throughout protocol ? (Y or N)				

Total Errors Page 1 & 2				

Picture Completion Errors	P1	P2	P3	P4
Failing to begin with the correct item according to the child's age				
Establishing a basal incorrectly				
Not recording responses verbatim				
Not recording incorrect response				
Incorrect point assignment				
Failing to show OT if response correct				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Information Errors	P1	P2	P3	P4
Failing to begin with the correct item according to the child's age				
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-28:				
1. nose (0 or 1)				
2. ears (0 or 1)				
3. legs (0 or 1)				
4. Thursday (0 or 1)				
5. boil (0 or 1)				
6. coins (0 or 1)				
7. March (0 or 1)				
8. week (0 or 1)				
9. seasons (0 or 1)				
10. dozen (0 or 1)				

Information Errors	P1	P2	P3	P4
11. hours (0 or 1)				
12. stomach (0 or 1)				
13. Columbus (0 or 1)				
14. oceans (0 or 1)				
15. leap year (0 or 1)				
16. sun (0 or 1)				
17. oxygen (0 or 1)				
18. Brazil (0 or 1)				
19. water (0 or 1)				
20. bulb(0 or 1)				
21. population (0 or 1)				
22. Frank (0 or 1)				
23. hieroglyphics(0 or 1)				
24. glass (0 or 1)				
25. Greece (0 or 1)				
26. rust (0 or 1)				
27. Barometer (0 or 1)				
28. Darwin (0 or 1)				
29. London (0 or 1)				
30. turpentine (0 or1)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Coding Errors	P1	P2	P3	P4

Using wrong form (A or B)				
Failure to give sample item				
Failure to record time				
Failure to award correct bonus points				
Failure to add scores correctly				
Incorrect point assignment				
Total Subtest Scoring Errors				

Similarities Errors	**P1**	**P2**	**P3**	**P4**
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-19:				
1. milk-water (0, 1, or 2)				
2. candle-lamp (0, 1, or 2)				
3. shirt-shoe (0, 1, or 2)				
4. piano-guitar (0, 1, or 2)				
5. wheel-ball (0, 1, or 2)				
6. apple-banana (0, 1, or 2)				
7. cat-mouse (0, 1, or 2)				
8. elbow-knee (0, 1, or 2)				
9. telephone-radio (0, 1, or 2)				
10. anger-joy (0, 1, or 2)				
11. family-tribe (0, 1, or 2)				
12. painting-statue (0, 1, or 2)				
13. ice-steam (0, 1, or 2)				
14. mountain-lake (0, 1, or 2)				
15. temperature-length (0, 1, or 2)				
16. first-last (0, 1, or 2)				
Similarities Errors	**P1**	**P2**	**P3**	**P4**

17. rubber-paper (0, 1, or 2)				
18. the numbers 9 and 25 (0, 1, or 2)				
19. salt-water (0, 1, or 2)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Picture Arrangement	**P1**	**P2**	**P3**	**P4**
Failing to begin with the correct item according to the child's age				
Failure to record time				
Failure to record correct sequence				
Failure to record incorrect sequence				
Incorrect point assignment (0, 1 or 2)				
Failure to award correct bonus points				
Failure to circle correct bonus points				
Establishing a basal incorrectly				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Arithmetic Errors	**P1**	**P2**	**P3**	**P4**

Failing to begin with the correct item according to the child's age				
Failure to record time				
Failure to record correct response				
Failure to record incorrect response				
Failure to award correct score (0 or 1)				
Failure to award correct bonus points				
Failure to circle correct bonus points				
Establishing a basal incorrectly				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Block Design Errors	**P1**	**P2**	**P3**	**P4**
Failing to begin with the correct item according to the child's age				
Failure to record time				
Establishing a basal incorrectly				
Failing to circle Yes or No				
Failure to award correct number of bonus points				
Failure to circle correct point assignment				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Incorrect point assignment				
Failure to fill in correct designs				
Failure to fill in incorrect designs				
Block Design Errors	**P1**	**P2**	**P3**	**P4**

Total subtest scoring errors				

Vocabulary Errors	P1	P2	P3	P4
Failing to begin with the correct item according to the child's age				
Establishing a basal incorrectly				
Not recording responses verbatim				
Incorrect point assignment on items 1-33:				
1. clock (0, 1, or 2)				
2. hat (0, 1, or 2)				
3. umbrella (0, 1, or 2)				
4. bicycle (0, 1, or 2)				
5. cow (0, 1, or 2)				
6. alphabet (0, 1, or 2)				
7. donkey (0, 1, or 2)				
8. thief (0, 1, or 2)				
9. leave (0, 1, or 2)				
10. brave (0, 1, or 2)				
11. island (0, 1, or 2)				
12. ancient (0, 1, or 2)				
13. nonsense (0, 1, or 2)				
14. absorb (0, 1, or 2)				
15. fable (0, 1, or 2)				
16. precise (0, 1, or 2)				
17. migrate (0, 1, or 2)				
18. mimic (0, 1, or 2)				
19. transparent (0, 1, or 2)				
Vocabulary	**P1**	**P2**	**P3**	**P4**
20. strenuous (0, 1, or 2)				
21. boast (0, 1, or 2)				
22. unanimous (0, 1, or 2)				

23. seclude (0, 1, or 2)				
24. rivalry (0, 1, or 2)				
25. amendment (0, 1, or 2)				
26. compel (0, 1, or 2)				
27. affliction (0, 1, or 2)				
28. imminent (0, 1, or 2)				
29. aberration (0, 1, or 2)				
30. dilatory (0, 1, or 2)				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Total subtest scoring errors				

Object Assembly Errors	**P1**	**P2**	**P3**	**P4**
Failure to record time				
Going overtime				
Failure to record the number of correct junctures				
Multiplying incorrectly				
Failure to circle correct point assignment				
Failure to award correct bonus points				
Failure to add scores correctly				
Total subtest scoring errors				

Comprehension Errors	**P1**	**P2**	**P3**	**P4**
Failure to establish correct basal				
Failure to record responses verbatim				
Incorrect point assignment on items 1-18:				

1. cut finger (0, 1, or 2)				
2. smoke (0, 1, or 2)				
3. seatbelts (0, 1, or 2)				
4. find wallet (0, 1, or 2)				
5. lose ball (0, 1, or 2)				
6. lights (0, 1, or 2)				
7. rules (0, 1, or 2)				
8. fight (0, 1, or 2)				
9. telephone book (0, 1, or 2)				
10. inspect meat (0, 1, or 2)				
11. licence plates (0, 1, or 2)				
12. newspaper (0, 1, or 2)				
13. secret ballot (0, 1, or 2)				
14. stamps (0, 1, or 2)				
15. paperback books (0, 1, or 2)				
16. promise (0, 1, or 2)				
17. senators (0, 1, or 2)				
18. freedom of speech (0, 1, or 2)				
Fail to obtain 2nd response				
1. cut finger				
2. smoke				
6. lights				
7. rules				
11. licence plates				
12. newspaper				
15. paperback books				
17. senators				
Comprehension Errors	**P1**	**P2**	**P3**	**P4**
18. freedom of speech				
Discontinuing subtest incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				

Total subtest scoring errors				

Symbol Search Errors	**P1**	**P2**	**P3**	**P4**
Failure to record the time				
Failure to record the number right/wrong on each page				
Failure to subtract the number incorrect from the number correct				
Failure to record the number correct and incorrect on protocol				
Incorrect point assignment				
Failure to give the sample item				
Failure to give the practice item				
Failure to give the correct part (A or B)				
Total subtest scoring errors				

Digit Span Errors	**P1**	**P2**	**P3**	**P4**
Failure to record correct response				
Failure to record incorrect response				
Discontinuing subtest incorrectly				
Incorrect point assignment (0 or 1)				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Failure to record trial scores				
Failure to record item scores				
Digit Span Errors	**P1**	**P2**	**P3**	**P4**
Total subtest scoring errors				

Mazes	**P1**	**P2**	**P3**	**P4**

Failing to begin with the correct item according to the child's age				
Failure to record time				
Not recording the number of errors				
Failure to award correct bonus points				
Failure to circle correct point assignment				
Failure to give sample item (age 6-7)				
Incorrectly gave sample item				
Establishing a basal incorrectly				
Exceeding discontinuance criteria				
Failure to add scores correctly				
Incorrect point assignment				
Total subtest scoring errors				
Failure to record demographic information				

	P1	P2	P3	P4
SUBTEST TOTAL ERRORS				
Scorer's Initial				