A COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC) AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED (WISC-R) FOR CHILDREN REFERRED TO SCHOOL PSYCHOLOGISTS BECAUSE OF CONCERNS ABOUT THEIR INTELLECTUAL ABILITY

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
MARK EDWARD SWERDLIK

1976

This is to certify that the

thesis entitled

A COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC) AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED (WISC-R) FOR CHILDREN REFERRED TO SCHOOL PSYCHOLOGISTS BECAUSE OF CONCERNS ABOUT THEIR INTELLECTUAL ABILITY

presented by

Mark Edward Swerdlik

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Educational Psychology

Don E. Hamachek
Major professor

Date July 22, 1976

O-7639

~~NOV 11 '78~~

~~NOV 20 '78~~

~~NOV 30 '78~~ 8

~~MAR 29 '79~~ 8

~~APR 12 '79~~

093

271

I 051

ABSTRACT

A COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC) AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED (WISC-R) FOR CHILDREN REFERRED TO SCHOOL PSYCHOLOGISTS BECAUSE OF CONCERNS ABOUT THEIR INTELLECTUAL ABILITY

By

Mark Edward Swerdlik

The Wechsler Intelligence Scale for Children (WISC), originally published in 1949, was the test most often chosen by school psychologists to assess the intelligence of children in the 7-13 age range and to select candidates for special education programs for the educable mentally retarded. Some have called the WISC the best test available that claims to measure intelligence. The WISC was revised 25 years after publication and entitled the Wechsler Intelligence Scale for Children--Revised (WISC-R). No comparative studies of the WISC and WISC-R are reported in the WISC-R manual. However, such a comparison is of practical importance because the WISC-R was designed to replace the WISC.

The essential purpose of this study was to compare scores resulting from the WISC and WISC-R for black, white, and Latino children aged 7 to 15.11 years who had been referred to school psychologists in a midwestern tri-state area because of suspected mental deficiency. Also investigated in the study were various conceptions of test bias as it applies to the WISC-R to determine if, for the

subjects in this study, the WISC-R is more, less, or equally biased compared to the WISC. A survey of participating school psychologists' views of what constitutes a meaningful IQ score difference between the WISC and WISC-R was conducted as part of this study. Further, data regarding how the obtained IQ scores for each test influenced decisions about the educational programming of the subjects involved in this study were also reported.

A total of 78% of the WISC-R items have been taken directly from the WISC, 5.9% are from the WISC but have undergone substantial modification, and 16.1% are new items. Like its predecessor, the WISC-R yields a Verbal, Performance, and Full Scale IQ with a mean of 100 and a standard deviation of 15. Both the Verbal and Performance Scales comprise six subtests, which yield scaled scores with a mean of 10 and a standard deviation of 3. The Full Scale IQ is an average of the Verbal and Performance Scales. Changes between the two tests have been made in terms of administration instructions including questioning, scoring criteria, standardization samples including incorporation of nonwhites in the WISC-R standardization sample, and provision of more statistical data in the WISC-R manual.

All previous studies comparing the WISC and the WISC-R have reported the revised test yielded lower scores. The majority of studies comprised a fairly restrictive sample of special education students, employed designs that did not adequately control for both growth and practice effects, and dealt with small numbers of children. No studies were found that attempted to generalize their results to a population of students referred to school psychologists for suspected mental deficiency, nor did any compare the performance of three different racial groups within a wide age range. However, this is the population with whom the test is most widely used.

In the present study, 72 school psychologists in the tri-state area of Michigan, Illinois, and Ohio administered both the WISC and the WISC-R to 164 children in a counterbalanced order with a specific test-retest interval of not less than a week nor more than a month.

WISC and WISC-R scaled and IQ scores and differences were reported for each of the three major scales and 12 subtests. Significant interactions were also discussed and diagrammed.

The data from this study can be summarized as follows:

1. Subjects obtained significantly higher IQ scores on the WISC than on the WISC-R.

2. WISC Verbal subtests' scaled scores were significantly higher than the WISC-R Verbal subtests' scaled scores.

3. WISC Performance subtests' scaled scores were significantly higher than the WISC-R Performance subtests' scaled scores for all the subtests except Object Assembly.

4. Overall, the differences between the WISC and WISC-R IQ scores were of equal magnitude for younger and older students.

5. A greater difference was found between scaled scores resulting from the WISC and WISC-R for younger than for older students on the Verbal subtests of Information and Arithmetic. The WISC scaled scores were higher for all but the older students on the Arithmetic subtest.

6. For all of the Performance subtests, the difference between WISC and WISC-R scaled scores was of equal magnitude for younger and older students.

7. WISC and WISC-R IQ score differences tended to vary significantly for blacks, whites, and Latinos. In all cases, each of the racial groups scored higher on the WISC than on the WISC-R. These data indicated that the racial IQ discrepancy is widening despite efforts to narrow it. Using the definition of test bias concerning differences among mean IQs of various racial groups, the present study found the WISC-R to be more biased than the WISC. However, those who subscribe to this definition assume that the groups are equal in ability to begin with.

8. There was no significant difference between Verbal-Performance IQ score discrepancies yielded by the WISC and the WISC-R. In all cases, the Performance Scale was higher. Utilizing the conception of test bias that assumes the Performance Scale is less culture loaded and therefore less biased than the Verbal Scale, this finding would lead one to conclude that the WISC-R is neither more nor less biased than the WISC, but is equally biased.

9. Blacks', whites', and Latinos' WISC/WISC-R Verbal subtests' scaled score differences did not vary significantly.

10. Blacks', whites', and Latinos' WISC/WISC-R Performance subtests' scaled score differences did not vary significantly.

11. Obtained WISC/WISC-R differences were not related to any examiner characteristics such as years of experience or training, nor to subject characteristics such as state of residence, size of community, or sex.

12. WISC/WISC-R differences increased as the ability of the students decreased. In all cases, the WISC yielded higher scores.

13. Participating school psychologists looked for a 6-8 point or greater IQ score difference in the 60-90 IQ range before their decisions regarding a particular case would be affected. In the 90-110 IQ score range, the examiners looked for a 9-11 IQ point difference between the WISC and the WISC-R.

14. After testing, the majority of cases included in the present study were enrolled in special education classes for the mentally impaired or learning disabled. For the majority of children who were not enrolled in special education classes, the testing led the school psychologists to make certain recommendations to the teacher.

15. Eighty-six percent of the participating school psychologists indicated the disposition of the case they submitted for the present study would <u>not</u> have changed if <u>only</u> the WISC results had been utilized in the decision-making process.

16. Implications may be drawn from this study for special education programs for the learning disabled (LD) and educable mentally impaired (EMI). If present state criteria are not adjusted, enrollments in LD programs may decline as a result of use of the WISC-R while programs for EMI may increase in number. There also may be fewer special education students integrated into the regular program if WISC-R scores are the major criteria for mainstreaming.

17. It remains necessary for school psychologists to exercise caution to use tests in a fair and sophisticated manner. In addition, criteria in addition to WISC-R scores must be utilized in making special education placement decisions.

A COMPARISON STUDY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN (WISC) AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED (WISC-R) FOR CHILDREN REFERRED TO SCHOOL PSYCHOLOGISTS BECAUSE OF CONCERNS ABOUT THEIR INTELLECTUAL ABILITY

By

Mark Edward Swerdlik

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1976

ACKNOWLEDGMENTS

I feel indebted to a number of individuals who have contributed a great deal in many different ways toward the completion of this dissertation and my doctoral program at Michigan State University.

Drs. Don Hamachek and Harvey Clarizio have served as respected and trusted advisors these past three years. They have always been empathetic, friendly, and cooperative as well as serving as models of professional and scholarly competence. They have demonstrated to me their interest in helping me realize my educational and professional objectives, and I have a great deal of respect for both of them. I hope this in some way expresses my appreciation and respect for them.

Dr. John Schweitzer has served as the director of this dissertation. Although I met John late in my doctoral program when I began this study, he has provided me with a great deal of support and assistance in the completion of this study. His expert knowledge of statistics and measurement was a major contributing factor to its successful completion. Further, he was never too busy to answer questions and provide assistance in any way he could. His empathy and emotional support have been greatly appreciated.

Dr. William Rice has served as my internship supervisor and colleague these past two years. He also was responsible for providing me with invaluable assistance during the initial development of this dissertation proposal. Although not a formal member of the doctoral committee, he has always been willing to take time to discuss this project and provide editorial assistance. His professional competence as a school psychologist has contributed immeasurably to increasing and refining both my clinical and consultation skills as a school psychologist.

Drs. Neal Schmitt and Frank Bruno have also provided me with invaluable assistance in the completion of this study. They both have always provided unlimited time and support.

I also wish to thank the participating school psychologists who provided the data for this study. It involved extra time and effort on their part, and provided additional evidence of their professional commitment to research and providing better service to children. Their help is very much appreciated. In addition, many others such as Dr. Tom Fagan of Western Illinois University; John Austin of the Muskegon Public Schools; Dr. I. L. Zimmerman of Whittier, California; and Dr. John Braccio of the Michigan Department of Education, among others, have provided input into the formulation of this study, helped in collecting the data, supplied references and support for the completion of this study. In addition, Sue Cooley has provided expert typing and editorial assistance in the completion of this manuscript.

Last, I wish to thank several people who have contributed to this study and my years here at Michigan State in ways that cannot adequately be expressed in words. Donna Hobart assisted in the

coding and keypunching of the data. However, more than this, for the past two years I have been fortunate to benefit from her sensitivity, understanding, good humor, and love during many happy and trying times. I wish to thank my parents and brother Jerry, whose constant love, support, and occasional financial assistance have helped me to complete my formal education. But more than this, they have always provided an example of kindness, understanding, and love which I can only hope I will demonstrate all through my professional and personal life.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	ix
Chapter	
I. INTRODUCTION	1
Need for the Study	1
Purpose of the Study	7
Organization of the Study	7
II. A BRIEF, DESCRIPTIVE OVERVIEW OF THE WISC AND AND WISC-R	9
The Tests	9
The Manual and Answer Sheets	10
Standardization Procedures	16
Administration and Scoring	18
Changes Within the Individual Subtests	19
Information Subtest	19
Comprehension Subtest	19
Arithmetic Subtest	20
Similarities Subtest	21
Vocabulary Subtest	22
Digit Span Subtest	23
Picture Completion Subtest	23
Picture Arrangement Subtest	24
Block Design Subtest	25
Object Assembly Subtest	25
Coding Subtest	26
Mazes Subtest	27
Summary	28
III. REVIEW OF THE LITERATURE	29
WISC/WISC-R Comparative Studies	29
Test Bias	45

Chapter Page

IV. METHOD 54

Sample of Examiners 54
Subject Sample 59
Design 62
Testable Hypotheses 63
Analysis 64
Procedure 64
Summary 69

V. RESULTS 70

Findings 70
Supplementary Analyses 83
Summary 90

VI. SUMMARY AND CONCLUSIONS 92

Results 94
Discussion 96
Implications and Recommendations for Test Users . . . 107
Recommendations for Further Research 112

APPENDICES 115

A. LETTERS TO RESPONDENTS AND QUESTIONNAIRE 116

B. ANOVA TABLES 125

BIBLIOGRAPHY 129

LIST OF TABLES

Table		Page
2.1	Organization of the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Intelligence Scale for Children--Revised (WISC-R)	11
2.2	Reliability and Standard Error of Measurement of the WISC Tests (N=200 for Each Age Level)	13
2.3	WISC-R Reliability Coefficients of the Tests and IQ Scales, by Age (N=200 for Each Age Group)	14
2.4	WISC-R Standard Errors of Measurement (SE_m) of the Scaled Scores and IQs, by Age (N=200 for Each Age Group)	15
3.1	Results of Doppelt and Kaufman's Study of Differences Between WISC and WISC-R IQs for Various Ability Levels	32
3.2	Summary of Zimmerman Study	33
3.3	Major Findings of WISC/WISC-R Comparison Studies . . .	43
4.1	Examiner Characteristics	56
4.2	Summary Statistics of Participants by States	58
4.3	Summary Data From Tri-State Area	58
4.4	Subject Characteristics	60
5.1	Comparison of WISC and WISC-R Verbal, Performance, and Full Scale Mean IQ Scores	71
5.2	Comparison of WISC and WISC-R Verbal Subtests' Scaled Score Means	72
5.3	Comparison of WISC and WISC-R Performance Subtests' Scaled Score Means	74
5.4	Comparison of WISC and WISC-R Verbal-Performance Scale Discrepancies	80
5.5	Correlations of WISC and WISC-R Scores	83

Table		Page
5.6	Frequency Counts	85
5.7	Correlation Coefficients of Various Subject and Examiner Characteristics and Obtained WISC/WISC-R Verbal, Performance, and Full Scale IQ Score Differences	87
5.8	Mean WISC/WISC-R Differences for Three Ability Groups .	88
5.9	Summary of Disposition of Cases and Changes if Only the WISC Scores Had Been Utilized by Examiners	89
5.10	Summary of Major Findings	91
B1.	ANOVA Table for Verbal-Performance IQ Score Discrepancies	126
B2.	ANOVA Tables for Verbal Scale	127
B3.	ANOVA Tables for Performance Scale	128

LIST OF FIGURES

Figure		Page
4.1	Design of the First Analysis of Variance and Cell Sizes	65
4.2	Verbal Subscales ANOVA	66
4.3	Performance Subscales ANOVA	67
5.1	Interaction of Age and Information and Arithmetic Subtests as Measured by the WISC and the WISC-R	76
5.2	Interaction of Race and WISC/WISC-R IQ Scores	77
5.3	Interaction of Order of Administration and WISC/WISC-R Scores	81
5.4	Interaction of Order, Test, and Verbal-Performance Scale	82

CHAPTER I

INTRODUCTION

Need for the Study

The use of standardized, individually administered intelligence tests by school psychologists in evaluating children for special education is well documented (Bardon & Bennett, 1975; Sattler, 1975). One particular intelligence test, the Wechsler Intelligence Scale for Children (WISC), has become the test most often chosen for use by school psychologists with children in the 7- to 13-year range (Osborne, 1972) and for placement in special education programs for the educable mentally retarded (Weise, 1960; Silverstein, 1963). Buros (1972) referred to the WISC as the best available test that claims to measure intelligence.

The WISC, originally published in 1949, was revised and published as the Wechsler Intelligence Scale for Children--Revised (WISC-R) 25 years later. The WISC-R test manual cites correlational studies of the WISC-R with the Wechsler Preschool Primary Scale of Intelligence (WPPSI), the Wechsler Adult Intelligence Scale (WAIS), and the Stanford-Binet Form L-M. However, the manual does not report any studies dealing with the obvious and important issue of how the WISC and WISC-R compare. Such a comparison is important, because the WISC-R was designed to replace the WISC. The present study is an attempt to fill this void.

There are many subtle and obvious differences between the WISC and the WISC-R. Four major differences are that the WISC-R standardization sample includes nonwhites and is therefore more representative than the WISC, the WISC-R has new administration and scoring criteria, and its sequence of subtest administration is different. However, the WISC-R, like the old WISC, still yields a Full-Scale, Verbal, and Performance IQ with a mean of 100 and a standard deviation of 15.

The WISC-R appears to have been adopted by most school psychologists for the intellectual assessment of school-age children. Many school psychologists who have had experience administering both tests have observed lower scores on the WISC-R as compared to the original WISC. Carvajal and McKnab (1975), in a survey of more than 70 Kansas school psychologists, reported differences in the neighborhood of 8 to 10 IQ score points with a test-retest interval of 1-2 years. Research on test-retest stability (Quereshi, 1968; Gehman & Matyas, 1956; Whatley & Plant, 1957; Zimmerman & Woo-Sam, 1973) has indicated that differences of 8 to 10 IQ points must indicate something other than measurement error.

This apparent difference between the two tests may be quite crucial, since important educational decisions are made partly on the basis of scores from the WISC and currently the WISC-R. Many state Special Education codes (e.g. those of Michigan and Ohio, among others) require the administration of an individual intelligence test for specific programs. In addition, one criterion for placement is that the candidate must score within a particular

range. For example, in the state of Michigan, one of the criteria for eligibility for the mentally impaired (EMR) program is "development at a rate approximately two to three standard deviations below the mean as determined through intellectual assessment" (Public Act 198--Mandatory Special Education Law). If indeed a WISC/WISC-R score difference exists, with the WISC-R yielding lower scores, this could reflect a number of real-life situations.

One speculation of what might be occurring relates to the possibility that the true mean of the population is higher on the WISC than it is on the WISC-R. For example, if one were to administer the WISC and the WISC-R to the entire population of appropriately aged children, the mean of the WISC would be 100 and the mean of the WISC-R would be 92. Briefly, there are two possible explanations of why this hypothesis may be correct. The first involves sampling error. It is possible that the WISC-R standardization sample may include more bright children. In this particular case, the standardization sample would score higher than the population it was designed to represent, thus artificially raising the norms. This would cause the scores on the WISC-R to be lower and the mean less than 100. Another explanation would entail the scoring of the test protocols of the standardization sample. It is possible that the scorers employed by the Psychological Corporation, publishers of the WISC-R, who scored the protocols of the WISC-R standardization sample, were more lenient than the school psychologists who are currently scoring the test in the field. This condition would also artificially raise the WISC-R norms.

A resulting implication, if this hypothesis is correct, is that the WISC scores are accurately assessing the ability of appropriately aged children and leads to a valid identification of a certain percentage of special education children who score below a particular cut-off score (i.e., 70). However, the WISC-R would be identifying and mislabeling a larger pool of special education youngsters. Those who believe this is what is responsible for the lower scores on the WISC-R are calling for a readjustment of state special education criteria, with a lowering of the IQ-score cutoff.

To determine the potential impact of the new test on the number of people who are eligible for special education, it is necessary to assume that there exists a fixed cut-off point, say 70, that is used to assign students to special education classes. On the WISC this score of 70 is 2 standard deviations below the mean and therefore identifies 2.28% of the population as special education students. If it is further assumed that the WISC-R has a mean seven points lower than the WISC across all age levels, an individual who would have obtained a score of 77 (1-1/2 standard deviations below the mean) on the WISC would now receive a score of 70 on the WISC-R. Thus if the WISC-R is used, all individuals who fall more than 1-1/2 standard deviations below the mean of the WISC (6.30%) would be eligible for special education. This represents an increase of 4% of those eligible for special education. The role of intelligence test scores in the misplacement of youngsters in classes for the mentally impaired is well documented and publicized

in both the literature and the courts (e.g., Prillaman, 1975; Larry P. vs. Wilson Riles).

A second speculation of what might account for the observed difference between the WISC and the WISC-R, which was noted by the Kansas psychologists, is that the WISC-R accurately assesses the student's ability and that the WISC overestimates it. This would assume that if the WISC and the WISC-R were administered to the entire population of appropriately aged children, the mean of the WISC would be 108 and the mean of the WISC-R would be 100. This difference would be attributed to a change in the characteristics of the two populations the WISC (1949) and WISC-R (1974) standardization samples were designed to represent. This interpretation was advanced by Thorndike (1975) and Larabee and Holroyd (1976).

After reviewing the relevant research, Anastasi (1968) concluded that a significant rise in mean intellectual performance occurs when cultural conditions, including increased educational opportunities, improve over time. Owens (1966), Tuddenham (1968), and Wheeler (1942) all consistently reported significant cross-generational increases in mean performance on IQ tests over time spans of 10 to 40 years. It is crucial to this interpretation to understand that if a mean increase in actual ability (ability to answer questions on an IQ test) occurs over time, a later restandardized and renormed test should produce lower IQ scores. This is true of the 1972 norms of the Stanford-Binet Form L-M, which yields lower IQs than the previous 1960 Pinneau norms. These lower Binet

scores have been documented by several researchers, including Zimmerman and Woo-Sam (1975) and Holroyd and Bickely (1976).

This explanation would then predict that children would obtain a higher mean score on the WISC than the WISC-R because of the 25-year interval between restandardizations. One of the implications, if this hypothesis correctly represents what is occurring in real life, is that by using the WISC we have recently been identifying a smaller pool of special education students and by now utilizing the WISC-R we will be correcting this and identifying an appropriate pool of special education youngsters.

If either of the previously discussed speculations represents reality, the current use of the WISC-R is leading to an increased number of children who are eligible for special education. According to one hypothesis, the WISC-R accurately assesses the student's intelligence and compares him meaningfully with his peers. However, if the other hypothesis is true, the WISC-R scores represent an inaccurate estimate of the student's intelligence and therefore misclassify students. Additional evidence relating to these hypotheses is presented in the final chapter of this dissertation.

Because intelligence tests are important and frequently used tools of the practicing school psychologist, they have great implications for children's future educational programs. Hence a study comparing scores resulting from the WISC and WISC-R is most appropriate at this time.

Purpose of the Study

The purpose of this study is to compare scores resulting from the WISC and the WISC-R for children referred to school psychologists because of suspected mental deficiency. It also examines how the tests have affected decisions regarding the educational programming of these children.

The five major purposes of the study are: (a) to investigate whether there is a difference between scores resulting from the WISC and the WISC-R and if the difference is the same for different ages and races, (b) to determine whether these differences or lack of them is related to the training and/or experience of the examiners (school psychologists) and demographic variables of the school setting, (c) to investigate how these scores affect the educational programming of students and if using scores resulting from the WISC or the WISC-R would lead to different decisions, (d) to assess the opinions of school psychologists concerning what constitute meaningful IQ score differences between scores resulting from the WISC and the WISC-R, and (e) to investigate in part whether the WISC-R is less culturally biased than the WISC.

Organization of the Study

In Chapter II, a brief, descriptive overview of the WISC and the WISC-R tests is presented. Chapter III contains a review of the pertinent literature, including a discussion of various definitions of cultural bias and previous research comparing the WISC and the WISC-R. Detailed in Chapter IV is the design of the study.

In Chapter V the results are presented, and Chapter VI includes a discussion of these results and a summary of the study.

CHAPTER II

A BRIEF, DESCRIPTIVE OVERVIEW OF THE WISC AND THE WISC-R

The WISC (1949) was renormed and revised 25 years after publication and entitled the WISC-R (1974). The WISC-R includes many improvements over its predecessor, the original WISC (Wechsler, 1974). The major improvements in the test include an increase in the number of items that compose each of the subtests to enhance reliability, omission or revision of items believed to be out of date or culturally biased, inclusion of nonwhites in the standardization sample, updating of norms, and clarification of administration and scoring criteria.

A discussion of these improvements, along with other changes in the WISC-R, will facilitate an understanding of the comparability of scores resulting from the two tests. Following is a discussion of these improvements and the differences between the WISC and the WISC-R.

The Tests

The WISC was appropriate for ages 5-15, whereas the WISC-R is administered to children and adolescents from the ages of 6 to 16.

Larrabee and Holroyd (1976) reported that 78% of the WISC-R items are taken directly from the WISC; 5.9% are from the WISC

but have undergone substantial modification and 16.1% of the WISC-R items are newly developed.

Both tests yield Verbal, Performance, and Full Scale IQ composite scores with a mean of 100 and a standard deviation of 15. The Verbal and Performance subscales consist of 12 subtests (see Table 2.1), each with a scaled score mean of 10 and a standard deviation of 3. However, on the WISC, the supplementary tests of Digit Span and Mazes, if administered, are averaged into the Verbal and Performance IQ composite scores; this is not done on the WISC-R.

WISC IQ scores range from 45-155 for the Verbal IQ, 44-156 for the Performance IQ, and 46-154 for the Full Scale IQ. WISC-R IQs range from 45-155 for the Verbal and Performance IQs and from 40-160 for the Full Scale IQ. Scaled score ranges for the 12 subtests are 0-20 for the WISC and 1-19 for the WISC-R (Wechsler, 1949; 1974).

The Manual and Answer Sheets

The WISC-R answer sheets are simplified and much improved over the WISC but lack space for recording responses on the Comprehension subtest.

Many reviewers (Carvajal & McKnab, 1975; Tittle, 1975; Kirchev, 1975) have claimed the WISC-R manual is a major improvement over its unclear and less complete predecessor, the WISC. Reviewers have evaluated the WISC-R manual as being more complete and readable. Its organization and content facilitate test administration and scoring by providing clearer and more complete test administration instructions and scoring criteria.

Table 2.1: Organization of the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Intelligence Scale for Children--Revised (WISC-R).

	WISC	WISC-R
VERBAL SUBTESTS YIELD VERBAL IQ	Information Comprehension Similarities Arithmetic Vocabulary Digit Span (optional)	Information Comprehension Similarities Arithmetic Vocabulary Digit Span (optional, but even if administered, not averaged into Verbal IQ)
PERFORMANCE SUBTESTS YIELD PERFORMANCE IQ	Picture Completion Picture Arrangement Block Design Object Assembly Coding Mazes (optional)	Picture Completion Picture Arrangement Block Design Object Assembly Coding Mazes (optional, but even if administered, not averaged into Performance IQ)

Full Scale IQ = Average of both Verbal and Performance subscales

Following is a discussion of some of the additional data that have led reviewers to conclude that the WISC-R manual is a major improvement over the WISC manual.

For the WISC, split-half reliability coefficients were reported for pupils aged 7-1/2, 10-1/2, and 13-1/2. For the younger ages, these reliability coefficients were lower for each of the three major scales and all 12 subtests. The WISC-R manual presents more complete data, reporting split-half reliabilities for each of 11 age groups in the standardization sample and by the test-retest method for three of the age groups. The reliability coefficients for the WISC-R are generally higher than the WISC, especially at the younger age levels. The reliability coefficients for both the WISC and WISC-R are presented in Tables 2.2 and 2.3.

The standard error of measurements for the WISC were reported for three age levels--7-1/2, 10-1/2, and 13-1/2. For the WISC-R, the SEMs are reported for all 11 age levels, 6-1/2 to 16-1/2 (Tittle, 1975; Carvajal & McKnab, 1975). The standard error of measurements are reported for the WISC in Table 2.2 and the WISC-R in Table 2.4.

No validity data have been reported for the WISC in the manual. Studies comparing the WISC-R with the WPPSI, WAIS, and Stanford-Binet Form L-M (1972 norms) have reported that WISC-R Full Scale IQs correlated at .82, .95, and .73 with the WPPSI, WAIS, and Stanford-Binet Form L-M, respectively.

The WISC provided intercorrelations of subtests for ages 7-1/2, 10-1/2, and 13-1/2. The WISC-R manual contains these data for all ages 6-1/2 to 16-1/2 (Carvajal & McKnab, 1975).

Table 2.2.--Reliability and standard error of measurement[a] of the WISC tests (N = 200 for each age level).

	Age 7½		Age 10½		Age 13½	
	r	SE_m	r	SE_m	r	SE_m
Information	.66	1.75	.80	1.34	.82	1.27
Comprehension	.59	1.92	.73	1.56	.71	1.62
Arithmetic	.63	1.82	.84	1.20	.77	1.44
Similarities	.66	1.75	.81	1.31	.79	1.37
Vocabulary	.77	1.44	.91	.90	.90	.95
Digit Span	.60	1.90	.59	1.92	.50	2.12
Verbal Score (without Digit Span)	.88	5.19	.96	3.00	.96	3.00
Picture Completion	.59	1.92	.66	1.75	.68	1.70
Picture Arrangement	.72	1.59	.71	1.62	.72	1.59
Block Design	.84	1.20	.87	1.08	.88	1.04
Object Assembly	.63	1.82	.63	1.82	.71	1.62
Coding**	.60	1.90	—	—	—	—
Mazes	.79	1.37	.81	1.31	.75	1.50
Performance Score (without Coding and Mazes)	.86	5.61	.89	4.98	.90	4.74
Full Scale Score (without Digit Span, Coding and Mazes)	.92	4.25	.95	3.36	.94	3.68

[a]The SE_m is in Scaled Score units for the tests and in IQ units for the Verbal, Performance, and Full Scale scores.

[b]Based on correlating Coding A and Coding B, 115 cases. For age 8-1/2 the value is .56 for 91 cases.

Table 2.3.--WISC-R reliability coefficients of the tests and IQ scales, by age (N = 200 for each age group).

Test	Age Group											Average
	6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½	r_{11}[a]
Information	.67	.80	.80	.81	.83	.88	.87	.87	.88	.90	.89	**.85**
Similarities	.87	.85	.79	.79	.79	.81	.84	.79	.81	.74	.83	**.81**
Arithmetic	.79	.75	.69	.80	.76	.81	.80	.81	.73	.80	.75	**.77**
Vocabulary	.74	.70	.86	.86	.84	.86	.88	.89	.91	.90	.92	**.86**
Comprehension	.69	.70	.73	.78	.71	.83	.87	.81	.82	.72	.78	**.77**
Digit Span	.76	.84	—[b]	—[b]	.71	.75	—[b]	—[b]	.79	.79	—[b]	**.78**
Picture Completion	.84	.81	.85	.78	.68	.80	.75	.75	.72	.68	.75	**.77**
Picture Arrangement	.77	.72	.69	.76	.72	.73	.78	.72	.74	.73	.70	**.73**
Block Design	.80	.82	.85	.80	.86	.89	.86	.86	.84	.85	.90	**.85**
Object Assembly	.76	.73	.66	.70	.64	.72	.63	.72	.72	.68	.71	**.70**
Coding	.63	.63	—[b]	—[b]	.76	.79	—[b]	—[b]	.65	.80	—[b]	**.72**
Mazes	.82	.81	.77	.71	.66	.75	.62	.65	.72	.65	.57	**.72**
Verbal IQ	.91	.92	.92	.94	.93	.95	.96	.95	.95	.94	.95	**.94**
Performance IQ	.91	.90	.91	.91	.89	.91	.91	.90	.89	.90	.91	**.90**
Full Scale IQ	.95	.95	.95	.96	.95	.96	.96	.96	.96	.95	.96	**.96**

Table 2.4.--WISC-R standard errors of measurement (SE_m) of the scaled scores and IQs, by age (N = 200 for each age group).

Test	Age Group											Average
	6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½	SE_M
Information	1.67	1.35	1.26	1.34	1.09	1.00	1.06	1.14	1.08	.93	1.12	**1.19**
Similarities	1.14	1.30	1.41	1.45	1.48	1.37	1.14	1.34	1.34	1.50	1.28	**1.34**
Arithmetic	1.33	1.41	1.48	1.28	1.38	1.29	1.38	1.30	1.45	1.25	1.59	**1.38**
Vocabulary	1.55	1.55	1.21	1.21	1.19	1.18	1.02	1.04	.96	.90	.87	**1.15**
Comprehension	1.59	1.61	1.48	1.40	1.43	1.21	1.08	1.27	1.22	1.52	1.51	**1.39**
Digit Span	1.52	1.15	—[a]	—[a]	1.63	1.49	—[a]	—[a]	1.41	1.41	—[a]	**1.44**
Picture Completion	1.23	1.24	1.13	1.36	1.59	1.37	1.50	1.61	1.54	1.89	1.50	**1.45**
Picture Arrangement	1.55	1.73	1.61	1.50	1.55	1.50	1.51	1.59	1.54	1.59	1.61	**1.57**
Block Design	1.38	1.27	1.20	1.31	1.12	1.08	1.13	1.14	1.14	1.09	.99	**1.17**
Object Assembly	1.58	1.68	1.70	1.59	1.71	1.67	1.84	1.71	1.68	1.82	1.74	**1.70**
Coding	1.89	1.82	—[a]	—[a]	1.46	1.38	—[a]	—[a]	1.79	1.44	—[a]	**1.63**
Mazes	1.29	1.38	1.45	1.63	1.81	1.62	1.97	1.98	1.68	1.83	2.08	**1.70**
Verbal IQ	4.08	4.02	3.86	3.69	3.65	3.34	3.13	3.42	3.40	3.42	3.57	**3.60**
Performance IQ	4.75	4.80	4.48	4.46	4.65	4.39	4.58	4.96	4.74	4.84	4.60	**4.66**
Full Scale IQ	3.41	3.39	3.23	3.14	3.21	2.98	2.96	3.23	3.15	3.19	3.16	**3.19**

Standardization Procedures

The WISC was standardized on a 1940 population of 2,200 whites, 100 of each sex in 11 age groups, with adjustment for western movement of the population. Nine categories of parental occupations were condensed from 14 included in the 1940 Census (Wechsler, 1949; Carvajal & McKnab, 1975). The absence of nonwhites in the WISC standardization sample has been one of the major criticisms of the WISC (Littell, 1960).

The WISC-R was standardized on a stratified sample of 2,200 children, 100 of each sex in each of 11 age groups. In each age range nonwhites were included in the same proportion as they existed in the 1970 Census. Five categories of parental occupations were condensed from 10 in the 1970 Census (Carvajal & McKnab, 1975; Wechsler, 1974; Kirchev, 1975).

Field supervisors in different regions of the country selected the individual children who composed the standardization samples of both the WISC and the WISC-R. These supervisors selected the children of different ages, races, and sexes under guidelines from the Psychological Corporation, publishers of the WISC and WISC-R. In many cases, the field supervisors hired other individuals to administer the tests.

Kaufman and Doppelt (1976) analyzed the Verbal, Performance, and Full Scale IQs of the WISC-R standardization sample of 2,200 children according to the variables of sex, race, geographic region, parental occupation, and type of residence (urban vs. rural). They then compared results with similar data obtained in a study of the

WISC standardization sample reported in 1950. For both the WISC and the WISC-R, the researchers found a relationship of IQ to sex and parental occupation. However, the differences in IQ for boys and girls were not related to age for the WISC-R standardization sample, as they were for the WISC. The average IQs of whites were approximately 1 standard deviation (15 points) higher than those of blacks. The relationship between IQ and type of residence changed for the WISC-R standardization sample, because the gap had closed between rural and urban areas; the authors explained this as being the result of mass media and improved educational opportunities in rural areas. The authors concluded that the WISC-R IQ differences between the sexes and between children who resided in rural or urban areas were too small to be meaningful. However, the IQ differences among the various parental occupation groups and between races (blacks and whites) were of enough magnitude to be considered meaningful. The authors also concluded that the results of their study of the WISC-R standardization sample were quite similar to those obtained in studies of the WISC standardization sample and supported the comparability of the tests.

Kaufman (1975) factor analyzed the WISC-R at 11 age levels between 6-1/2 and 16-1/2 years using the WISC-R standardization sample as his sample. He concluded that the factors on the WISC-R (Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility) resembled the factors identified for the WISC.

Administration and Scoring

The WISC manual indicated that Verbal and Performance subtests may be intermixed; but both the manual and score sheets listed all Verbal subtests followed by all Performance subtests, thereby not implementing this suggestion. The WISC-R manual dictates that the Verbal and Performance subtests must be alternated, and complies with this requirement in both the manual and score sheets.

The general rule regarding questioning on the WISC was to question marginal zero responses if the main idea was presented. The examiner was also allowed to question one-point responses if the item was followed by a (Q) in the manual, but few such cases were listed. Probing and testing the limits were not allowed for determination of scores; only neutral questions were allowed. The WISC-R provides clearer rules regarding spoilage of responses and when to question. As was true for the WISC, only neutral questions are permitted unless otherwise specified.

The WISC-R requires at least three Verbal and three Performance subtests to have raw scores above zero to permit calculation of the Verbal and Performance IQs, respectively. In addition, to permit calculation of the Full Scale IQ, there must be at least three Verbal and three Performance subtests combined that have raw scores above zero (Carvajal & McKnab, 1975).

The WISC manual provided examples of two-point responses, which were generally the poorest answers that could be given two points. Most of the zero-point responses were marginal and many were to be questioned. This is similar to the WISC-R, except that

in the latter, many one-point responses are also considered marginal and are followed by a (Q), which means that they should be further questioned in a neutral fashion. Generally, the scoring criteria and examples for the Vocabulary, Similarities, and Comprehension subtests are much improved, in terms of clarity, on the WISC-R test (Carajal & McKnab, 1975; Kirchev, 1975).

Changes Within the Individual Subtests

Information Subtest

Eleven of the original 30 items on the WISC Information subtest have been significantly modified or replaced on the WISC-R; outdated or culturally unfair items have been replaced. The WISC had two beginning points for different age subjects, whereas the WISC-R has four. The examiner did not explain any missed items on the WISC, but on the WISC-R he is instructed to explain item 1 if it is missed. On the WISC, items 4, 5, and 6 needed to be correct to assume credit for items 1 through 3. On the WISC-R, if the subject earns a perfect score on the first two items administered, credit is given for all preceding items not administered. If he does not earn a perfect score on the first two items, the examiner administers items in reverse order until the subject answers two consecutive questions correctly, not including the beginning item (Kirchev, 1975; Carvajal & McKnab, 1975).

Comprehension Subtest

The WISC-R contains three more Comprehension items than the WISC; eight items of the WISC-R are new. Regarding test administration,

the WISC examiner was not instructed to ask for an additional response if the subject gave only one answer and two were required for a perfect score. The WISC-R has similar scoring for many items, but the examiner is required to ask for an additional correct response if the subject originally provides only one. The WISC manual contained few examples, whereas many scoring sample items are provided in the WISC-R manual. No missed items were explained on the WISC; however, the WISC-R examiner explains item 1 if the subject provides less than a two-point response (Kirchev, 1975; Carvajal & McKnab, 1975).

Arithmetic Subtest

WISC Arithmetic subtest materials included blocks for counting, but there were no directions regarding spacing of the blocks. Story problems were included, but no specific directions were given concerning what to do if a subject had difficulty reading. The WISC-R provides trees printed on a card for counting, and highly specific directions; it too has story problems, and the examiner can assist the subject if he has difficulty reading. The WISC had two starting points, whereas the WISC-R has four. There are two more items on the WISC-R than on the WISC. The number of times the examiner could repeat an item was not limited on the WISC, whereas the WISC-R specifies that an item be repeated only once. No items could be explained on the WISC, but the WISC-R examiner is allowed to explain item 1 if it is missed and may define what is meant by "cover up" on items 2 and 3 if the subject fails to comprehend the instructions. On the WISC, if either item 4 or 5 was answered

correctly, the subject was given credit for items 1, 2, and 3. The WISC-R, however, requires that a perfect score be achieved on the first two items administered, to assume credit for all previous items not administered. Otherwise, the examiner goes in reverse order until the subject has two consecutive items correct, not including the beginning item. Credit is assumed for all earlier items not administered. The exception is items 5 and 6, for which the examiner must go back to item 1, because they deal with the tree card. More child-oriented items and modern price and wage standards compose the WISC-R story problems as compared to those of the WISC (Kirchev, 1975; Carvajal & McKnab, 1975).

<u>Similarities Subtest</u>

Seventeen items make up the WISC-R Similarities subtest, as compared to 16 on the WISC. The four analogy items on the WISC have been omitted on the WISC-R; several WISC items have also been replaced or modified on the revised instrument. The WISC included two starting points for two different ages, whereas the WISC-R has only one. On the WISC-R, the examiner is allowed to clarify the question on item 1, adding "how are they the same"; this was not allowed on the WISC. On the WISC, item 5 was explained if the subject scored a zero; item 6 was explained if the subject scored zeroes on items 5 and 6. The WISC-R examiner explains items 1 and 2 if they are missed. If the subject gives a one-point response to item 5 and/or 6, the examiner explains. Only marginal zero responses were questioned on the WISC, but many one-point as well as marginal

zero responses are questioned on the WISC-R (Kirchev, 1975; Carvajal & McKnab, 1975).

Vocabulary Subtest

The WISC Vocabulary subtest comprised 40 items, many of which were judged to be more difficult than the WAIS items. The WISC-R comprises 32 items, with 11 new ones felt to be more appropriate than former WISC items. The instrument has four starting points, compared to two on the WISC. Out-of-date and possible slang words have been eliminated on the WISC-R, and more parts of speech are included. All words are scored 2, 1, or 0 on the WISC-R; on the WISC items 1-5 were scored 2 or 0 and items 6-40 were scored 2, 1, or 0. If two one-point responses were given on the WISC, the score remained 1, whereas on the WISC-R such responses are scored as two points. No items were explained on the WISC, whereas the WISC-R instructs the examiner to explain item 1 if the subject gives less than a two-point response. If the subject does not hear a word correctly, the WISC-R instructs the examiner to say, "Listen carefully. What does ________ mean?" The WISC did not include this possibility.

On the WISC, if the subject began with number 10, he must have made five consecutive two-point responses. Otherwise, the examiner went in reverse order until five consecutive two-point responses were given. The examiner assumed credit for those items not administered. The WISC-R requires that the subject earn perfect scores on the first two items administered in order to assume credit for all preceding items not administered. Otherwise, the examiner goes in reverse order until the subject earns two consecutive

two-point responses, not including the beginning item. The WISC-R examiner assumes credit for all preceding items not administered (Kirchev, 1975; Carvajal & McKnab, 1975).

Digit Span Subtest

The items on the Digit Span subtest are unchanged on the WISC-R, although both trails of each item are administered even if the child passes the first trail. On the WISC, the second trail was administered only if the subject failed trail one. The WISC directions did not state whether the examiner should drop his voice on the last digit; this question remains unanswered on the WISC-R. Scoring on the WISC was based on the highest number of digits repeated successfully, whereas the WISC-R scoring is based on the total number of trails passed. On the backward digits, the WISC started with the three-digit item if the subject passed either of the sample items; the WISC-R, however, has the subject start with item 1 (two-digit series), whether or not he succeeded or failed on the sample(s) (Kirchev, 1975; Carvajal & McKnab, 1975).

Picture Completion Subtest

Six WISC items have been eliminated on the WISC-R Picture Completion subtest, and it has been lengthened by six items. The WISC instrument had one starting point for all ages; the WISC-R has two starting points. The time limit for each item has been changed from 15 seconds on the WISC to 20 seconds on the WISC-R. Assuming credit for unadministered items was unnecessary on the WISC; however, credit is assumed for items 1-4 if the subject responds

correctly to items 5 and 6 on the WISC-R. WISC-R test reliability is greater at the younger age levels than was the WISC. Also, the WISC-R contains more items picturing blacks and women than did the former test. Whereas the WISC instructed the examiner to introduce each item by saying, "Now what is missing in this one?" the WISC-R drops or shortens the instructions if the subject understands the task. Regarding inquiry, more specific instructions are provided on the WISC-R and the examiner may say, "Show me where you mean" rather than only the neutral questioning that was allowed on the WISC (Kirchev, 1975; Carvajal & McKnab, 1975).

Picture Arrangement Subtest

Sample items differed on the WISC Picture Arrangement Subtest, with separate ones provided for ages 5-7 and 8-15; the WISC-R gives one sample item for all ages. On the WISC, the subject had to pass items 1 or 2 to receive credit for items A-D, whereas on the WISC-R the subject must pass the first trail of item 3 to receive credit for items 1 and 2.

All items on the WISC-R have the same format. On this test, two WISC items have been eliminated, four shortened by one card, and several redrawn. Most of the five new items on the WISC-R replace the cut-up pieces used at the younger age levels on the WISC. The WISC-R examiner is also allowed to encourage the child to work faster to earn the time bonus points (this is also true of Block Design and Object Assembly); this was not specifically covered in the WISC manual (Kirchev, 1975; Carvajal & McKnab, 1975).

Block Design Subtest

The WISC Block Design subtest used blocks with six color combinations; the WISC-R utilizes red, white, and red-white combinations. On the WISC, the examiner had to introduce every design by saying, "Now make one like this." The WISC-R instructions may be shortened if the subject understands the task. On the WISC, if the subject passed either trail of item C, he was given credit for items A and B. In contrast, to earn credit for items 1 and 2 on the WISC-R, the subject must pass the first trail of item 3. Reversals on items A, B, and C could be explained on the WISC, whereas the WISC-R examiner may show the subject the correct arrangement only once if the subject has rotated any design. No examples of rotations were given in the WISC scoring manual, but the WISC-R provides five examples plus written instructions. Some new items and modifications in time bonuses are included in the WISC-R (Kirchev, 1975; Carvajal & McKnab, 1975).

Object Assembly Subtest

No sample items were provided on the Object Assembly subtest of the WISC, whereas the WISC-R gives one sample item for all ages. Although the WISC manual did not state whether to remove the shield before or after the directions were given, the WISC-R instructs the examiner to arrange the puzzle pieces behind a shield, then remove it, state the directions, and begin timing. The WISC examiner did not explain any missed items, but the WISC-R examiner shows the correct response if the subject fails the first item. The scoring is somewhat different between the WISC and WISC-R; points for

various puzzle arrangements differ, depending on the year in which the WISC manual was published. It most frequently instructed the examiner to score one point for each correct juxtaposition on the manikin, horse, and auto items; the face item was scored one-half point for each correct juxtaposition. The WISC-R scores one point for each cut on the girl and horse items and one-half point for each cut on the auto and face. Regarding the items more specifically, the WISC-R auto style has been updated and the manikin has been changed slightly to a little girl (Kirchev, 1975; Carvajal & McKnab, 1975).

Coding Subtest

The WISC Coding subtest did not instruct the subject to use any special pencils. The WISC-R requires the subject to use a red pencil without any eraser. The permissibility of praise was not dealt with in the WISC instructions; on the other hand, the WISC-R instructs the examiner to praise the subject for each sample item success. The WISC-R examiner instructs the subject to go from line to line, continuing to work until time has expired. The WISC Coding directions were brief, compared to the more complex directions given on the WISC-R. Whereas only brief scoring directions were given in the WISC manual, the WISC-R is more complete, instructing the examiner to score as correct any figure that is identifiable. The WISC-R record book is separate from the Coding answer sheet, and is printed in two colors, as compared to the monocolor WISC Coding answer sheet, which was printed on the back of the record book (Kirchev, 1975; Carvajal & McKnab, 1975).

Mazes Subtest

On the WISC-R Mazes subtest, a more difficult item has been added and a boy or girl is printed in the center of each maze. As was true with the Coding subtest, the WISC required no special pencils, whereas the WISC-R requires that the subject use a red pencil without an eraser. Concerning lifting the pencil, the WISC manual stated that the examiner must, as often as necessary, inform subjects under 8 years old to keep their pencil points on the paper. The WISC instructions for subjects older than age 8 did not say anything about this, except in the initial instructions for maze C. The WISC-R manual instructs the examiner to remind all subjects, as often as necessary, to keep their pencil points on the record form.

The WISC Maze subtest directions were very brief, whereas the WISC-R general directions are more detailed. The WISC-R manual lists specific statements that cover six cautions to the subject if he encounters specific difficulties. On the WISC, credit was given for items A and B if C was accomplished with not more than one error. On the WISC-R, however, the subject must complete item 4 correctly to receive credit for items 1 through 3. Errors on the WISC consisted in entering a blind alley, crossing a line, or lifting the pencil; no scoring examples were provided. Scoring criteria for the WISC-R Mazes subtest are different, in that the only type of error scored is entering into any blind alley. In addition, 14 scoring examples are provided. A maximum of three points could be given on some WISC mazes, but a maximum of five is allowed on some WISC-R items. The WISC-R record book also provides a chart with the maze

number, number of errors, and number of points listed, which facilitates scoring. As was true of the WISC-R Coding subtest, the answer sheet is separate from the record book (Kirchev, 1975; Carvajal & McKnab, 1975).

Summary

This chapter presented a brief overview of the changes that have been made between the WISC and the WISC-R tests. The WISC-R possesses the same format as the WISC, with identical subtests yielding Verbal, Performance, and Full Scale IQ scores. Major changes on the WISC-R include more complete presentation of statistical data and clearer scoring and administration instructions in the manual, inclusion of nonwhites in the WISC-R standardization sample, and updated norms. Seventy-eight percent of the items on the WISC-R are taken directly from the WISC, 5.9% are taken from the WISC but have undergone substantial modification, and 16.1% of the items are new. Many of the items that were either modified or replaced include those that were judged to be out of date or culturally biased.

CHAPTER III

REVIEW OF THE LITERATURE

In this chapter the literature pertinent to the present study is reviewed. Examined first are previous studies that have compared the WISC and other standardized intelligence tests--more specifically, WISC/WISC-R comparative studies. Then the issue of test bias, including a discussion of various definitions and studies related to this issue, is examined.

WISC/WISC-R Comparative Studies

One of the common ways researchers attempt to determine the usefulness of a recently developed instrument is by comparing it to older, more established tests. This is also true in the case of the WISC-R. Studies (Wechsler, 1974) have indicated the WISC-R has validity similar to the WISC. These studies have shown correlations of .82 with the Wechsler Preschool Primary Scale of Intelligence (WPPSI), .95 with the Wechsler Adult Intelligence Scale (WAIS), and .73 with the Stanford-Binet Form L-M (1972 norms) using the WISC-R Full Scale IQ score.

Doppelt and Kaufman (in press) identified those WISC items that remained substantially unchanged and were administered in the same manner on the WISC-R (Verbal Scale: Information--19 items, Arithmetic--8 items, and Vocabulary--21 items; Performance Scale: Object Assembly--2 items, Coding--entire test, and Mazes--8 items).

The authors developed regression equations in an attempt to answer the question: If the WISC-R standardization sample had been tested with the WISC, what would their IQs have been? Those items that could be scored differently in the WISC and WISC-R were scored according to the 1974 WISC-R scoring criteria. Doppelt and Kaufman described their procedure for estimating WISC IQs for the WISC-R sample as follows:

> Scores on the three sets of items that constituted the Verbal common core were used in a multiple regression equation to predict the WISC Verbal IQ sum of scaled scores of the children in the 1949 standardization sample. This was done separately for each age group. Corresponding equations were developed to predict the Performance core items. The coefficients of multiple correlation for the Verbal and Performance scales are provided. The coefficients for the Full Scale were computed by correlating the estimated Full Scale Score (the sum of the estimates for the Verbal and Performance scales) with the actual WISC Full Scale Score. The regression equations that were obtained from the analysis of the WISC standardization data were applied to the core scores of the WISC-R standardization sample to obtain estimated WISC scores for those children. Correlations between the estimated WISC scores and the obtained WISC-R scores for the 1974 sample are [shown along with] the coefficients between estimated WISC and actual WISC scores for the 1949 sample. It is apparent that the estimated WISC IQs . . . correlate with the actual WISC IQs of the 1949 sample to about the same extent as they do with the WISC-R IQs of the 1974 sample (p. 4).

Doppelt and Kaufman's statistical analysis predicted for the age range 6.5 to 15.5 the mean WISC IQ would be higher by 1.5 points on the Verbal Scale, 6 points on the Performance Scale, and 4 points on the Full Scale. At age 11, the differences between the two tests were much smaller. Older subjects were predicted to show less difference between the two tests than younger subjects. (Barclay and Carolan [1966] termed this the "specific age effect.") The study also predicted a more marked difference for lower ability

groups (this is termed a specific ability effect, and was discussed by Hannon and Kicklighter [1970]). Table 3.1 gives a more detailed presentation of the findings. In summary, Doppelt and Kaufman (in press) found significant score differences between the WISC and WISC-R tests, with the WISC-R being lower in all cases. They also predicted greater IQ score differences between the two tests for younger and lower-ability students.

Limitations of the Doppelt and Kaufman (in press) statistical prediction study include the lack of actual cases, the use of only a small number of common items, and the fact that for items that could be scored differently by the WISC and WISC-R, the WISC-R criteria were subjectively decided upon. In addition, the study overlooked crucial differences between the two tests, including order of subtest administration, different administration instructions, item modifications, and changes in scoring criteria, among others. However, their study did raise questions about the comparability of scores on the WISC and WISC-R.

Zimmerman (1975) analyzed 86 cases of educationally handicapped and educationally mentally retarded (EMR) children in California to whom the WISC-R had been administered within the past year and the WISC some time previously. Zimmerman found differences between the two tests, with the WISC-R being lower in terms of IQ score in all cases. In addition, Zimmerman found these differences to be more marked for educationally handicapped (a combination of learning disabled and emotionally impaired students) and younger subjects than for EMR and older students. For the

Table 3.1.--Results of Doppelt and Kaufman's study of differences between WISC and WISC-R IQ's for various ability levels.[a]

WISC-R IQ	Classification	Difference (WISC IQ Minus WISC-R IQ)		
		Verbal	Performance	Full Scale
Ages 6.5-10.5				
130 and above	Very Superior	+3	+5	+4
120-129	Superior	+3	+5	+4
110-119	High Average (Bright)	+4	+6	+5
90-109	Average	+4	+7	+6
80-89	Low Average (Dull)	+5	+8	+7
70-79	Borderline	+6	+9	+7
69 and below	Mentally Deficient	+6	+9	+8
	Total	+4	+7	+6
Ages 11.5-15.5				
130 and above	Very Superior	-2	+6	+2
120-129	Superior	-1	+6	+2
110-119	High Average (Bright)	-1	+6	+2
90-109	Average	-1	+6	+3
80-89	Low Average (Dull)	-1	+6	+3
70-79	Borderline	0	+6	+3
69 and below	Mentally Deficient	0	+6	+3
	Total	-1	+6	+2

[a]A positive sign means the WISC IQ is higher.

educationally handicapped children, major subscale differences between the WISC and WISC-R were 4.9 for the Verbal Scale, 3.0 for the Performance Scale, and 4.1 for the Full Scale IQ. For the EMR sample, Zimmerman reported WISC/WISC-R differences of 3.3 for the Verbal Scale, 2.2 for the Performance IQ, and 2.1 for the Full Scale IQ. In all cases, the WISC-R results were lower than the previous WISC results. A summary of these results is presented in Table 3.2.

Table 3.2.--Summary of Zimmerman (1975) study.

	WISC	WISC-R	Difference
Young Educationally Handicapped			
Verbal IQ	90.9	84.5	6.4
Performance IQ	89.2	86.7	2.5
Full Scale IQ	89.3	84.2	5.1
Older Educationally Handicapped			
Verbal IQ	81.9	79.3	2.6
Performance IQ	90.2	87.0	3.2
Full Scale IQ	84.4	81.8	2.6
Young Educationally Mentally Retarded			
Verbal IQ	67.2	62.9	4.3
Performance IQ	68.7	63.0	5.7
Full Scale IQ	64.9	59.8	5.1
Older Educationally Mentally Retarded			
Verbal IQ	64.0	61.0	3.0
Performance IQ	66.1	65.5	0.6
Full Scale IQ	61.6	60.5	1.1

Limitations of the Zimmerman (1975) study are the lack of control for both order of administration and growth effects. The researcher provided no test-retest summary information such as ranges, means, or medians, nor did she impose any test-retest interval limits. It is most likely that the two tests were administered many years apart. In addition, Zimmerman utilized a sample of special education students who, in many cases, had been enrolled in special education for some time. This also affects the results and limits their generalizability.

Swerdlik and Rice (1975) found results similar to Zimmerman's in their analysis of 41 EMR and non-EMR children who had been administered the WISC-R within the past year and the WISC from one to four years previously. Children identified as emotionally impaired or learning disabled were eliminated from the sample. The researchers reported significant mean WISC/WISC-R differences of 3.80 for the Verbal Scale, 2.74 for the Performance Scale, 3.05 for the Full Scale IQ, and 1.31 for the Vocabulary subtest. In all cases, the WISC-R yielded lower scores than the WISC. The mean WISC/WISC-R difference for the Comprehension subtest was not significant. In addition, no specific age or ability effects were noted.

Deficiencies of the Swerdlik and Rice (1975) study are similar to those of the Zimmerman research. They include a lack of control for order of administration and growth effects, although a specific test-retest time limit was imposed. In addition, special

education students made up a large proportion of the sample, thereby limiting generalizability of the results.

Hamm et al. (1975) estimated WISC/WISC-R differences for 48 10- and 13-year-old special education students enrolled in EMR classes in rural southeast Georgia. They found significant overall WISC/WISC-R differences of 6 points for the Verbal IQ, 10 points for the Performance IQ, and 7.5 points for the Full Scale IQ. In all cases the WISC-R yielded lower scores. The authors reported no significant age effects.

The study attempted to improve upon previous studies by utilizing a semi-counterbalanced design with a specific test-retest interval of not less than 14 days and a mean of 39 days. This was an attempt to control for both growth and practice effects--a substantial design improvement over the previous studies reviewed.

Limitations of the study conducted by Hamm et al. (1975) are that no differences were reported for the individual subtests, and the authors dealt with a small, restricted sample of children who had already been identified and enrolled in EMR classes for at least six months. This may have had a profound impact on the final results and severely limits the generalizability of the results.

Berry and Sherrets (1976) conducted both a pilot study and a major study comparing the WISC and the WISC-R. The pilot study compared the scores of 14 special education students who had been administered the WISC-R and the WISC some time previously by other examiners. Their method was similar to that employed by both Zimmerman (1975) and Swerdlik and Rice (1975). The findings

indicated that the two scales were measuring similar abilities, with the Full Scale correlation coefficient equal to .90. The students, however, performed significantly lower on the WISC-R, with the largest difference being 7.79 points on the Verbal Scale ($p<.001$). The limitations of the Zimmerman (1975) and Swerdlik and Rice (1975) efforts apply to this pilot study as well.

The major Berry and Sherrets (1976) study had a sample of 28 special education students from an urban school district. The sample had originally comprised 30 subjects, but the authors reported that a tornado had struck the area the day before testing and they felt two subjects were too emotionally upset to produce valid results. The age range for the 28 subjects was from 8.7 to 15.6 years, with a mean of 11.8. The tests were administered at two-week intervals in a counterbalanced order to control for both growth and practice effects. The results were similar to those obtained in the pilot study. Moderately strong correlations between the WISC and WISC-R were reported for the Verbal Scale (.74), Performance Scale (.85), and Full Scale (.86). The following mean significant WISC/WISC-R differences were obtained: 4.43 points for the Verbal Scale, 3.25 for the Performance Scale, and 3.43 for the Full Scale IQ. In all cases, the WISC-R results were lower than the WISC results. The authors concluded that a larger number of children would be classified as retarded using the WISC-R and a larger number would be placed in special education classes.

Deficiencies of the Berry and Sherrets (1976) study include the limits to the generalizability of the results because of the

small, restricted sample of special education students. In addition, no analysis of subtest scores was reported, the possible age effects were not examined, and the reader was not informed of the possible effects of the tornado on the subjects included in the sample.

Kaufman and Weiner (1976) compared the results of the WISC and the WISC-R for 46 low-SES black children aged 7 to 10 years who had been referred to a Brooklyn, New York, clinic for suspected learning and/or behavioral disorders. The tests were administered in a counterbalanced order, with a mean test-retest interval of seven weeks. The WISC-R consistently yielded lower IQ scores than the WISC. The IQ score differences were 7 points for the Verbal Scale and 8 points for both the Performance Scale and the Full Scale. Between the WISC and WISC-R, correlations of .90 and .82 for the Verbal and Performance Scales, respectively, were found. A final result reported was a .63 correlation between the WISC-R Full Scale IQ and the Wide Range Achievement Test (WRAT) Reading subtest, which was significant at the .01 level. This is similar to the .59 correlation found for the WISC Full Scale IQ and the WRAT Reading subtest.

Limitations of the Kaufman and Weiner (1976) study include the lack of analysis of the 12 subtests. In addition, their sample comprised children referred to a Brooklyn clinic for behavioral difficulties. Such children tend to have unstable test scores to begin with; hence this would affect the results of the study, because the researchers employed a test-retest design. The study

did, however, include validity data for the WISC-R and is consistent with previous studies, reporting a fairly large WISC/WISC-R score difference with the WISC-R yielding lower scores.

In comparing the test results of 22 deaf children between the ages of 9 and 11, who were tested in the spring of 1974 on the WISC and one year later on the WISC-R, Davis (personal communication) found that five subjects obtained higher Performance IQ's on the WISC-R (an average increase of 7 points) and 17 subjects obtained lower Performance IQ's on the WISC-R (an average decrease of 8.4 points). The Verbal Scale was not administered because the children were deaf.

Limitations of the Davis study are similar to those of the Zimmerman (1975) and Swerdlik and Rice (1975) research and Berry and Sherrets' (1976) pilot study. In addition, Davis utilized a small, restricted sample, which limits the generalizability of the results. Also, no subtest analysis was provided. However, the study did deal with a population (deaf children) for which WISC/WISC-R comparison data previously had not been available.

Solway et al. (1976) compared WISC and WISC-R results for a group of juvenile delinquents, and found significant differences between the two tests for Verbal ($p<.05$), Performance ($p<.0001$), and Full Scale IQs ($p<.01$) and all 10 subtests administered except Information, Comprehension, Vocabulary, and Picture Arrangement. The differences for Verbal, Performance, and Full Scale IQs were 2.35, 3.67, and 3.05, respectively. WISC/WISC-R differences for the subtests ranged from .46 on Information to 8.39 on Similarities.

In all cases the WISC-R results were lower than the WISC results. No significant differences were found for different sexes, races (whites, blacks, and Mexican-Americans), ages, or grades.

A limitation of the Solway et al. (1976) study involves a limit to the generalizability of the results. Further, their experimental design did not have each subject receive both tests. They took a large sample of juvenile delinquents and randomly selected those who would be administered only the WISC and a group who would only be administered the WISC-R. The results were then compared, assuming the two groups were identical.

Reschly and Davis (in press) attempted to determine the comparability of WISC and WISC-R scores among children aged 7.9 to 16.1 from borderline and educable levels of intelligence. They administered the WISC and WISC-R to 48 children in Tuscon, Arizona, who had been referred and evaluated for special education placement. Davis originally administered the WISC; the WISC-R was administered by different graduate students enrolled in an intellectual assessment course. The time interval between testings ranged from 5 to 26 months, with a mean of 17.3 months. The WISC was always administered first. In almost every case the WISC-R scores were lower than the WISC scores. The largest differences were reported on the Verbal IQ Scale and on several Verbal subtests. Performance IQ scores on both tests were comparable; in fact, scores on three of the five Performance subtests (Picture Completion, Picture Arrangement, and Object Assembly) were either at

similar levels on both tests or were significantly higher on the WISC-R. The difference between Full Scale IQ scores on the two tests was 4 points.

Limitations of this study are similar to those previously discussed. They include the lack of control for both growth and practice effects, and the fact that the WISC-R was always administered by relatively inexperienced, noncertified examiners. Also, the WISC/WISC-R examiners were always different.

All of the studies previously described have dealt with samples of low-ability students and/or students enrolled in special education EMR classes. In contrast, Larrabee and Holroyd (1976) compared scores earned by 38 highly intelligent fifth graders on both the WISC and the WISC-R. These children attended a private school in a Pasadena, California, suburb; the school had a reputation for academic excellence. Significant WISC/WISC-R differences were reported for Verbal, Performance, and Full Scale IQs, with the WISC score being higher in all cases. The mean differences between the WISC and the WISC-R were large: 9.6 points for the Verbal IQ, 8.4 points for the Performance IQ, and 9.4 points for the Full Scale IQ. Correlations between WISC and WISC-R subtests ranged from .269 for Picture Arrangement to .936 for Similarities.

Limitations of the Larrabee and Holroyd (1976) study include the restricted sample and small sample size. However, it does provide data for the upper IQ ranges, which were not previously available. The differences found are the highest reported for the Verbal and Full Scale IQ Scales. This also contradicts previous

studies, which predicted and found less difference between the WISC and the WISC-R for higher-ability students.

Schwarting (1976) administered sets of the WISC and WISC-R to 58 randomly selected children aged 6 through 15 years in a suburban Omaha, Nebraska, school containing grades one through eight. The order of administration was counterbalanced to control for practice effects. The test-retest interval between the two tests for each child ranged from 60 to 67 days. Omitted were supplementary tests of Digit Span and Mazes. Significant mean WISC minus WISC-R differences were reported for Verbal (mean difference=4.86), Performance (mean difference=8.74), and Full Scale IQs (mean difference=7.49). All of the WISC/WISC-R mean differences for the 10 subscales were significant except for Vocabulary. The mean WISC-R scores were lower in all cases except Comprehension, on which the mean WISC score was lower than the WISC-R score. The following regression equations were also computed to predict WISC-R scores from WISC results:

WISCR Verbal IQ = .91 x (WISC Verbal IQ) + 5

WISC-R Performance IQ = .77 x (WISC Performance IQ) + 17.75

WISC-R Full Scale IQ = .91 x (WISC Full Scale IQ) + 2.72

The limitations of Schwarting's (1976) study include restrictions on the generalizability of the results because of the location of the sample. In addition, supplementary tests were omitted. Because of the small sample size, the author did not investigate the possible specific age or ability effects or racial

differences. However, Schwarting's study is the only one to date that permits generalization of the results to the entire school population of one school building.

A summary of the major findings of studies comparing the WISC and the WISC-R is presented in Table 3.3. The WISC/WISC-R Verbal IQ differences ranged from 1.5 to 9.6 IQ points, Performance IQ differences ranged from 2.2 to 10 IQ points, and Full Scale IQ differences were from 2.1 to 9.4 points. Two of the 10 studies reviewed reported a specific age effect (a greater difference between the WISC and the WISC-R for younger than for older students) and a specific ability effect (a greater difference between the WISC and the WISC-R for lower than higher ability students). In all cases the WISC-R yielded lower scores. In several studies, the Coding and Similarities subtests showed the largest WISC/WISC-R differences.

In the present study, an attempt was made to improve upon previous studies reviewed in this chapter in the following ways:

1. It controlled for both growth and practice effects by utilizing a counterbalanced design with a specific test-retest interval.

2. It employed a larger, less restrictive sample of children referred to school psychologists primarily because of concerns about their intellectual ability. The sample was drawn from both rural and urban areas of a midwestern tri-state area. This is the population with whom the test has most frequently been utilized.

3. The study examined how nonwhites perform on the WISC-R.

Table 3.3.--Major findings of WISC/WISC-R comparison studies.

Researcher	Sample	Procedure	WISC - WISC-R Differences
Doppelt and Kaufman (in press)	WISC & WISC-R standardization sample.	Statistical prediction study.	Age: 6.5-15.5 Verbal IQ: 1.5 pts. Performance IQ: 6 pts. Full Scale IQ: 4 pts.
Zimmerman (1975)	86 cases of EH and EMR children in California.	No test-retest time limit. WISC always administered first.	EH Students: Verbal IQ: 4.9 pts. Performance IQ: 3 pts. Full Scale IQ: 4.1 pts. EMR Students: Verbal IQ: 3.3 pts. Performance IQ: 2.2 pts. Full Scale IQ: 2.1 pts.
Swerdlik and Rice (1975)	41 cases of EMR and non-EMR children in Lansing, Michigan area.	Test-retest time limit of 4 years. WISC always administered first.	Verbal IQ: 3.80 pts. Performance IQ: 2.74 pts. Full Scale IQ: 3.05 pts. Vocabulary: 1.31 pts.
Hamm et al. (1975)	48 EMR students in rural southeast Georgia.	Test-retest interval of 39 days; counterbalanced order.	Verbal IQ: 6 pts. Performance IQ: 10 pts. Full Scale IQ: 7.5 pts.
Berry and Sherrets (1976)	28 special education EMR students from urban school district. Ages = 8.7-15.6 years Mean age = 11.8 years	Test-retest interval of 14 days; counterbalanced order.	Verbal IQ: 4.43 pts. Performance IQ: 3.25 pts. Full Scale IQ: 3.43 pts.
Kaufman and Weiner (1976)	46 low-SES black children aged 7 to 10 years, referred for suspected learning and/or behavioral problems to a New York clinic.	Test-retest mean interval of 7 weeks; counterbalanced order.	Verbal IQ: 7 pts. Performance IQ: 8 pts. Full Scale IQ: 8 pts.

Table 3.3.--Continued.

Researcher	Sample	Procedure	WISC - WISC-R Differences
Davis (personal communication)	22 deaf children aged 9-11 years; urban area.	Test-retest interval of not less than 1 year; WISC always administered first.	Performance IQ: 8.4 pts. Verbal Scale not administered because Ss were deaf.
Solway et al. (1976)	Juvenile delinquents in Texas urban area.	Randomly administered WISC and WISC-R to equal numbers of subjects.	Verbal IQ: 2.35 pts. Performance IQ: 3.67 pts. Full Scale IQ: 3.05 pts.
Reschly and Davis (in press)	48 borderline and educable children aged 7.9-16.1 in rural Iowa	Test-retest range = 5 to 26 months; mean = 17.3 months; WISC always administered first.	Verbal IQ: Performance IQ: Full Scale IQ: 4 pts.
Larrabee and Holroyd (1976)	38 high-ability fifth grade students in a California suburb.	Counterbalanced order.	Verbal IQ: 9.6 pts. Performance IQ: 8.4 pts. Full Scale IQ: 9.4 pts.
Schwarting (1976)	58 randomly selected children aged 6-15 years in a suburban Omaha school.	Test-retest range = 60-67 days; counterbalanced order.	Verbal IQ: 4.86 pts. Performance IQ: 8.74 pts. Full Scale IQ: 7.49 pts.

4. It controlled for age to examine the specific age effect (Barclay & Carolan, 1966) and to examine differences across abilities so as not to mask any real differences (Hannon & Kicklighter, 1970).

5. The study explored WISC/WISC-R differences for each of the three major scales and the 12 individual subtests.

6. It identified what school psychologists consider a "meaningful difference" between the two tests.

7. It examined how these tests affect the educational programming of children, by reporting the disposition of the cases involved in this study.

Test Bias

The inclusion of nonwhites in the WISC-R standardization sample; greater ethnic group representation in subtest items; and revision of Vocabulary, Similarities, and Comprehension items to omit seemingly biased items and substitute others have led some researchers to conclude that the WISC-R is a fairer and much improved test for minority groups than was the WISC (Kirchev, 1975; Carvajal & McKnab, 1975). The present study addressed the question of whether the WISC-R is in fact fairer for minority groups by comparing the performance of whites and nonwhites on the two tests.

Reschly et al. (1976) provided a useful summary of the current literature on test bias, which contains highly conflicting views. They discussed three general conceptualizations of test bias that are dominant in the literature.

One viewpoint defines a test as biased whenever mean differences in performance occur among several groups. For example, a test is said to be biased if different racial or ethnic groups score lower, on the average, than the population means. According to this definition, any detected difference is attributed to measurement error. This definition also makes the critical assumption that there are no real differences among the groups at the outset. This definition has been endorsed by many researchers (e.g., Jackson, 1975; Williams, 1974).

However, when considering this first viewpoint, Jensen (1975) believes it is important to distinguish between two concepts: culture loading and culture bias. They are not synonymous. Culture loading refers to the specificity or generality of the informational content of the test items that compose a particular test. The more specific the culture from which the test's information could be taken, the more culture loaded it is. The amount of specificity or lack of it in the content of the test items corresponds to its culture loading. Jensen (1975) gives examples of two questions that differ in their degree of culture loading. "Name three parks in New York City" is more culture loaded than "How many 10¢ postage stamps can you buy for $1?" Culture content and degree of culture loading of items in a particular test is a different matter from whether the particular content of the test items causes the test to be biased with regard to the performance of any number of groups within the population. Jensen's (1975) position with regard to considering differences between means as an indicator of test bias

is clear. He states,

> To the extent that the test contains cultural content that is generally peculiar to the members of one group but not to members of another group, it is liable to be biased with respect to comparisons of test scores between the groups or predictions based on their scores. Score differences per se, whether between individuals, social classes, or racial groups, obviously cannot be a proper criterion of bias. There is no basis for assuming a priori that any two populations should be equal in whatever it is that the test is supposed to measure (p. 5).

The second viewpoint of test bias addresses itself to external validity and the use of standardized tests in predicting some future outcome, such as academic or employment success. This definition assumes a test is biased or unbiased, based upon the accuracy with which it predicts future performance for all groups. From this viewpoint of test bias, even if various racial or ethnic groups obtain different mean scores on the test, the instrument is still not considered biased if it can be shown to predict accurately and fairly for all groups. A test is thought to be fair if it does not consistently over- or under-predict the criterion score for any racial or ethnic group. Cleary's (1968) definition of test bias best represents this viewpoint:

> A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test is designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup. With this definition of bias, there may be a connotation of "unfair" particularly if the use of the test produces a prediction that is too low (p. 115).

Cleary's definition of test bias has become the most widely accepted by the courts, educational and industrial psychologists, textbook authors, and governmental agencies (Seaton, 1975).

Thorndike (1971) argued that the Cleary definition which selects "the best man for the job" is unfair to minority groups because it will select a smaller proportion of their group that meet a particular criterion level as compared to the majority group. Thorndike's definition of test bias states, "An alternate definition would specify that the qualifying scores on a test should be set at levels that will qualify applicants in the two groups in proportion to the fraction of the two groups reaching a specified level of criterion performance" (p. 63). This definition of test bias, if operationalized as a means of selecting applicants for a job, would result in the selection of a greater proportion of minority group members than the Cleary definition.

The third viewpoint regarding test bias involves the social-policy implications of test use. Definitions advanced by Darlington (1971) and expanded by Novick and Peterson (1976) suggested that predictor scores be adjusted in the direction of socially desirable outcomes to eliminate past inequities among groups. Some have argued against this viewpoint, saying it leads to reverse discrimination. Peterson and Novick (1976) felt this criticism can be overcome if the amount of disadvantage, rather than particular ethnic or racial group membership, is utilized in adjusting scores.

A fourth viewpoint looks at an internal validity measure. Many researchers determine test bias on the basis of item content. Jensen (1976) found the rank order of item difficulty levels of WISC-R items was not significantly different for whites and blacks. In fact, one particular WISC-R item that many critics

have claimed to be culturally biased against blacks ("What is the thing to do if a fellow [girl] much smaller than yourself starts to fight with you?") actually was found to be relatively easier for blacks than for whites. It ranked forty-second in difficulty for blacks, compared to forty-seventh for the white group. Jensen referred to this as an example of "armchair analysis of cultural bias in a specific test item" (p. 16). The item-content approach was the major method used to omit "culturally biased" items from the WISC-R.

Often it is assumed that the Performance score of an intelligence test like the WISC is less "culture bound" and therefore less negatively affected by a deprived social and educational background than is the Verbal Scale. Many researchers have termed the Performance Scale less biased (Telford & Sawrey, 1967; Holland, 1960; Teahan & Drews, 1962). However, most research has suggested that blacks score equal to or higher than whites on the WISC Verbal as compared to Performance subtests (Atchinson, 1955; Caldwell & Smith, 1968; Cole & Hunter, 1971; Hughes & Lessler, 1965; Young & Bright, 1954; Loehlin, Lindzey, & Spahler, 1975).

Previous studies have reported that nonwhites generally perform lower than whites on the WISC (Carson & Rabin, 1960; Simpson, 1960; Zimmerman & Woosam, 1973; Holland, 1960; Webb, 1965; Ortiz, 1968). Many researchers have concluded that blacks score approximately one standard deviation below the mean as compared to whites (Shuey, 1966; Dreger & Miller, 1960; Tyler, 1965).

Specifically, Young and Bright (1954) tested southern Negro children aged 10 to 13 years, and reported a WISC Full Scale IQ mean score of 67.74. They concluded that the WISC was inappropriate for testing southern Negro children because of the lack of nonwhites in the standardization sample.

Caldwell (1954) tested 420 Negro children ranging in age from 6 to 12, with equal numbers of males and females. The sample was selected from towns in five deep-southern states, and was randomly selected from various schools. A difference was found between southern Negro children and the white standardization group. The Full Scale IQ mean obtained for the black sample was 85.52, which is considerably higher than the mean obtained in the Young and Bright (1954) study. Caldwell concluded that cultural bias resulted from using the WISC, which had been standardized on a white population.

Holland (1960) studied a sample of 36 Spanish-speaking children in the first through fifth grades in Tuscon, Arizona. The children were referred for testing because of academic and emotional problems. The WISC subtests were first administered in English, just as in the standard procedure. Only when the instructions were not understood were they repeated in Spanish. Correct answers in either language were credited. Verbal IQ scores for tests administered only in English ranged from 45 to 118; the group mean was 80.6. The Bilingual Verbal IQ scores (mixed English and Spanish as outlined) ranged from 48 to 118, with a mean of 85.2. There was a large discrepancy between the Verbal and Performance IQ scores.

The Performance results were, on the average, 10.2 points higher than the English Verbal IQ for the group. The Bilingual Performance IQ results were, on the average, 5 to 6 points higher than the Bilingual Verbal IQ. Both of these discrepancies were significant at the $p<.01$ level.

Simpson (1970) administered the WISC to 120 Anglo, Mexican, and black 16 year olds of below-average ability. He reported mean Full Scale IQ scores of 85.55 for the Anglo sample, 81.02 for the Mexican sample, and 81.00 for the blacks. Mean Performance IQ scores were 91.72 for the Anglos, 88.25 for the Mexicans, and 85.57 for the blacks. For all groups, the Performance IQs were highest.

Concerning the WISC-R, some evidence suggests that nonwhites continue to score below whites as they did on the WISC. Mercer (1975) gathered WISC-R data on 688 Anglo-American, 620 Chicano/Latino, and 616 black children in a California suburb. She reported mean WISC-R Full Scale IQ scores of 103 for the white sample, 91.5 for the Latino sample, and 87.6 for the black sample.

Kaufman and Doppelt (1976), in their analysis of data from the 2,200 children included in the WISC-R standardization sample, found whites scored approximately one standard deviation (15 points) higher than blacks on the WISC-R Verbal, Performance, and Full Scale IQ measures.

In their analysis of WISC-R results, Jensen and Figueora (1975) found that the backward digits portion of the Digit Span subtest correlated more highly with total IQ than did forward digit span, and that blacks and whites differed more on the backward than

the forward digit span portions. Because Digit Span is considered to be one of the least culturally loaded subtests of the WISC-R, the authors used this finding as evidence of true differences in intellectual ability between blacks and whites.

As is evident from the preceding discussion, there are many definitions and conceptions of what constitutes test bias. An attempt was made to cite some of the research pertaining to each of these definitions and conceptions.

The present study attempted to investigate the question of whether the WISC-R is less biased than the WISC by examining the performance of nonwhites on both tests. Because of the characteristics of the data collected in the present study, two of the previously discussed conceptions of test bias have been employed in this study: (a) A test is said to be biased if the means among several racial and/or ethnic groups differ; this definition was supported by Jackson (1975) and Williams (1974). (b) The second conception of test bias utilized in the present study involves comparing the Verbal-Performance Scale discrepancies between the WISC and WISC-R. This conception of test bias assumes the Verbal Scale of the WISC is more "culture bound" and more negatively affected by a deprived social and educational background than the Performance Scale. Many researchers (Telford & Sawrey, 1967; Holland, 1960; Teahan & Drews, 1962) consider the Performance Scale to be less biased than the Verbal Scale. In the present study, if the WISC-R had a smaller Verbal-Performance Scale discrepancy than the WISC,

advocates of this position would interpret this to mean that the WISC-R is less biased than the WISC.

This chapter presented a detailed review of previous studies that compared the WISC and the WISC-R tests, a discussion of various definitions and conceptions of test bias, and a review of previous studies that have compared the performance of whites and nonwhites on the WISC and the WISC-R.

CHAPTER IV

METHOD

Presented in this chapter is a detailed description of the sample of examiners, subjects, and general procedures employed in data collection and analysis. In addition, the study's hypotheses are set forth.

Sample of Examiners

The sample of examiners was drawn from a pool of school psychologists in Michigan, Illinois, and Ohio. A letter (see Appendix A) was mailed to approximately 800 members of the local school psychology organizations in the tri-state area during the first week in September, 1975. Further, ads soliciting examiners were placed in the APA Monitor and the weekly newsletter, Behavior Today. Announcements were also made and letters distributed at the 1975 annual meeting of the American Psychological Association in Chicago and the fall, 1975, meeting of the Illinois Psychological Association.

Potential participants were asked to return a form (see Appendix A) in a self-addressed stamped envelope. On this form, the school psychologist indicated his intent to participate, the number of children he was willing to test, and certain demographic information. Seventy-two school psychologists in the tri-state

area volunteered to test 164 children aged 6-15.11, who had been referred to them for suspected intellectual deficiency. A summary of examiner characteristics, as well as demographic data pertaining to their schools, is presented in Table 4.1.

A comparison of participant examiners to nonparticipants, defined as those who returned forms indicating they would participate and in fact did not and those who returned forms indicating they would not participate, on the variables of state, school district size, and years of experience revealed virtually no differences between the groups. Comparing the characteristics of the participating examiners with summary characteristics of school psychologists obtained from state-wide survey information indicated that:

1. The participating examiners from the state of Michigan tended to be slightly younger, more possessed master's degrees and fewer held doctorates, and they were less experienced than the average Michigan school psychologist. However, the Michigan summary statistics are somewhat dated (1970-71), and with the implementation of mandatory special education the state has recently employed an increased number of school psychologists at the master's and specialist levels. This would indicate that the participating school psychologists were not significantly different from the average Michigan school psychologist (see Tables 4.2 and 4.3).

2. The participating examiners from Illinois tended to be slightly younger and less experienced than the average Illinois school psychologist. However, in regard to their training (highest

Table 4.1.--Examiner characteristics (N=72).

Examiner Age	
Range: 22-63	
Mean: 36.28	
No response: 9	
Highest Degree Earned	Frequency
Master's	41
Specialist	19
Doctorate	6
No response	6
Years of Training Including Internship	
1 year	10
2 years	25
3 years	15
4 years	7
5 years	3
6 years	8
No response	4
Mean = 2.9 years	
Years of Experience as a School Psychologist	
0 years	8
1 year	7
2 years	9
3 years	8
4 years	7
5 years	7
6 years	9
7 years	2
8 years	5
9 years	1
10 years	3
11 years	1
15 years	1
20 years	1
No response	3
Mean = 4.66 years	

Table 4.1.--Continued.

Number of Psychological Batteries Administered Last Year

Range: 0-350
Mean: 90.12
No response: 4

Number of Psychological Batteries Expected to Administer This Year

Range: 0-350
Mean: 88.24
No response: 5

Number of Children Tested by Examiners	Number of Examiners
1 child	27
2 children	27
3 children	7
4 children	7
5 children	2
6 children	1
16 children	1

degree earned), they were similar to the average school psychologist in their state (see Tables 4.2 and 4.3).

3. The participating examiners from Ohio had equal training (highest degree earned) and experience compared to the average Ohio school psychologist. However, participating school psychologists tended to be older. It should be noted that the Ohio state-wide

Table 4.2.--Summary statistics of participants by states.

	Michigan	Illinois	Ohio
Age			
Less than 25	9%	10%	0%
25-35	44%	50%	25%
36-45	32%	20%	42%
Over 46	15%	20%	33%
Highest Degree Earned			
BA	0%	0%	0%
MA	54%	82%	67%
Specialist	41%	0%	25%
Doctorate	5%	18%	8%
Years Experience			
1-3	51%	41%	≤5=42%
4-6	30%	23%	
7+	19%	36%	≥5=58%

Table 4.3.--Summary data from tri-state area.

	Michigan[a]	Illinois[b]	Ohio[c]
Age			
23-25	0%	≤30 =30%	10%
25-35	36%	31-40=38%	38%
36-45	39%	41-50=19%	26%
Over 46	26%	51-60= 6%	26%
		60+ = 2%	
		NA[d] = 6%	
Highest Degree Earned			
BA	2%	0%	0%
MA	20%	84%	89.3%
Specialist	38%	Not Reported	Not Reported
Doctorate	40%	16%	10.7%
Years Experience			
1-3	40%	28%	≤5=49%
4-6	27%	32%	
7+	33%	40%	≥5=51%

[a]Lesiak (1971).

[b]Klemt and Peterson (1975); Illinois Office of Education (1975).

[c]Cardee et al. (1976).

[d]NA = No answer.

data were obtained from a survey in which only 255 of 800 Ohio school psychologists participated. It is not possible to determine whether they adequately represented the entire population of Ohio school psychologists (see Tables 4.2 and 4.3).

Subject Sample

The 72 participating school psychologists agreed to administer the WISC and WISC-R to 164 children. The examiners were asked to select children of particular races and ages who had been referred to them primarily because of suspected mental deficiency. A summary of subject characteristics appears in Table 4.4. It is difficult to determine if the children referred and utilized in this study are representative of those referred to school psychologists throughout the tri-state area because of concerns about their intellectual ability. No data are available to make this comparison.

The sampling procedure employed in this study had the following limitations:

1. Participating examiners consisted of those who were interested in the question of the equivalency between the WISC and WISC-R, and conceivably could have suspected a difference between the two tests to begin with.

2. It was not a random sample of examiners in the tri-state area.

3. The examiners did not randomly choose children within particular age ranges and of particular racial groups.

4. Those who composed the examiner pool belonged to the professional local school psychology organization and/or had attended

Table 4.4.--Subject characteristics (N=164).

Characteristic	Frequency
States	
Michigan	74
Illinois	52
Ohio	38
Community Type	
1 (Metropolitan core: one or more adjacent cities with a population of 50,000 or more that serve as the focal point of their environs)	71
2 (City: Community of 10,000 to 50,000)	28
3 (Town: Community of 2,500 to 10,000)	8
4 (Urban fringe: A community of any population size that has as its economic focal point a metropolitan core of a city)	43
5 (Rural community: A community of less than 2,500)	14
School District Size	
Small (less than 1,000)	9
Medium (1,000-2,500)	19
Large (greater than 2,500)	136
Sex	
Male	106
Female	58
Age Level	
Young (7-11.0 years)	100
Old (11.1-15.11 years)	64

Table 4.4.--Continued.

Characteristic	Frequency
Race	
White	104
Black	39
Latino	21
Test-Retest Interval	
Range: 7-31 days	
Mean: 20 days	
Grade Level	
1st	24
2nd	19
3rd	28
4th	24
5th	16
6th	16
7th	17
8th	11
9th	7
10th	2
Ability Classification	
Above average (F.S. IQ above 115)	7
Average (F.S. IQ 90-114)	58
Below average (F.S. IQ less than 90)	99

professional conventions. This may indicate that they were more active, more interested in research, and had greater professional identity than the total population of school psychologists in the tri-state area.

Design

The design over subjects included two levels of age (young = 7-11.0 years) and old = 11.1-15.11 years), three levels of race (white, black, and Latino), and two levels of order (WISC first and WISC-R first) all completely crossed. The design over measures included two levels of test (WISC and WISC-R) crossed with two scales (Verbal and Performance). The Full Scale is a composite of the Verbal and Performance Scales.

Age was included in this study as a variable to investigate the specific age effect (Barclay & Carolan, 1966), which refers to the existence of a greater difference between two tests for different age levels. Barclay and Carolan found a greater difference between the Stanford-Binet Form L-M and the WISC for younger than for older subjects. The research reviewed in Chapter III, which dealt with previous comparison studies between the WISC and the WISC-R, showed conflicting results. Zimmerman (1975) and Doppelt and Kaufman (in press) found a specific age effect, with more of a difference between the WISC and WISC-R for younger than older students. However, other researchers (e.g., Hamm et al., 1975; Swerdlik & Rice, 1975) have reported no specific age effects.

Previous research has also pointed to the possible specific ability effect (Hannon & Kicklighter, 1970; Doppelt & Kaufman, in press). These researchers suggested that the specific ability effect be explored so as not to mask any real differences between two tests. They found more of a difference between two tests for lower than higher ability students (WISC vs. WAIS and WISC vs.

WISC-R). To explore this effect in the present study, one-way analyses of variance were conducted between the WISC minus WISC-R differences for Verbal, Performance, and Full Scale IQ measures and ability levels (above average, average, and below average, determined by the average of the WISC and WISC-R Full Scale IQs). These findings are reported in Chapter V under the heading Supplementary Analysis.

Testable Hypotheses

The following hypotheses were formulated and tested in this study:

Hypothesis 1a: There is no significant difference between WISC and WISC-R IQ scores.

Hypothesis 1b: There is no significant difference between WISC and WISC-R Verbal subtests' scaled scores.

Hypothesis 1c: There is no significant difference between WISC and WISC-R Performance subtests' scaled scores.

Hypothesis 2a: There is no significant interaction between age and IQ test scores measured by the WISC and WISC-R.

Hypothesis 2b: There is no significant interaction between age and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 2c: There is no significant interaction between age and Performance subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 3a: There is no significant interaction between the factors of race and IQ test scores as measured by the WISC and WISC-R.

Hypothesis 3b: There is no significant interaction between the factors of race and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 3c: There is no significant interaction between the factors of race and Performance subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 4a: There is no significant second-order interaction among the factors of age, race, and IQ scores as measured by the WISC and WISC-R.

Hypothesis 4b: There is no significant second-order interaction among the factors of age, race, and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 4c: There is no significant second-order interaction among the factors of age, race, and Performance subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 5: There is no significant interaction between the repeated-measures factors of the WISC and WISC-R and the Verbal-Performance subscales.

Analysis

Three repeated-measures analyses of variance were conducted to test the hypotheses of the study. The first analysis of variance included three crossed factors (age, race, and order of test administration) over subjects and two crossed factors (WISC/WISC-R and Verbal-Performance subscales) over measures. This design is shown in Figure 4.1. The second analysis was a multivariate repeated-measures ANOVA computed for the six Verbal subscales of the WISC and WISC-R. The third ANOVA was the same as the second, except that it was performed on the six Performance subscales. The design of the second and third ANOVA's is shown in Figures 4.2 and 4.3.

Procedure

After the initial data form was received (Appendix A), the respondents were sent a letter describing the procedure, a

Race	Age	Order	WISC Verbal	WISC Perf.	WISC-R Verbal	WISC-R Perf.
White	Young	1	28	28	28	28
		2	28	28	28	28
	Old	1	13	13	13	13
		2	13	13	13	13
Black	Young	1	12	12	12	12
		2	12	12	12	12
	Old	1	6	6	6	6
		2	6	6	6	6
Latino	Young	1	4	4	4	4
		2	4	4	4	4
	Old	1	3	3	3	3
		2	3	3	3	3

Key:
Order: 1 = WISC administered first
2 = WISC-R administered first

Figure 4.1.--Design of the first analysis of variance and cell sizes.

			WISC						WISC-R					
Race	Age	Order	Information	Comprehension	Arithmetic	Similarities	Vocabulary	Digit Span	Information	Comprehension	Arithmetic	Similarities	Vocabulary	Digit Span
White	Young	1												
	Old	2												
Black	Young	1												
	Old	2												
Latino	Young	1												
	Old	2												

Key:
Order: 1 = WISC administered first
2 = WISC-R administered first

Figure 4.2.--Verbal subscales ANOVA.

			WISC						WISC-R					
Race	Age	Order	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes	Picture Completion	Picture Arrangement	Block Design	Object Assembly	Coding	Mazes
White	Young	1												
	Old	2												
Black	Young	1												
	Old	2												
Latino	Young	1												
	Old	2												

Key:
Order: 1 = WISC administered first
2 = WISC-R administered first

Figure 4.3.--Performance subscales ANOVA.

data-collection form, and a stamped self-addressed envelope (see Appendix A). This procedure entailed counterbalancing the order of administration of the WISC and the WISC-R and scheduling a specific test-retest interval of not less than a week or more than a month.

The order of administration was counterbalanced to correct for practice effects; this was similar to the procedures used in previous comparison studies (Hannon & Kicklighter, 1970; Barclay & Carolan, 1966; Quereshi & Miller, 1970; Rohrs & Haworth, 1962; Quereshi, 1963; Hamm et al., 1975). After the data were collected, it was found that there were not always equal numbers of each order of administration within each racial and age cell. Therefore, 32 subjects were randomly eliminated to balance the order of administration within each cell.

The specific test-retest time interval was an attempt to control for both practice and growth effects. It was similar to the intervals employed in the studies by Hannon and Kicklighter (1970) and Quereshi and Miller (1970).

After the test data were received, a questionnaire was forwarded to each examiner (see Appendix A). This questionnaire included items to assess the relationship between examiner expectations and final test results, opinions of what constitutes a meaningful difference between the WISC and WISC-R Full Scale IQs for various IQ ranges, disposition of the particular case, assessment of whether the disposition of the case would have changed if only the WISC results had been utilized and in what manner they

would have been changed, and other demographic and personal information. Frequency counts were tabulated; meaningful correlation coefficients are reported in Chapter V.

Summary

This chapter described in detail the sample of school psychologist examiners drawn from a tri-state area who administered sets of WISC's and WISC-R's in a counterbalanced order within a specific test-retest time interval to children referred to them for suspected mental deficiency. In addition, the research design, data analysis, and testable hypotheses were discussed. The results of these procedures are presented in the next chapter.

CHAPTER V

RESULTS

Findings

The findings of the tests of the hypotheses of this study plus supplementary analysis are presented in this chapter.

Hypothesis 1a: There is no significant difference between WISC and WISC-R IQ scores.

The difference between WISC and WISC-R IQ score means (see Table 5.1) was statistically significant. The complete ANOVA table is presented in Table B1 on page 116. Hypothesis 1a was rejected at the $p<.0001$ level ($F=108.03$; df 1,120). It was concluded that the subjects obtained significantly higher IQ scores on the WISC than on the WISC-R.

Hypothesis 1b: There is no significant difference between WISC and WISC-R Verbal subtests' scaled scores.

The mean differences between WISC and WISC-R Verbal subtests' scaled scores (see Table 5.2) were found to be statistically significant. Hypothesis 1b was also rejected at the $p<.0001$ level ($F=21.28$; df 6,107). All of the individual Verbal subtest differences were significant. Inspecting the univariate F ratios presented in Table B2 on page 117, two of the Verbal subtests (Comprehension and Arithmetic) were significant at the $p<.01$ level, one (Vocabulary) was significant at the $p<.002$ level, and the remaining three subtests were all significant at $p<.0001$. The mean scaled

Table 5.1.--Comparison of WISC and WISC-R Verbal, Performance, and Full Scale mean IQ scores.

	Verbal IQ Scale			Performance IQ Scale			Full Scale IQ		
	WISC	WISC-R	Difference	WISC	WISC-R	Difference	WISC	WISC-R	Difference
Age									
Young	88.19	82.90	5.29	94.75	88.59	6.16	91.47	85.74	5.73
Old	84.09	79.77	4.32	98.02	92.39	5.63	91.05	86.08	4.97
Race									
White	88.52	84.30	4.22	97.87	93.00	4.87	93.20	88.65	4.55
Black	85.14	78.64	6.50	84.25	81.47	7.78	87.20	80.06	7.14
Latino	81.21	75.79	5.42	100.90	93.00	7.90	91.06	84.40	6.66
Order									
WISC first	86.09	82.88	2.05	93.92	94.26	-.34	90.00	88.57	1.43
WISC-R first	87.56	80.83	6.73	97.76	85.45	12.31	92.66	82.89	9.77
All subjects	86.83	81.86	4.97	95.84	89.86	5.98	91.33	85.86	5.47

Table 5.2.--Comparison of WISC and WISC-R Verbal subtests' scaled score means.

	Information			Comprehension			Arithmetic			Similarities			Vocabulary			Digit Span		
	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.
<u>Age</u>																		
Young	7.429	6.440	.989	8.381	7.905	.476	7.238	6.381	.857	9.548	7.369	2.179	8.542	7.893	.649	7.238	5.917	1.321
Old	6.952	6.905	.047	7.857	6.976	.881	6.548	7.000	-.452	9.238	6.690	2.548	6.976	6.667	.309	7.929	7.405	.524
<u>Race</u>																		
White	7.695	7.171	.524	8.390	8.000	.390	7.061	6.659	.402	9.463	7.293	2.170	8.512	7.866	.646	7.488	6.512	.976
Black	6.833	5.600	1.233	7.767	6.967	.800	6.567	6.067	.500	10.070	7.233	2.837	7.500	6.767	.733	7.033	6.100	.933
Latino	5.714	5.357	.357	8.071	6.561	1.500	7.643	7.286	.357	8.000	6.071	1.929	6.143	7.786	-.643	8.286	6.500	1.786
<u>Order</u>																		
WISC First	7.048	6.952	.096	8.238	7.762	.476	7.079	6.778	.301	9.095	7.317	1.778	7.937	7.524	.413	7.254	6.254	1.000
WISC-R First	7.442	6.238	1.254	8.175	7.429	.746	6.437	6.397	.540	9.794	6.968	2.826	8.079	7.444	.635	7.683	6.571	1.112
All subjects	7.270	6.595	.675	8.206	7.595	.611	7.008	6.587	.421	9.444	7.142	2.302	8.008	7.484	.524	7.468	6.412	1.056

scores of the six WISC Verbal subtests exceeded those of the WISC-R.

Hypothesis 1c: There is no significant difference between WISC and WISC-R Performance subtests' scaled scores.

The mean differences between WISC and WISC-R Performance subtest scores (see Table 5.3) were found to be significant at the $p<.0001$ level ($F=14.36$; df 6,77); therefore the hypothesis was rejected. An inspection of the univariate F's presented in Table B3 on page 118 revealed the mean differences were significant for all of the subtests except Object Assembly. Two of the subtests (Block Design and Mazes) were significant at $p<.002$, Picture Completion was significant at $p<.01$, Picture Arrangement at $p<.001$, and Coding at $p<.0001$. In the great majority of differences between the WISC and WISC-R, the WISC-R yielded lower scores than the WISC.

Hypothesis 2a: There is no significant interaction between age and IQ test scores measured by the WISC and WISC-R.

This hypothesis was not rejected at the $p<.5036$ level ($F=.4501$; df 1,120). This indicated that the WISC and WISC-R IQ scores for younger and older subjects did not differ significantly (see Table 5.1).

Hypothesis 2b: There is no significant interaction between age and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

This hypothesis was rejected at the $p<.004$ level ($F=4.6280$; df 6,109). By looking at the univariate F ratios, the interactions were found to be significant at $p<.005$ for the Verbal subtest of Information, whereas the interaction was significant at a lower

Table 5.3.--Comparison of WISC and WISC-R Performance subtests' scaled score means.

	Picture Completion			Picture Arrange.			Block Design			Object Assembly			Coding			Mazes		
	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.	WISC	WISC-R	Diff.
Age																		
Young	9.750	8.933	.817	9.267	7.650	1.617	9.100	8.317	.783	9.308	9.650	-.342	9.650	7.933	1.717	9.500	8.517	.983
Old	9.765	9.324	.441	9.706	9.882	-.176	8.912	8.029	.883	10.850	10.320	.530	9.735	8.206	1.529	9.618	8.824	.794
Race																		
White	9.800	9.400	.400	9.667	9.033	.634	9.500	8.867	.633	10.230	10.120	.110	9.917	8.283	1.634	9.583	8.767	.816
Black	8.450	7.600	.850	7.700	6.350	1.350	7.550	6.150	1.400	8.400	8.300	.100	7.700	6.150	1.550	9.300	8.050	1.250
Latino	11.430	9.786	1.644	10.860	9.000	1.860	9.143	8.357	.786	10.360	11.210	-.850	11.500	9.643	1.857	9.714	8.857	.857
Order																		
WISC First	9.255	9.511	-.256	9.404	9.149	.255	9.191	9.255	-.064	9.574	10.980	-1.406	9.277	8.809	.468	9.851	9.021	.830
WISC-R First	10.260	8.638	1.622	9.447	7.766	1.681	8.872	7.170	1.702	10.150	8.809	1.341	10.090	7.255	2.835	9.234	8.234	1.000
All subjects	9.757	9.074	.683	9.425	8.457	.968	9.031	8.212	.819	9.862	9.894	-.032	9.683	8.032	1.651	9.542	8.627	.915

level--p<.0006--for the Arithmetic Verbal subtest. This indicates there was a greater difference between the WISC and WISC-R for younger than older subjects on the Verbal subtest of Information. On the Arithmetic subtest, older subjects actually scored higher on the WISC-R. In all cases except for Arithmetic, the WISC mean scores were higher (see Table 5.2). These interactions are represented graphically in Figure 5.1.

> Hypothesis 2c: There is no significant interaction between age and Performance subtests' scaled scores as measured by the WISC and WISC-R.

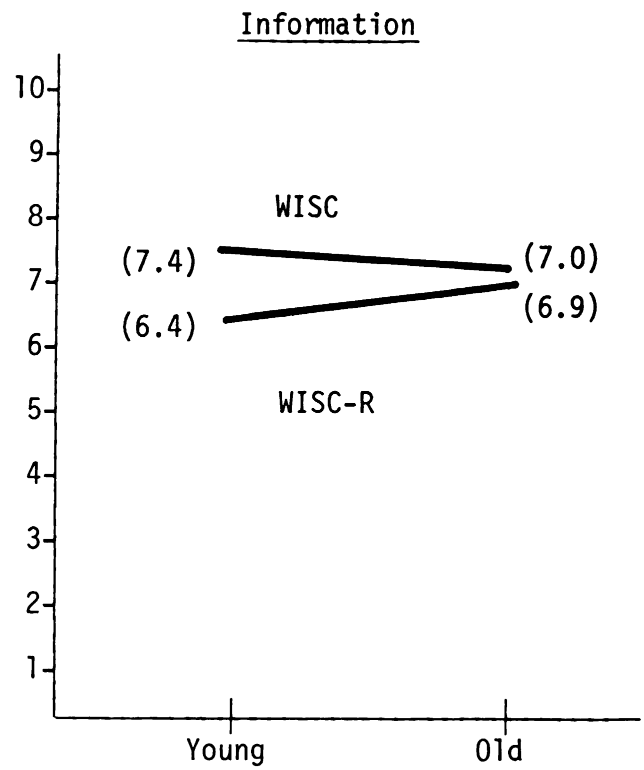
This hypothesis was not rejected at the $p<.1230$ level ($F=1.74$; df 6,77). The data led to the conclusion that there is no difference between WISC and WISC-R Performance subtests' scaled scores for younger and older subjects (see Table 5.3).

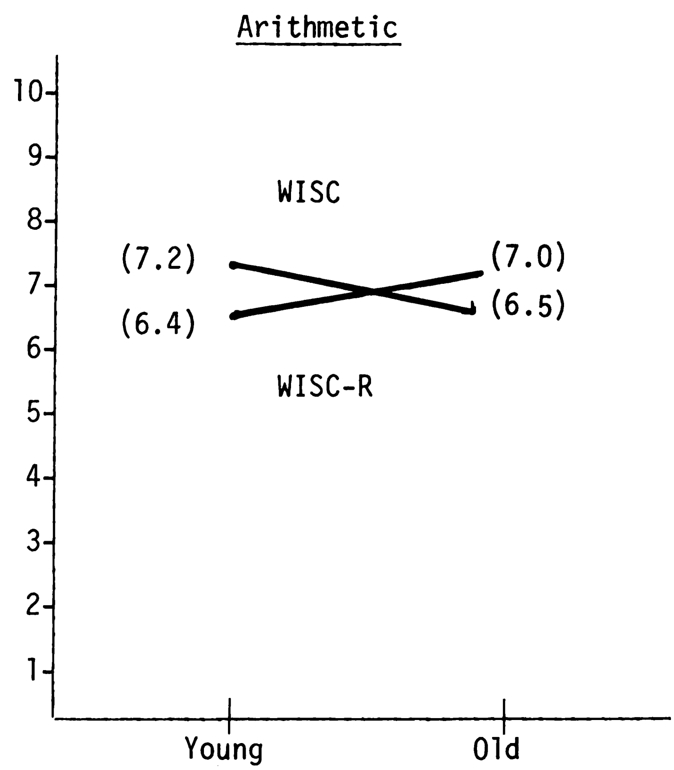
> Hypothesis 3a: There is no significant interaction between the factors of race and IQ test scores as measured by the WISC and WISC-R.

The test of this hypothesis approached statistical significance at the $p<.0733$ level ($F=2.6719$; df 2,120). This provided some tentative evidence that the IQ discrepancy among blacks, whites, and Latinos has increased on the WISC-R as compared to the WISC, despite the effort to narrow it. On the WISC-R, whites lost an average of four points, blacks seven points, and Latinos five points (see Table 5.1). This interaction is presented graphically in Figure 5.2.

> Hypothesis 3b: There is no significant interaction between the factors of race and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

This hypothesis was not rejected at the $p<.2441$ level ($F=1.2652$; df 12,218). WISC/WISC-R Verbal subtests' scaled score





Firuge 5.1.--Interaction of age and Information and Arithmetic subtests as measured by the WISC and the WISC-R.

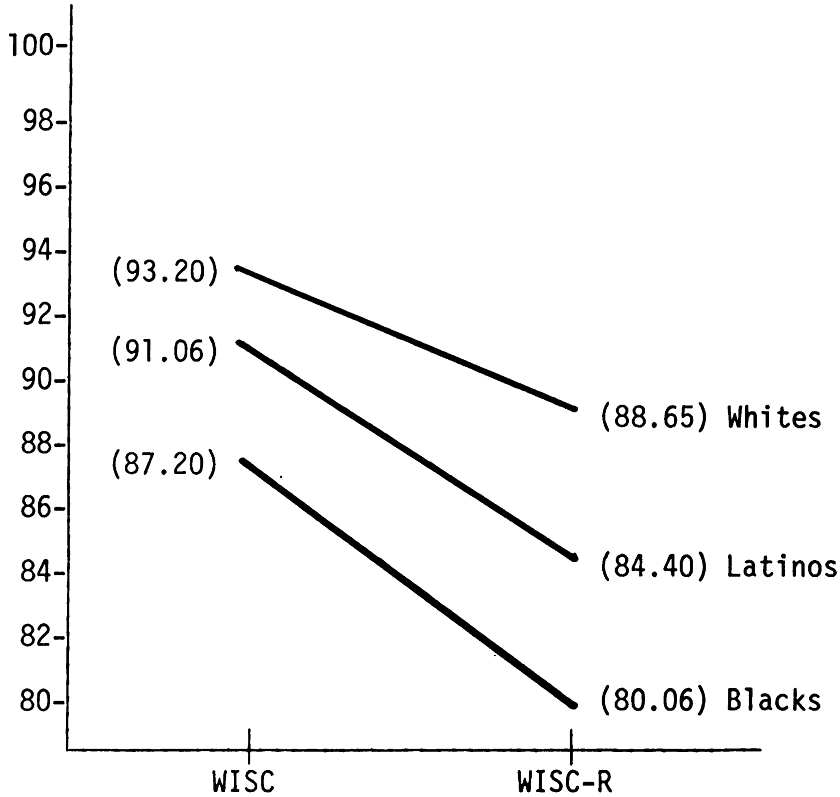


Figure 5.2.--Interaction of race and WISC/WISC-R IQ scores.

differences among blacks, whites, and Latinos did not vary significantly (see Table 5.2).

> <u>Hypothesis 3c</u>: There is no significant interaction between the factors of race and Performance subtests' scaled scores as measured by the WISC and WISC-R.

Hypothesis 3c was also not rejected at the $p<.7760$ level ($F=.6724$; df 12,154). Blacks' whites', and Latinos' WISC/WISC-R Performance subtest scaled score differences did not vary significantly (see Table 5.3).

> <u>Hypothesis 4a</u>: There is no significant second-order interaction among the factors of age, race, and IQ scores as measured by the WISC and WISC-R.

The second-order interaction was found not to be significant at the $p<.7591$ level ($F=.2763$; df 2,120); thus the hypothesis was not rejected.

> <u>Hypothesis 4b</u>: There is no significant second-order interaction among the factors of age, race, and Verbal subtests' scaled scores as measured by the WISC and WISC-R.

This hypothesis was not rejected at the $p<.5931$ level ($F=.8556$; df 12,218) and the data indicated this second-order interaction was not significant.

> <u>Hypothesis 4c</u>: There is no significant second-order interaction among the factors of age, race, and Performance subtests' scaled scores as measured by the WISC and WISC-R.

This second-order interaction was not significant and the hypothesis was not rejected at the $p<.9735$ level ($F=.3659$; df 12,154).

<u>Hypothesis 5</u>: There is no significant interaction between the repeated-measures factors of the WISC and WISC-R and the Verbal-Performance subscales.

This hypothesis was not rejected at the $p<.2495$ level ($F=1.3392$; df 1,120). The Verbal-Performance score discrepancies did not differ for the WISC and the WISC-R (see Table 5.4). In all cases, the Performance IQ score was higher.

A significant interaction at $p<.0001$ was found between the factors of order of administration and WISC/WISC-R scores ($F=58.7136$; df 1,120). This interaction is represented pictorially in Figure 5.3. It was expected because the practice effect would artificially inflate the scores of the test administered second, thus either increasing or decreasing the differences between the WISC and WISC-R scores.

The second-order interaction of order, test, and Verbal-Performance Scale was significant at the $p<.0001$ level ($F=27.0285$; df 1,120). This interaction is represented pictorially in Figure 5.4. It illustrates that the Verbal-Performance discrepancy differed between the WISC and the WISC-R depending on which test was administered first (order). This interaction was partially a result of the fact that practice effects have a greater influence on the Performance Scale than on the Verbal Scale. Thus, when the WISC was given first, WISC-R Performance scores were higher because of practice but WISC-R Verbal scores were lower because of the hypothesized effect. When the WISC-R was given first, WISC Performance scores were much higher as a result of the combination of practice and test effects.

Table 5.4.--Comparison of WISC and WISC-R Verbal (V)-Performance (P) Scale discrepancies.

	WISC Verbal	WISC Performance	WISC V-P Discrepancy	WISC-R Verbal	WISC-R Performance	WISC-R V-P Discrepancy	WISC V-P Discrepancy	WISC-R V-P Discrepancy	WISC Minus WISC-R V-P Discrepancy
<u>Age</u>									
Young	88.19	94.75	-6.56	82.90	88.59	-5.69	-6.56	-5.69	.87
Old	84.09	98.02	-13.93	79.77	92.39	-12.62	-13.93	-12.62	1.31
<u>Race</u>									
White	88.52	97.87	-9.35	84.30	93.00	-8.70	-9.35	-8.70	.65
Black	85.14	89.25	-4.11	78.64	81.47	-2.83	-4.11	-2.83	1.28
Latino	81.21	100.90	-19.69	75.79	93.00	-17.21	-19.69	-17.21	2.48
<u>Order</u>									
WISC First	86.09	93.92	-7.83	82.88	94.26	-11.38	-7.83	-11.38	3.55
WISC-R First	87.56	97.76	-10.20	80.83	85.45	-4.62	-10.20	-4.62	5.58
All subjects	86.83	95.83	-9.00	81.86	84.86	-8.00	-9.00	-8.00	1.00

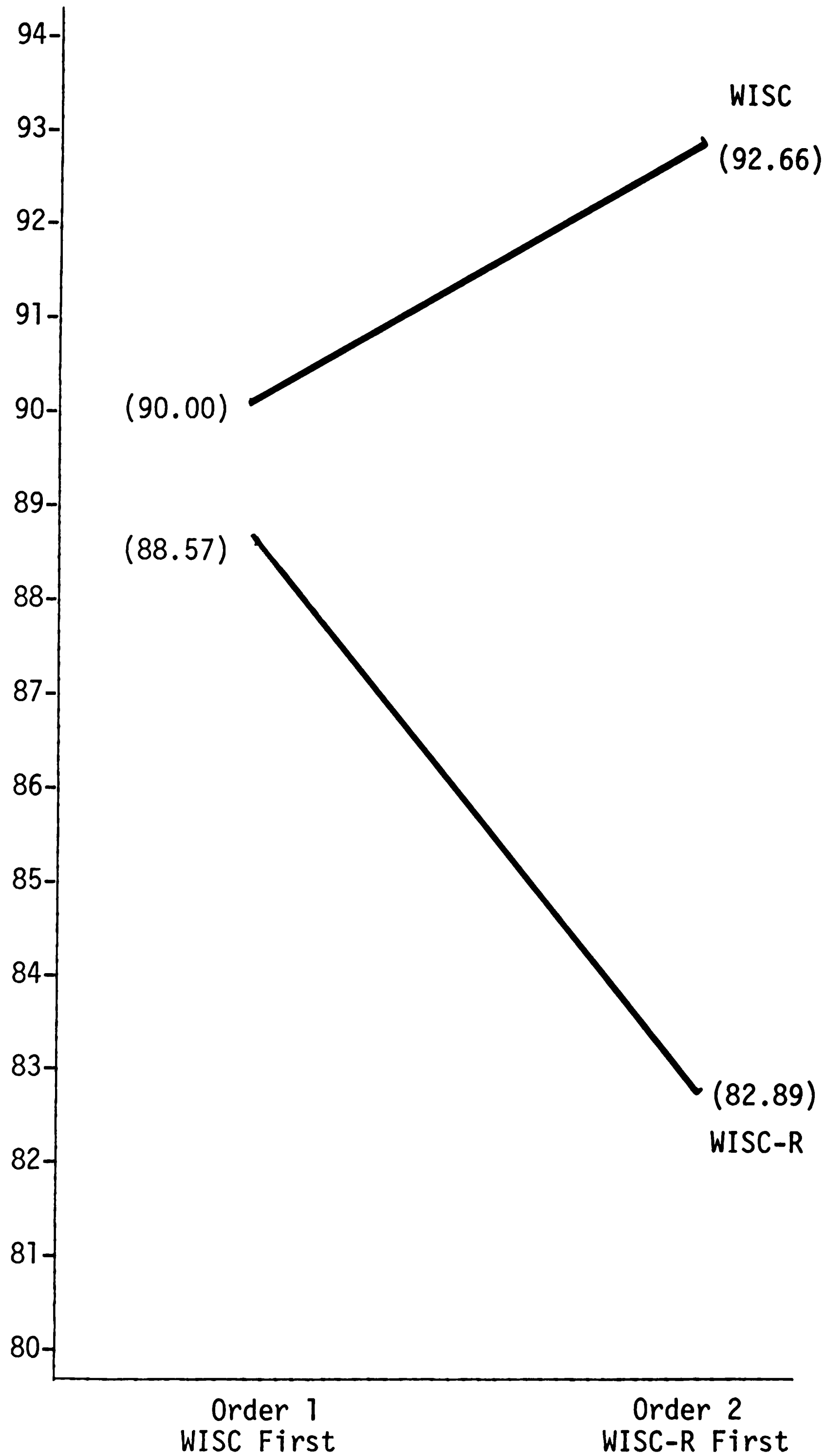
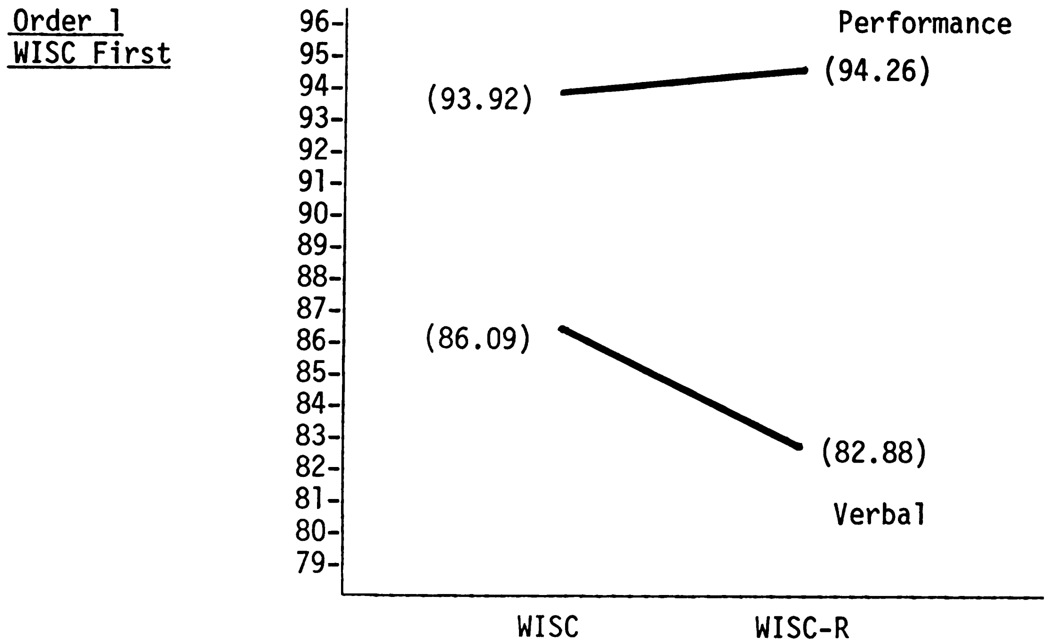


Figure 5.3.--Interaction of order of administration and WISC/WISC-R scores.



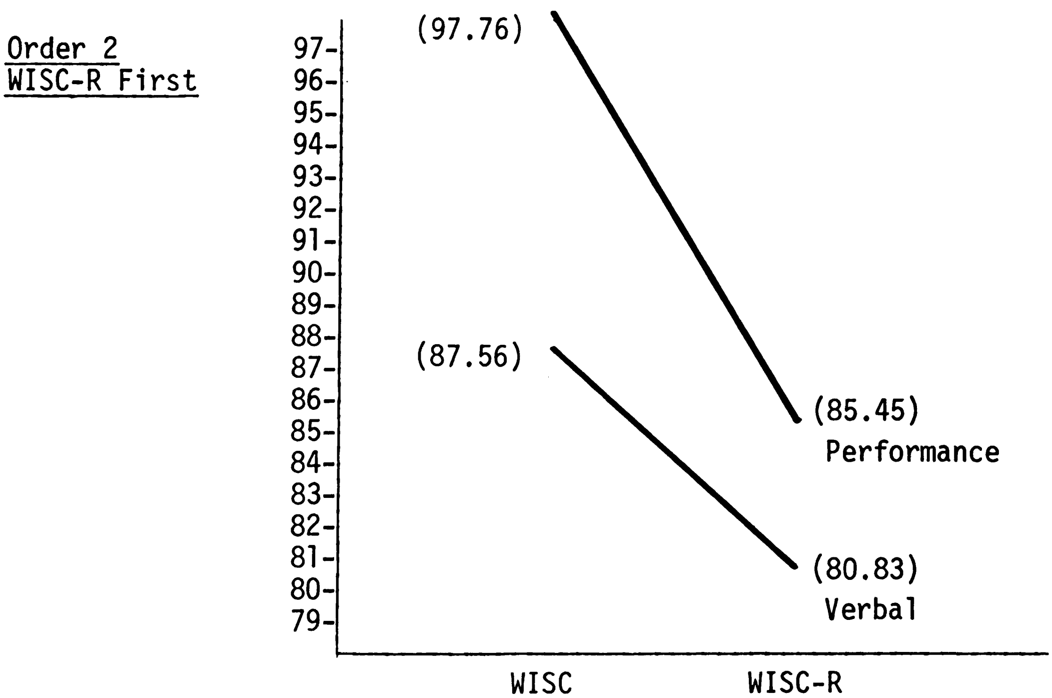


Figure 5.4.--Interaction of order, test, and Verbal-Performance Scale.

Supplementary Analyses

To determine the magnitude of the relationship between the WISC and WISC-R tests, the correlations between the corresponding scales on the WISC and WISC-R were computed separately for the subjects receiving each order of administration. Using r to Z transformations, the average correlation of the two orders for WISC and WISC-R scores was computed, to eliminate the practice effects. The correlations are presented in Table 5.5. The magnitude of the correlations for the three major scales indicates the two tests are highly related.

Table 5.5.--Correlations of WISC and WISC-R scores.

	Correlation
Verbal IQ	.90
Performance IQ	.87
Full Scale IQ	.92
Information	.78
Comprehension	.70
Similarities	.72
Vocabulary	.77
Arithmetic	.73
Digit Span	.70
Picture Completion	.62
Picture Arrangement	.65
Block Design	.72
Object Assembly	.61
Coding	.72
Mazes	.53

A correlation coefficient was computed to assess the degree of the relationship between WISC/WISC-R Full Scale IQ score differences obtained and examiners' expectations of WISC/WISC-R Full Scale IQ score differences. This correlation coefficient of -.17 was found to be statistically significant at the $p<.05$ level (see Table 5.6, questionnaire item #1 results). However, the magnitude of the coefficient was not large enough to be meaningful. This small but statistically significant correlation should not be interpreted as evidence that the examiners' expectations influenced the results. Since the expectations were obtained after the testing, a more likely explanation is that the results of the testing influenced the examiners' expectations.

A one-way analysis of variance was performed to look for differences among subjects in the three-state area on obtained WISC/WISC-R differences. There were no significant differences among subjects from the states of Michigan, Illinois, and Ohio for WISC/WISC-R Verbal, Performance, or Full Scale IQ differences.

No significant differences were found between community type of the subject and WISC/WISC-R Verbal, Performance, or Full Scale IQ score differences obtained. This was also true for various school district sizes.

Nonsignificant correlation coefficients were computed between obtained WISC/WISC-R Verbal, Performance, and Full Scale IQ score differences and examiner's age, number of psychological test batteries administered last year, examiner's years of training and

Table 5.6.--Frequency counts.

---

1. What would you estimate our overall finding regarding WISC/WISC-R Full Scale (FS) IQ score differences will be? Please try to base your response on your intuitive feelings, past experience, reading of the literature, and/or conversations with colleagues.

   Please check one:

   12 WISC FS IQ score higher by 10 or more points

   25 WISC FS IQ score higher by 7-9 points

   15 WISC FS IQ score higher by 4-6 points

   3 WISC FS IQ score higher by 1-3 points

   7 No difference between WISC and WISC-R FW IQ scores

   1 WISC-R FS IQ score higher by 1-3 points

   4 WISC-R FS IQ score higher by 4-6 points

   3 WISC-R FS IQ score higher by 7-9 points

   1 WISC-R FS IQ score higher by 10 or more points

   No response = 1

---

2. How large would the Full Scale (FS) IQ score difference between WISC and WISC-R in each of the following FS IQ ranges have to be before they would affect your decisions regarding a particular case?

I. FS IQ range 60-75	II. FS IQ range 75-90	III. FS IQ range 90-110
0 1-2 points	1 1-2 points	0 1-2 points
12 3-5 points	8 3-5 points	0 3-5 points
28 6-8 points	28 6-8 points	12 6-8 points
19 9-11 points	22 9-11 points	36 9-11 points
10 over 11 points	10 over 11 points	21 over 11 points
No response=3	No response=3	No response=3

experience, and examiner's highest degree earned. In addition, for the various subject characteristics of grade level and sex, no significant correlations were found with obtained WISC/WISC-R Verbal, Performance, or Full Scale IQ score differences. The magnitude of these correlation coefficients is reported in Table 5.7. The only coefficient that was statistically different from zero was the correlation between the number of psychological test batteries predicted for the current year and the WISC/WISC-R Verbal IQ score difference. The magnitude of the correlation (-.15) did not indicate a meaningful relationship.

The preceding results of the supplementary analysis have documented that the obtained WISC/WISC-R differences cannot be attributed to obvious examiner characteristics (e.g., training, age, case load, etc.) and therefore are a result of characteristics of the two tests themselves.

A test of differences across ability groups was conducted for each of the major IQ scales. There appeared to be greater WISC/WISC-R differences for lower ability students. An inspection of the mean differences in Table 5.8 for each of the major scales revealed that the mean WISC/WISC-R difference increased as the student's ability decreased. In all cases, the WISC-R mean scores were lower.

The participating school psychologist examiners' opinions of what constitutes a meaningful difference between WISC and WISC-R Full Scale IQ scores within various IQ ranges are presented in Table 5.6. The majority of school psychologists looked for a 6-8

Table 5.7.--Correlation coefficients of various subject and examiner characteristics and obtained WISC/WISC-R Verbal, Performance, and Full Scale IQ score differences.

Examiner Characteristics	WISC/WISC-R Verbal IQ Differences	WISC/WISC-R Performance IQ Differences	WISC/WISC-R Full Scale IQ Differences
Examiner Characteristics			
Age	-.11	-.02	-.08
Number of psychological test batteries administered last year	-.12	-.05	-.11
Number of psychological test batteries predicted for this year	-.15*	-.02	-.10
Years of experience	-.07	-.07	-.10
Years of training	.002	-.08	-.04
Highest degree earned	-.04	.04	-.02
Subject Characteristics			
Grade level	-.06	-.05	-.08
Sex	-.07	-.01	-.04

*Significant at $p<.05$.

point or greater difference between the two tests in the 60-90 IQ range before their decisions regarding a particular case would be affected. In the 90-110 IQ range, the examiners looked for a 9-11 point or greater WISC/WISC-R IQ score difference.

The dispositions of the cases in this study are shown in Table 5.9. The majority of children were enrolled in special education after testing, primarily classes for the mentally impaired

and learning disabled. For the majority of those who were not placed in special education, the testing led to teacher recommendations.

Table 5.8.--Mean WISC/WISC-R differences for three ability groups.

Scale	Ability Classification	Mean Difference	Significance
Verbal Scale	Above Average[a]	3.14	$p<.18$ ($F=1.732$; df 2,129)
	Average[b]	3.68	
	Below Average[c]	5.83	
Performance Scale	Above Average	2.00	$p<.03$ ($F=3.57$; df 2,129)
	Average	3.00	
	Below Average	7.95	
Full Scale	Above Average	2.29	$p<.07$ ($F=2.66$; df 2,129)
	Average	4.16	
	Below Average	6.93	

[a]Above Average = Full Scale IQ above 115.

[b]Average = Full Scale IQ 90-114.

[c]Below Average = Full Scale IQ less than 90.

Eighty-six percent of the examiners indicated that the disposition of the case they submitted for this study would <u>not</u> have changed if only the WISC results had been available (see Table 5.9). Those who responded that using only the WISC results would have produced a different disposition said approximately 56% of the cases appeared to be eligible for classes that required higher IQs and 44% lost their eligibility for special education. For example, many subjects became ineligible for the mentally impaired classes, or, rather than becoming eligible for classes for the mentally impaired, they became eligible for classes for the learning disabled.

Table 5.9.--Summary of disposition of cases and changes if only the WISC scores had been utilized by examiners.

99	Special education placement
40	Mentally impaired (EMR or EMH)
0	Trainable mentally impaired (TMR or TMH)
41	Learning disabled (LD)
0	Physically handicapped
7	Emotionally impaired (EI)
3	EI/MI
1	MI/LD
7	Resource room--educationally handicapped
36	Teacher recommendations only
4	Referral for outside resources
25	None of the above
20	Disposition would have changed if <u>only</u> WISC results had been utilized.
128	Disposition would <u>not</u> have changed if only WISC results had been utilized.
16	No response

<u>How Disposition Would Have Changed if Only WISC Results Had Been Utilized</u>

8	Not eligible now
1	Special ed. MI to EI
4	Not eligible to LD
1	MI to EH
4	MI to LD
2	No response

To further understand the significance of the WISC/WISC-R differences, these differences were compared to the standard error of measurement of each of the two IQ tests. In 66% of the cases, the WISC Full Scale IQ was four or more points higher than the WISC-R Full Scale IQ. This represents more than one standard error of measurement (Average SEM=3.19). This was also true for 65% of the cases involving the Verbal IQ and 63% of the cases involving the Performance IQ (Average SEM: Verbal Scale=3.60, Performance Scale=4.66). This finding tends to provide additional evidence that the reported WISC/WISC-R differences can be attributed to something other than measurement error.

Summary

This chapter presented the results of the study for each of the test hypotheses and supplementary analyses. Overall, the results indicated that there exists a significant difference between WISC and WISC-R scores, with the WISC-R yielding lower scores. The differences obtained did not appear to be related to any specific examiner characteristics but rather to characteristics of the two tests. A summary of other significant findings is included in Table 5.10.

Discussed in the final chapter are possible explanations for these obtained results, implications for school psychologists and other test users, suggestions for further research, and a summary of the research.

Table 5.10.--Summary of major findings.

Hypothesis	Effect	Scale	p< Sig. Level
1a	WISC/WISC-R differences	All scales	<.0001
1b	WISC/WISC-R differences for Verbal subtests	Verbal	<.0001
		Comprehension	<.01
		Vocabulary	<.002
		Information	<.0001
		Digit Span	<.0001
		Similarities	<.0001
		Arithmetic	<.02
1c	WISC/WISC-R differences for Performance subtests	Performance	<.0001
		Block Design	<.002
		Mazes	<.002
		Picture Compl.	<.02
		Picture Arr.	<.001
		Coding	<.0001
2b	WISC/WISC-R differences by age	Verbal	<.0004
		Information	<.005
		Arithmetic	<.0006
3a	WISC/WISC-R differences by race	All scales	<.0733
	WISC/WISC-R differences by order of administration	All scales	<.0001
	Interaction of order, test, and Verbal-Performance Scale	---	<.0001

Supplementary Analysis		Scale	Mean WISC/ WISC-R Dif.	p<
WISC/WISCR Differences Across Ability Groups	Above Average	Verbal	3.14	
	Average	Verbal	4.19	<.181
	Below Average	Verbal	6.40	
	Above Average	Performance	2.00	
	Average	Performance	4.84	<.031
	Below Average	Performance	8.28	
	Above Average	Full Scale	2.29	
	Average	Full Scale	5.36	<.074
	Below Average	Full Scale	7.55	

CHAPTER VI

SUMMARY AND CONCLUSIONS

The Wechsler Intelligence Scale for Children (WISC), originally published in 1949, was the test most often chosen by school psychologists to assess the intelligence of children in the 7-13 age range and to select candidates for special education programs for the educable mentally retarded. Buros (1972) referred to the WISC as the best test available that claims to measure intelligence. The WISC was revised 25 years after publication and entitled the Wechsler Intelligence Scale for Children--Revised (WISC-R). No comparative studies of the WISC and WISC-R are reported in the WISC-R manual. However, such a comparison is of practical importance because the WISC-R was designed to replace the WISC.

The essential purpose of this study was to compare scores resulting from the WISC and WISC-R for black, white, and Latino children aged 7 to 15.11 years who had been referred to school psychologists in a midwestern tri-state area because of suspected mental deficiency. Also investigated in the study were various conceptions of test bias as it applies to the WISC-R to determine if, for the subjects in this study, the WISC-R is more, less, or equally biased compared to the WISC. A survey of participating school psychologists' views of what constitutes a meaningful IQ

score difference between the WISC and WISC-R was conducted as part of this study. Further, data regarding how the obtained IQ scores for each test influenced decisions about the educational programming of the subjects involved in this study were also reported.

Larrabee and Holroyd (1976) reported that a total of 78% of the WISC-R items have been taken directly from the WISC, 5.9% are from the WISC but have undergone substantial modification, and 16.1% are new items. Like its predecessor, the WISC-R yields a Verbal, Performance, and Full Scale IQ with a mean of 100 and a standard deviation of 15. Both the Verbal and Performance Scales comprise six subtests, which yield scaled scores with a mean of 10 and a standard deviation of 3. The Full Scale IQ is an average of the Verbal and Performance Scales. Changes between the two tests have been made in terms of administration instructions including questioning, scoring criteria, standardization samples including incorporation of nonwhites in the WISC-R standardization sample, and provision of more statistical data in the WISC-R manual.

All previous studies comparing the WISC and the WISC-R have reported the revised test yielded lower scores. The majority of studies comprised a fairly restrictive sample of special education students, employed designs that did not adequately control for both growth and practice effects, and dealt with small numbers of children. No studies were found that attempted to generalize their results to a population of students referred to school psychologists for suspected mental deficiency, nor did any compare the performance of

three different racial groups within a wide age range. However, this is the population with whom the test is most widely used.

In the present study, 72 school psychologists in the tri-state area of Michigan, Illinois, and Ohio administered both the WISC and the WISC-R to 164 children in a counterbalanced order with a specific test-retest interval of not less than a week nor more than a month.

WISC and WISC-R scaled and IQ scores and differences were reported for each of the three major scales and 12 subtests. Significant interactions were also discussed and diagrammed.

Results

The data from this study can be summarized as follows:

1. Subjects obtained significantly higher IQ scores on the WISC than on the WISC-R.

2. WISC Verbal subtests' scaled scores were significantly higher than the WISC-R Verbal subtests' scaled scores.

3. WISC Performance subtests' scaled scores were significantly higher than the WISC-R Performance subtests' scaled scores for all the subtests except Object Assembly.

4. Overall, the differences between the WISC and WISC-R IQ scores were of equal magnitude for younger and older students.

5. A greater difference was found between scaled scores resulting from the WISC and WISC-R for younger than for older students on the Verbal subtests of Information and Arithmetic. The WISC scaled scores were higher for all but the older students on the Arithmetic subtest.

6. For all of the Performance subtests, the difference between WISC and WISC-R scaled scores was of equal magnitude for younger and older students.

7. WISC and WISC-R IQ score differences tended to vary significantly for blacks, whites, and Latinos. In all cases, each of the racial groups scored higher on the WISC than on the WISC-R. These data indicated that the racial IQ discrepancy is widening despite efforts to narrow it. Using the definition of test bias concerning differences among mean IQs of various racial groups, the present study found the WISC-R to be more biased than the WISC.

8. There was no significant difference between Verbal-Performance IQ score discrepancies yielded by the WISC and the WISC-R. In all cases, the Performance Scale was higher. Utilizing the conception of test bias that assumes the Performance Scale is less culture loaded and therefore less biased than the Verbal Scale, this finding would lead one to conclude that the WISC-R is neither more nor less biased than the WISC, but is equally biased.

9. Blacks', whites', and Latinos' WISC/WISC-R Verbal subtests' scaled score differences did not vary significantly.

10. Blacks', whites', and Latinos' WISC/WISC-R Performance subtests' scaled score differences did not vary significantly.

11. Obtained WISC/WISC-R differences were not related to any examiner characteristics such as years of experience or training, nor to subject characteristics such as state of residence, size of community, or sex.

12. WISC/WISC-R differences tended to increase as the ability of the students decreased. In all cases, the WISC yielded higher scores.

13. Participating school psychologists looked for a 6-8 point or greater IQ score difference in the 60-90 IQ range before their decisions regarding a particular case would be affected. In the 90-110 IQ score range, the examiners looked for a 9-11 IQ point difference between the WISC and the WISC-R.

14. After testing, the majority of cases included in the present study were enrolled in special education classes for the mentally impaired or learning disabled. For the majority of children who were not enrolled in special education classes, the testing led the school psychologists to make certain recommendations to the teacher.

15. Eighty-six percent of the participating school psychologists indicated the disposition of the case they submitted for the present study would not have changed if only the WISC results had been utilized in the decision-making process.

Discussion

Significantly different scores resulting from the WISC and the WISC-R have consistently been reported in the literature, with the WISC-R always yielding lower scores of approximately one-third to one-half standard deviation for the three major scales. The present study allowed generalization of this finding to a new population of children aged 7-15.11 who had been referred to school psychologists within the midwestern tri-state area of Michigan, Illinois, and Ohio on the basis of suspected mental deficiency.

Because these differences between the tests have consistently been reported, it is important to speculate why these differences are occurring and to explore the resulting implications for the practicing school psychologists. What follows is a discussion of the possible explanations of these obtained differences, including those that presently may appear remote.

On the surface, it may appear to many observers that the lower WISC-R scores are consistent with the recent observed decline in aptitude and achievement test scores (Harnischfeger & Wiley, 1976). However, on the contrary, these lower WISC-R scores are consistent with the explanation of score differences resulting from a cross-generational increase in IQ (the ability to answer questions on IQ-type tests). This appears to be the most plausible explanation, as it has the most evidence in its support. This explanation would hypothesize that the WISC is currently overestimating the ability of school-aged children and the WISC-R is providing an accurate assessment. It further assumes that if the WISC and WISC-R were administered to the entire population of appropriately aged children, the mean of the WISC would, for example, be 108 and the mean of the WISC-R would be 100.

When children today (1976) are administered the WISC, which was standardized on a 1949 sample of children who had been exposed to very different cultural conditions, they tend to score higher because they are, on the average, better able to answer IQ-type questions than children of 25 years ago. They score higher and appear brighter because, among other things, they have been raised with an increased availability of

of manipulative materials similar to those used in the Performance subtests, greater test sophistication and awareness, an earlier and faster rate of maturation; they have been exposed to organized preschool and kindergarten programs, have received better diet and health care, and have been exposed to television. In summary, children today have generally experienced improved cultural and educational conditions, compared to their 1949 counterparts who made up the WISC standardization sample (Carvajal & McKnab, 1975; Larrabee and Holroyd, 1976; Reschly & Davis, 1976; Schwarting, 1976).

However, when children today are compared with the WISC-R standardization sample, which was collected in 1974, and is composed of their contemporaries who have been exposed to these same improved cultural and educational conditions, their scores on the WISC-R appear relatively lower in relation to their scores on the WISC.

Evidence is available that supports this explanation. By developing regression equations, Doppelt and Kaufman (in press) predicted that the contemporary WISC-R sample would have obtained higher scores on the WISC than those obtained by the original standardization sample. Their study was discussed in greater detail in Chapter III.

Other comparative studies between tests, which have been standardized on different populations at different times, provide additional evidence in support of the cross-generational increase in IQ (ability to respond to questions on IQ-type tests). If this explanation is valid, those tests that were standardized on two

populations separated by a long time interval should display larger differences than two tests that were standardized within a relatively short time interval. This is the case.

Wechsler (1974) reported findings of a comparison of the WISC-R (1974) and the Wechsler Adult Intelligence Scale (WAIS), which was standardized 19 years earlier. In his study utilizing a sample of 16 year olds of average ability, he found WAIS scores were 6.3 points higher on the Verbal Scale, 5.2 points higher on the Performance Scale, and 6.2 points higher on the Full Scale. These differences are similar to those observed between the WISC and the WISC-R, whose standardization samples are separated by 25 years.

Holroyd and Bickley (1976) and Zimmerman and Woo-Sam (1975) also reported higher IQs for the Pinneau norms of the Stanford-Binet Form L-M, which were obtained from a sample in the mid-1930's, versus the more recent 1972 norms.

Thorndike (1975) observed a similar phenomenon in comparing scores yielded by the Stanford-Binet 1960 and 1972 norms. Comparing IQs for the 1972 standardization sample using 1960 norms, the IQs were approximately 5-12 IQ points higher for various age groups. This he interpreted as a general rise in IQ level.

Additional evidence in support of this explanation is provided by the raw and scaled scores of the Coding subtest of the WISC and WISC-R shown in Table 6.1. This subtest underwent only minor changes in instructions in the revision of the WISC; therefore the two tests are comparable (Sattler, 1974). When the Coding subtest was administered first, the mean raw score was 33.91 for the WISC

and 34.07 for the WISC-R. When the Coding subtest was administered second, the mean raw scores were 36.88 and 38.06 for the WISC-R and WISC, respectively. The latter scores were higher because of the practice effect. The similarity between the WISC and WISC-R raw scores on the Coding subtest supports Sattler's assertion that the two tests are essentially the same. An examination of the mean scaled scores reveals a fairly large discrepancy between the WISC and WISC-R (almost two scaled score points--over one-half standard deviation), pointing to the influence of the standardization samples from which the scaled scores were derived. It now takes more raw score points on the WISC-R to earn the identical WISC scaled score.

Table 6.1.--Obtained mean raw and scaled scores on the Coding subtest by order of administration.

	First Subtest Administration		Second Subtest Administration	
	Raw Score	Scaled Score	Raw Score	Scaled Score
WISC	33.91	9.308	38.06	10.20
WISC-R	34.07	7.393	36.88	8.69
Difference	- .16	1.915	1.18	1.51

As predicted by this explanation, research has shown that individual IQ tests developed and normed at approximately the same time produce similar IQs. In a study comparing the 1974 WISC-R with the 1972 Stanford-Binet Form L-M, Wechsler (1974) found mean differences of two points or less between the tests at four different age levels.

Further, in the present study, participating school psychologists indicated they would only consider at the minimum a six to eight point difference between the WISC and WISC-R Full Scale IQ scores, with the WISC-R being lower, as a difference that would alter their decisions regarding a particular case. One way to interpret this finding is to conclude that participating school psychologists recognize that even though the IQ scores are different, the new test accurately estimates the intelligence of school-age children and compares children meaningfully with their peers, and therefore would lead them to the same conclusions regarding these children.

The present study reported for the Verbal subtests of Information and Arithmetic a greater WISC/WISC-R difference for younger than older students. The cross-generational increase in IQ can also account for this finding as being specifically a result of the influence of educational television programs such as Sesame Street. By being exposed to these programs, whose objectives include increasing the viewer's fund of general information such as days of the week and parts of the body (what the Information subtest taps on both tests) and increasing basic arithmetic skills such as adding and subtracting (this is tapped heavily at the younger age levels on both the WISC and WISC-R) (Evans, 1975), children today tend, on the average, to score higher on the WISC and therefore show more of a difference between the WISC and WISC-R. In this study, young was operationally defined as ages 7-11 and old as 11.1-15.11. The greatest majority

of students who could have benefited from a show such as Sesame Street are included in the young category, since the program began in the autumn of 1969. Evaluations of Sesame Street (Bogatz & Ball, 1971) provide additional evidence that the program is meeting their objectives in these two areas.

One explanation of why only the Information and Arithmetic subtests show the positive effects of an educational television program like Sesame Street is that these two subtests tend to tap the lower-level cognitive skills that are relatively easy to translate into educational programming. In contrast, a skill such as abstract reasoning, which is tapped on the Similarities subtest, showed no WISC/WISC-R differences according to age.

The results of this investigation also noted a tendency for students of lower ability to show a greater difference between the WISC and WISC-R than those of average ability. Further, students of average ability tended to show a greater WISC/WISC-R difference than those of above-average ability. This was evident on all of the major scales, and the WISC scores were always higher.

An explanation for this reported ability effect is provided by an inspection of various internal-consistency measures of the two tests. Given that these reported WISC/WISC-R differences are both reliable and replicable, one characteristic of the WISC-R that is hypothesized to be contributing to the observed greater WISC/WISC-R difference at the lower ability levels is the higher split-half reliabilities and intercorrelations between the subtests of each of the major scales as compared to the WISC (Wechsler, 1949; 1974).

Because the WISC-R items are more reliable and more interrelated than the WISC items (if a subject misses one item he is more likely to miss another on the WISC-R as compared to the WISC), this has the effect of lowering the scores of the lower ability students because they tend to miss more items to begin with. However, this explanation does not entirely account for the obtained differences, because it would also predict that the WISC-R scores would be higher than the WISC scores at the upper ability levels (not only would the higher reliabilities and intercorrelations cause the lower ability groups to get more items wrong but also would predict that the upper ability group would get more items correct). The latter prediction was not evident in the reported results. This explanation might also be used to explain the overall WISC/WISC-R differences reported in the present study as the vast majority of subjects fell into the lower IQ ranges (i.e., borderline). However, the differences reported in this study (e.g., 5.5 IQ score points) are too large a difference to be explained by higher WISC-R reliability coefficients of only two or three points (see Tables 2.2 and 2.3).

The obtained differential WISC/WISC-R differences for different ability groups could be explained by using the preceding explanation in combination with the following one. In looking at the amount of growth each ability group has made over the past 25 years, it could be hypothesized that there is greater growth in ability to answer items on IQ-type tests at the lower ability levels or that the lower ability students have benefited more than higher ability students from the improved cultural and educational conditions

described previously. This in combination with the first explanation may account for the observed WISC/WISC-R differences that varied according to ability level of the students.

A remaining explanation for the obtained WISC/WISC-R differences assumes that the WISC-R underestimates students' ability, whereas the WISC accurately assesses it. It further would predict that if the WISC and WISC-R were administered to the entire population of appropriately aged children, the mean of the WISC would, for example, be 100 while the mean of the WISC-R would be 93. There is, however, no evidence for this explanation.

Another explanation for the obtained WISC/WISC-R differences is related to examiners being more familiar and more experienced with the old WISC than they are with the WISC-R. This might lead them to administer and score the WISC-R in a less standardized manner, which could result in lower scores on that test. However, at present, there is no evidence to support this explanation. Further, one might argue that there is no reason why inexperienced examiners would not produce higher scores.

Another possible explanation might be that the individuals who scored the test protocols of the standardization sample (employees of the Psychological Corporation) were more lenient than the average school psychologist in the field who is now scoring the test. This would tend to raise raw scores but not really intelligence. At present, there is no evidence to support this reasoning.

Another explanation involves the sampling of the standardization samples. The WISC-R standardization sample could be brighter

than the population it was designed to represent. Another sampling error could also involve the WISC standardization sample, who may have been duller than the population they were designed to represent. This would lead one to conclude that the obtained WISC/WISC-R differences are a result of sampling error. However, because the WISC standardization sample did not include minorities, who typically score lower on the average than whites (approximately one standard deviation), if anything, the WISC standardization sample was probably brighter than the total population.

Chapter III contained a summary of the various definitions and conceptions of test bias found in the literature. Because of the limitations of the data collected in this study, only two definitions of test bias were employed to determine if the WISC-R is biased. One operational definition of test bias involved differences among the means of several racial and/or ethnic groups. The second conception of test bias used in this study concerned whether the WISC and WISC-R Verbal-Performance scale discrepancies varied significantly among the various racial groups. In using the latter definition, it was assumed that the Performance Scale is a less "culture loaded" part of the intelligence test and therefore is less biased.

The results of this study using the first definition of test bias were somewhat surprising to many, because of claims that the WISC-R is a fairer test for minority groups. The WISC-R tends, in fact, to widen the racial IQ discrepancy rather than narrow it. Blacks and Latinos lost more IQ points than whites, when their WISC

and WISC-R scores were compared. According to this definition of test bias, the WISC-R is more rather than less biased than its predecessor, the WISC. This is contrary to Jensen's (1975) position that it is important to make the distinction between culture loading and culture bias and that they are two separate issues. He also believed that score differences alone cannot be used as a proper criterion of test bias as there is no basis for assuming that any two groups should be equal in what the test is measuring to begin with.

The second definition relating to discrepancies between the Verbal-Performance Scale scores on the two tests leads one to conclude that the WISC-R is neither more nor less biased than the WISC. The Verbal-Performance Scale score discrepancies did not differ significantly for the WISC and WISC-R among the various racial groups.

In addition, for the subjects in this study the scores on the Performance Scale were consistently higher than their scores on the Verbal Scale. This is inconsistent with the majority of studies cited earlier (e.g., Loehlin et al., 1975) that found children within the general population score higher on the Verbal as compared to the Performance Scale. This inconsistency is a result of the fact that children included in the present study were referred because of concerns about their intellectual ability and resulting difficulty in school. Schools are primarily verbal institutions and in order to succeed, pupils require verbal skills such as those tapped on the Verbal Scale of both the WISC and the WISC-R. The referred children included in the present study are more likely to score lower on the

Verbal and higher on the Performance Scale as compared to the general population of children, who score the opposite. There appears to be a general trend of Performance IQ scores higher than Verbal IQ scores in the lower levels of intelligence (Reschly & Davis, in press).

Implications and Recommendations for Test Users

1. The results of the survey of participating school psychologists indicated that the mean Full Scale IQ score difference across all ages and races of 5.5 points is of practical significance in altering a decision pertaining to an individual student's educational program. The use of the WISC-R might have opposite effects on special education programs for the learning disabled (LD) and educable mentally impaired (EMI).

If the present criteria for LD, which are employed in many states, remain in effect (i.e., requires a certain percentage of discrepancy between a child's ability and his achievement in school), the use of the WISC-R will decrease the number of children who would be eligible for programs for the learning disabled. This is because of the fact that it appears that the WISC-R yields lower scores than the WISC, thereby decreasing the discrepancy between ability and achievement for most youngsters. If discrepancies are not adjusted, this might suggest that fewer students will be included in such programs in the future.

The use of the WISC-R would have the opposite effect on programs for the educable mentally impaired, increasing the numbers

of those eligible compared to the number identified in the recent past, but not misclassifying these students. In recent years, by using the WISC, school psychologists have been overestimating the ability of school-age children and thereby identifying fewer children as EMI. By now using the WISC-R, which accurately assesses the ability of these children and compares them meaningfully with their peers, school psychologists will be identifying an appropriate number (those who fall two to three standard deviations below the mean), thus during this transition period of the WISC to the WISC-R, increasing the number of children identified as EMI. This increase in number and resulting overcrowding of classrooms is a fear of many directors of special education.

To prevent this overcrowding during this transition period, school psychologists need to be particularly alert to using data other than WISC-R scores for the placement of youngsters in this program. However, this should always occur and children should never be placed solely on the basis of an IQ score.

The use of the WISC-R might also have implications for mainstreaming efforts of attempting to integrate special education youngsters into the regular program as much as possible. If scores on the WISC-R are the major factor in mainstreaming decisions, fewer children will be integrated out of special education classrooms as they will now be scoring lower. As was true for decisions regarding placement in the EMI program, it is important for school psychologists to look for additional data when providing input into

mainstreaming decisions if their goal is to mainstream as many special education children as possible.

2. The difference of approximately 1/3 to 1/2 standard deviation between the WISC and WISC-R scores, which was reported in most of the studies reviewed in Chapter III and this study, should be interpreted carefully by test users when making certain kinds of judgments. For example, Kaufman and Weiner (1976) urged the user to be cautious before inferring a loss in the child's intellectual functioning if he scores lower on the WISC-R when compared with scores from a previous WISC. These lower scores are to be expected and a difference even of one standard deviation may not be meaningful for this type of judgment.

3. Further, those children who had previously been evaluated using the WISC and scored in the borderline classification range (IQ 70-80), when re-evaluated using the WISC-R will likely become eligible for special education. The clinician should remember that these differences and resulting conflicting conclusions do not reflect negatively on the validity of the WISC-R but rather reflect the influence of different norm groups.

4. The issue of test bias and the WISC-R is quite controversial. From a review of the various definitions of test bias presented in Chapter III, it is obvious that there are many different conceptions of test bias and resulting implications for policy formation. These implications relate to different selection procedures for a particular job, educational program, or class.

Which definition one chooses may be more related to one's values and to whom one wants to be fair than anything else. One can be fair to the individual, the institution, or the minority group. For example, the Cleary (1968) definition of test bias is fair to the institution as it will select the "best man for the job," whereas the Thorndike (1971) definition is fair to the minority group as it will select a fair proportion of each group competing for the job.

Neither of the definitions explored in the present study including differences in mean performance among several group and Verbal-Performance Scale discrepancies seems adequate to determine if the WISC-R is biased.

The data from the present study point to a trend toward the widening of the IQ discrepancy between blacks, whites, and Latinos on the WISC-R compared to the WISC. Some would argue that this evidence indicates that the WISC-R is more biased than the WISC. However, people making this statement assume that there are no differences in ability to answer IQ-type questions between the various racial and ethnic groups to begin with. Currently, there is no evidence to suggest this is true. In addition, this definition of culture bias does not make the distinction between culture loading and culture bias. Jensen (1975) distinguished between culture loading as the specificity or generality of the informational content of the test items and whether the particular content of the test items causes the test to be biased with regard to the performance of any

number of groups within the population. Jensen believed these two concepts are not synonymous.

The results of this study also show that the Performance is higher than the Verbal Scale. Many believe that the Performance Scale is a more "pure" measure of intelligence and less "culture bound" than the Verbal Scale. However, in the present study the Performance Scale was higher than the Verbal Scale for all the racial groups including the majority whites, who usually score higher on the Verbal Scale. It appears that this is more a function of the type of children included in the present study (those children referred because of concerns about their intellectual ability) than a function of test bias.

It is the writer's opinion that the most adequate measures of test bias lie in external validity measures of test bias such as the definitions authored by Cleary (1968) and Thorndike (1971). However, the particular definition one chooses will be a function of one's values and to whom he wants to be fair--the individual, the minority group, or the institution. Each of the definitions will lead to different selection procedures and resulting numbers of individuals from each minority group chosen.

Although the predictive validity literature relating to the WISC-R is limited, the preliminary evidence seems to suggest that the WISC-R has predictive validity similar to the WISC (Reschly, 1976; Kaufman & Weiner, 1976). It remains necessary for school psychologists to continue to exercise caution to use tests in a fair and sophisticated manner.

5. The WISC/WISC-R differences reported in this study, with subjects scoring higher on the WISC as compared to the WISC-R, indicate that intelligence tests must be kept up to date. In suggesting this, Larrabee and Holroyd (1976) recommended that the maximum time between restandardizations and/or revisions should be 10 years. They concluded that "reasonably contemporary normative tables are essential for making valid estimates of a child's level of intellectual functioning if the goal is to compare him meaningfully with his peers" (p. 1080).

Recommendations for Further Research

On the basis of the findings of this study, the following recommendations for further research are set forth:

1. Several of the possible explanations for the obtained WISC/WISC-R differences discussed earlier in this chapter might be investigated. The research might assess their contribution to the obtained WISC/WISC-R differences.

a. A more detailed investigation might be undertaken to determine the effect of the examiner's familiarity, formal training, and experience with the new instrument on obtained WISC/WISC-R differences. This was partially addressed in the present study, which found no significant relationships between the WISC/WISC-R obtained differences and the examiner's years of training, years of experience, or highest degree earned.

b. A more detailed investigation of the 1949 and 1974 WISC and WISC-R norm groups, similar to the Kaufman and Doppelt (1976) study, should be conducted. This study might include an

investigation of the scoring procedures used by the scorers employed by the Psychological Corporation, who scored the protocols of the standardization samples.

c. A more detailed study examining why there are greater WISC/WISC-R differences at the lower ability levels should be conducted. This might include studying the two hypotheses advanced in this study relating to the higher reliabilities, intercorrelations, and whether lower ability students have gained more over the past 25 years than have students of higher ability.

2. States such as Michigan are moving toward a more operational definition of learning disabilities that includes a certain percentage of discrepancy between ability and achievement and subtest scatter. Research might be undertaken to determine the differences in subtest scatter and patterns between the WISC and WISC-R.

3. Kaufman (1975) reported the results of a factor analysis of the WISC-R at 11 age levels using the original WISC-R standardization sample. It would be interesting to compare the WISC and WISC-R factor analytic structures for the same pool of subjects.

4. An investigation might be undertaken to determine in a more precise and controlled manner the influence of the WISC or WISC-R scores on the disposition of a particular case. This issue was addressed in the present study, but in an after-the-fact manner, with obvious limitations.

5. Further investigation needs to be undertaken in the area of external validity measures of test bias and the WISC-R. Cleary's (1968), Thorndike's (1971), Darlington's (1971), and Peterson and

Novik's (1976) broader definitions and conceptions of test bias need to be studied, specifically in relation to the WISC-R. Reschly (1976) and Kaufman and Weiner (1976) addressed this issue, studying the relationship of the WISC-R to both the Reading subtest of the Wide Range Achievement Test and the Metropolitan Achievement Test.

6. Additional WISC/WISC-R studies with a variety of well-defined samples from different areas of the country are necessary before questions relating to score comparability between the WISC and the WISC-R can be answered conclusively for children in general. Further, these studies might profitably look more closely at various reasons why these differences might be occurring.

APPENDICES

APPENDIX A

LETTERS TO RESPONDENTS AND QUESTIONNAIRE

APPENDIX A

September 6, 1975

Dear Colleague:

I am a doctoral student in School Psychology at MSU as well as a School Psychologist at Ingham Intermediate School District. I am aware that you are a practicing School Psychologist and I would like to request your help in a study I am conducting.

I am interested in determining the equivalency of the WISC and WISC-R. Many School Psychologists, perhaps you among them, are suspecting that the WISC-R is yielding significantly different results than the original WISC. If, in fact, the tests are not equivalent, important questions that are raised include: Is the test identifying a different pool of special education students than it did previously? What implications might this have for the labeling of youngsters? What effect does this have on the differential diagnosis that we are required by law to make? For example, would a profile for a learning disabled child look relatively the same on both scales? I hope you agree that these questions are indeed important enough to try to answer.

I am in need of your help in terms of data collection. If you agree to participate, your responsibilities would include:

1. As part of your regular test battery, administer a WISC-R to one or more children who have been referred primarily because of concerns about their intellectual ability.

2. To these same children, also administer a WISC. (I will supply this test if you cannot locate one in your office.) We will not be using children's names in this study.

It would be most helpful at this time if you would complete and return the enclosed form in the self-addressed stamped envelope as soon as possible. As far as the number of students for whom you would submit data, I leave it up to you. Any number would be greatly appreciated. I realize that you are busy in the schools, but I hope by contacting you early in the school year this problem will be minimized. If you agree to participate, you will be first to receive a complete copy of the final results.

If you have any questions concerning this research project before you can reach a final decision, I would be quite willing to discuss the matter with you or provide you with whatever information you need. Please feel free to write or phone collect (517-351-3778 evening is the best time to call.)

Thank you very much for your consideration of this project. I do hope you will agree to participate and I look forward to hearing from you.

Sincerely,

Mark Swerdlik

MICHIGAN STATE UNIVERSITY

COLLEGE OF EDUCATION · DEPARTMENT OF COUNSELING, PERSONNEL SERVICES AND EDUCATIONAL PSYCHOLOGY

EAST LANSING · MICHIGAN · 48824

Please complete this form and return as soon as possible in the enclosed self-addressed stamped envelope.

I would be interested in participating in the WISC and WISC-R research study. I would be willing to administer the following number of WISC's to children who have been referred for problems of which intellectual ability is of primary concern and to whom I will also administer a WISC-R.

_____one_____two_____three_____four_____five_____six_____seven or more

Please include the following information about yourself:

Name:______________________________

Address:______________________________

City & State:______________________________

Zip Code:______________________________

Phone:______________________________

Number and type(s) of buildings serviced:_____elem._____J.H.S._____Sr. High

Number of students (approximately) in school district:____________________

Years of experience as a School Psychologist:____________________

I will need a WISC kit:_____Yes_____No

Thank you for your time and cooperation.

Mark Swerdlik

MICHIGAN STATE UNIVERSITY

COLLEGE OF EDUCATION • DEPARTMENT OF COUNSELING,
PERSONNEL SERVICES AND EDUCATIONAL PSYCHOLOGY

EAST LANSING • MICHIGAN • 48824

September 12, 1975

Dear

Thank you for your willingness to participate in my WISC-WISC-R equivalency study. On your return form, you indicated that you would administer both tests to _____ children.

If at all possible, please administer a WISC and WISC-R to the following children who have been referred primarily because of concerns about their intellectual ability (i.e. having difficulty meeting the academic demands of the classroom).

_______________ children within the ages 6.0 to 11.0

_______________ children within the ages 6.0 to 11.0

_______________ children within the ages 11.1 to 15.11

_______________ children within the ages 11.1 to 15.11

If you wish to do additional tests or cannot locate children within the assigned categories, please distribute your sample as equally as possible among the above categories.

The following guidelines are necessary in order to allow the final results to be as accurate and meaningful as possible.

1.) Please counterbalance the order of test administration. For example, if you test four children please administer the WISC followed by the WISC-R to two children and the WISC-R followed by the WISC to the remaining two children in your sample. Please divide the order up equally for each category listed above. For example, if you are testing two white children within the ages 6.0 to 11.0, please administer to one child the WISC followed by the WISC-R and reverse the order for the second child in that category.

2.) Please administer your second test not less than one week nor more than a month after the first test. Ideally, the closer to one month the better. This is necessary in order to control for both practice and growth effects.

3.) Please administer all subtests of both tests. This will greatly enhance the usefulness of the final results for school psychologists.

4.) Please be sure to specifically follow the directions for administration and scoring guidelines provided in the WISC and WISC-R manuals. There are many obvious and subtle differences in both administration and scoring of

the two tests. In the test manuals, note the differences in the administration of the similarities (WISC-R has no analogies), comprehension (on the WISC-R if child gives only one reason to several items, you ask him for another), digit span, coding and mazes subtests. In addition, the starting points and scoring criteria are different for various subtests. Be sure to utilize the appropriate normative tables for each test. As you probably realize, correct administration and scoring of both the WISC and WISC-R is crucial to this study.

5.) Please fill out the enclosed data recording sheet for each child as completely as possible.

6.) If it is necessary to deviate from any of the previously mentioned guidelines, please indicate this on the data recording sheet that you return for each child you test.

7.) Please return all forms by Christmas in the enclosed, self-addressed, stamped envelope provided for your convenience.

Again, thank you for your time and interest in this study. It is greatly appreciated. As soon as the data is analyzed, you will be first to receive a complete copy of the final results.

If there is ever anything I might be able to assist you with or if you have any further questions or concerns regarding this study, please do not hesitate to contact me. I look forward to hearing from you by Christmas.

Sincerely,

Mark Swerdlik

WISC-WISC-R EQUIVALENCY STUDY

Please return to: Mark Swerdlik
6243 EndenHall Way, Apt. #11
East Lansing, MI 48823

Please try to have all forms returned by Christmas

Subject #________ Examiner:________ Sex:________ Grade:________

Date of Birth:____________ CA:__________ Race:_________(Black, White, Latino)

Order of Administration: (Check one) _________WISC first _________WISC-R first

Years of experience as a school psychologist:__________

WISC Date:______ Verbal IQ:______ Performance IQ:______ Full Scale IQ:______

Verbal Scale	Raw Score	Scaled Score	Performance Scale	Raw Score	Scaled Score
Information:			Picture Completion:		
Comprehension:			Picture Arrangement:		
Arithmetic:			Block Design:		
Similarities:			Object Assembly:		
Vocabulary:			Coding:		
Digit Span			Mazes:		

WISC-R Date:______ Verbal IQ:______ Performance IQ:______ Full Scale IQ:____

Verbal Scale	Raw Score	Scaled Score	Performance Scale	Raw Score	Scaled Score
Information:			Picture Completion:		
Comprehension:			Picture Arrangement:		
Arithmetic:			Block Design:		
Similarities:			Object Assembly:		
Vocabulary:			Coding:		
Digit Span:			Mazes:		

COMMENTS:

Thank you very much for your cooperation:

Dear

Thank you very much for submitting your test data for my WISC-WISC-R equivalency study. In order for me to complete my final data analysis, I need your answers to the following questions. These questions had to be deferred until your testing was completed. I would appreciate it very much if you would promptly complete the attached form and return it in the enclosed self-addressed stamped envelope. I hope to be able to mail you a copy of the final results shortly.

Thank you,

Mark Swerdlik

1) What would you estimate our overall finding regarding WISC-WISC-R full scale (FS) IQ score differences will be? Please try to base your response on your intuitive feelings, past experience, reading of the literature and/or conversations with colleagues.

Please check one:

_______WISC F.S. IQ score higher by 10 or more points

_______WISC F.S. IQ score higher by 7-9 points

_______WISC F.S. IQ score higher by 4-6 points

_______WISC F.S. IQ score higher by 1-3 points

_______No difference between WISC and WISC-R F.S. IQ scores

_______WISC-R F.S. IQ score higher by 1-3 points

_______WISC-R F.S. IQ score higher by 4-6 points

_______WISC-R F.S. IQ score higher by 7-9 points

_______WISC-R F.S. IQ score higher by 10 or more points

---

2) How large would the Full Scale (F.S.) IQ score difference between WISC and WISC-R in each of the following F.S. IQ ranges have to be before they would affect your decisions regarding a particular case?

I. F.S. IQ range 60-75	II. F.S. IQ range 75-90	III. F.S. IQ range 90-110
_____1-2 points	_____1-2 points	_____1-2 points
_____3-5 points	_____3-5 points	_____3-5 points
_____6-8 points	_____6-8 points	_____6-8 points
_____9-11 points	_____9-11 points	_____9-11 points
_____over 11 points	_____over 11 points	_____over 11 points

3) Approximately how many psychologicals did you administer last year? ________

4) Approximately how many psychologicals do you expect to administer this year?_

5) What were the dispositions of the cases that you submitted for this study?

____

______Special Education placement (Check one)

______Mentally Impaired (EMR or EMH)
______Trainable Mentally Impaired (TMR or TMH)
______Learning Disabled (LD)
______Physically Handicapped
______Other (please specify)__________________

_______Teacher recommendations only

_______Referral for outside services (please specify)_________

5b) If you had used <u>only</u> the <u>WISC</u> scores <u>(not the WISC-R)</u> would the disposition of this case have changed? _______yes _______no

If yes, please specify in what way:

_______Special Education placement (Check one)

_______Mentally Impaired (EMR or EMH)
_______Trainable Mentally Impaired (TMR or TMH)
_______Learning Disabled (LD)
_______Physically Handicapped
_______Other (please specify)__________________

________Referral for outside services (please specify)__________________
________Teacher recommendations only

Comments:

6)

Name:____
Age:____
Highest degree earned:____
Years of training as a School Psychologist including internship:__________
Years of experience as a practicing School Psychologist:______________
Additional Comments:

Thank you very much for your time and cooperation. You will be receiving a copy of the final results shortly.

APPENDIX B

ANOVA TABLES

APPENDIX B

ANOVA TABLES

Table B1.--ANOVA table for Verbal-Performance IQ score discrepancies.

Effects	df	MS	Multivariate F	p
Between Subjects				
Age	1	.1856	.0003	.9874
Race	1	2689.2803	3.6828	.0281
Order	1	253.7045	.3474	.5567
Age x race	2	1855.7583	2.5413	.0830
Age x order	1	196.9091	.2697	.6046
Race x order	2	452.0998	.6191	.5402
Age x race x order	2	197.6158	.2706	.7639
ERROR	120	730.232978		
Within Subjects				
Test	1	3960.0682	108.0342	.0001
Age x test	1	16.5000	.4501	.5036
Race x test	2	97.9388	2.6719	.0733
Order x test	1	2152.1894	58.7136	.0001
Age x race x test	2	10.1283	.2763	.7591
Age x order x test	1	13.1856	.3597	.5498
Race x order x test	2	34.6762	.9460	.3912
Age x race x order x test	2	8.9428	.2440	.7840
ERROR	120	36.655704		
Scale	1	9554.0076	67.0901	.0001
Age x scale	1	1498.6402	10.5237	.0016
Race x scale	2	1077.0056	7.5635	.0009
Order x scale	1	159.2803	1.1185	.2924
Age x race x scale	2	19.4049	.1363	.8728
Age x order x scale	1	221.8333	1.5578	.2145
Race x order x scale	2	43.0306	.3022	.7398
Age x race x order x scale	2	27.2597	.1914	.8261
ERROR	120	142.405643		
Test x scale	1	34.0076	1.3392	.2495
Age x test x scale	1	1.5152	.0597	.8075
Race x test x scale	2	5.3107	.2091	.8116
Order x test x scale	1	686.3712	27.0285	.0001
Age x race x test x scale	2	.8611	.0339	.9667
Age x order x test x scale	1	23.6402	.9309	.3366
Race x order x test x scale	2	23.7436	.9350	.3955
Age x race x order x test x scale	2	31.1538	1.2268	.2969
ERROR	120	25.394395		

Table B2.--ANOVA tables for Verbal scale.

Source	Multivariate Tests df	Multivariate Tests F	Multivariate Tests p	Univariate Post Hoc Variable	MSH	MSE	df	F	p
Between Subjects									
Age	6,109	4.6764	.0003	Vocabulary	215.43	28.442	1,114	7.5745	.0069
				Digit Span	132.89	33.057	1,114	4.0201	.0474
Race	12,218	2.6840	.0022	Information	126.84	30.153	2,114	4.2166	.0173
				Vocabulary	88.84	28.442	2,114	3.1235	.0478
Order	6,109	.5182	.7935						
Age x race	12,218	.6585	.7900						
Age x order	6,109	.7612	.6020						
Race x order	12,218	1.0600	.3955						
Age x race x order	12,218	.6661	.7829						
ERROR	109								
Test	6,109	21.2774	.0001	Information	57.34	3.033	1,114	18.9028	.0001
				Comprehension	47.06	5.941	1,114	7.9205	.0058
				Arithmetic	22.29	3.775	1,114	5.9064	.0167
				Similarities	667.46	6.728	1,114	99.2121	.0001
				Vocabulary	34.57	3.451	1,114	10.0168	.0020
				Digit Span	140.39	5.319	1,114	26.3930	.0001
Age x test	6,109	4.6280	.0004	Information	24.77	3.033	1,114	8.1642	.0051
				Arithmetic	48.02	3.775	1,114	12.7211	.0006
Race x test	12,218	1.2602	.2441						
Order x test	6,109	2.8138	.0139	Information	42.29	3.033	1,114	13.9423	.0003
				Similarities	34.57	6.728	1,114	5.1387	.0253
Age x race x test	12,218	.8556	.5931						
Age x order x test	6,109	.3083	.4315						
Race x order x test	12,218	.7317	.7196						
Age x race x order x test	12,218	.7553	.6960						
ERROR	109								

Table B3.--ANOVA tables for Performance scale.

Source	Multivariate Tests			Univariate Post Hoc					
	df	F	p	Variable	MSH	MSE	df	F	p
Between Subjects									
Age	6,77	1.6793	.1374						
Race	12,154	2.0266	.0253	Picture Completion	119.616	27.114	2,82	4.4116	.0152
				Picture Arrangement	177.464	31.973	2,82	5.5504	.0055
				Block Design	166.821	35.209	2,82	4.7310	.0113
				Object Assembly	115.480	36.971	2,82	3.1235	.0493
				Coding	237.304	47.162	2,82	5.0316	.0087
Order	6,77	1.0414	.4054						
Age x race	12,154	1.7886	.0545						
Age x order	6,77	.1252	.9930						
Race x order	12,154	.4683	.9306						
Age x race x order	12,154	.7531	.6975						
ERROR	77								
Test	6,77	14.3608	.0001	Picture Completion	43.575	6.584	1,82	6.6187	.0019
				Picture Arrangement	88.096	8.300	1,82	10.6137	.0017
				Block Design	63.075	6.557	1,82	9.6197	.0027
				Coding	255.585	5.633	1,82	45.3738	.0001
				Mazes	78.681	7.838	1,82	10.0389	.0022
Age x test	6,77	1.7400	.1230						
Race x test	12,154	.6724	.7760						
Order x test	6,77	10.2436	.0001	Picture Completion	82.383	6.584	1,82	12.5135	.0007
				Picture Arrangement	47.755	8.300	1,82	5.7535	.0188
				Block Design	73.287	6.557	1,82	11.1773	.0013
				Object Assembly	177.032	8.171	1,82	21.6670	.0001
				Coding	131.075	5.633	1,82	23.2696	.0001
Age x race x test	12,154	.3659	.9735						
Age x order x test	6,77	.5912	.7365						
Race x order x test	12,154	.9528	.4967						
Age x race x order x test	12,154	.6771	.7716						
ERROR	77								

BIBLIOGRAPHY

BIBLIOGRAPHY

Anastasi, A. Psychological testing. 3rd ed. New York: Macmillan, 1968.

Atchison, C. O. Use of the Wechsler Intelligence Scale for Children with 80 mentally defective Negro children. American Journal of Mental Deficiency, 1955, 60, 378-79.

Ball, S., & Bogatz, G. The first year of Sesame Street: An evaluation. Princeton, N.J.: Educational Testing Service, 1970.

Barclay, A., & Carolan, P. A. A comparative study of the Wechsler Intelligence Scale for Children and the Stanford-Binet Intelligence Scale Form L-M. Journal of Consulting Psychology, 1966, 30, 563.

Bardon, J. I., & Bennett, V. School psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Berry, K., & Sherrets, S. A comparison of the WISC and WISC-R for special education students. Pediatric Psychology, in press.

Buros, K. (Ed.). Seventh mental measurement yearbook. Highland Park, N.J.: Gryson Press, 1972.

Caldwell, M. B. An analysis of responses of a southern urban Negro population to items on the Wechsler Intelligence Scale for Children. Unpublished Doctoral dissertation, Pennsylvania State College, 1954.

Cardee, B., McPherson, S., Gray, E., & Slomba, T. Unpublished survey of Ohio school psychologists. Mental Health Committee of the Ohio School Psychologists Association, 1976.

Carson, A. C., & Rubin, A. E. Verbal comprehension and communication in Negro and white children. Journal of Educational Psychology, 1960, 51, 47-51.

Carvajal, H., & McKnab, P. A comparison of the WISC and WISC-R. Unpublished paper, Emporia, Kansas, State College, 1974.

Cleary, A. T. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-24.

Cole, N. Bias in selection. Journal of Educational Measurement, 1973, 10, 237-55.

Cole, S., & Hunter, M. Pattern analysis of WISC scores achieved by culturally disadvantaged children. Psychological Reports, 1971, 29, 191-94.

Darlington, R. Another look at "cultural fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Davis, C. Comparison of WISC and WISC-R performance scale for deaf students. Personal communication.

Doppelt, J. E., & Kaufman, A. S. Estimation of the differences between WISC-R and WISC IQ's. Journal of Educational and Psychological Measurement, in press.

Dreger, R. M., & Miller, K. S. Comparative studies of Negros and whites in the U.S. Psychological Bulletin, 1960, 51, 361-402.

Evans, E. D. Contemporary influences in early childhood education. 2nd ed. New York: Holt, Rinehart and Winston, Inc., 1975.

Gehman, & Matyas, R. Stability of the WISC and Binet tests. Journal of Consulting Psychology, 1956, 20(2), 150-52.

Hamm, H., Wheeler, J., McCallum, S., Herrin, M., Hunter, D., & Catoe, C. A comparison between the WISC and WISC-R among educable mentally retarded students. Psychology in the Schools, 1976, 13(1), 4-8.

Hannon, J. E., & Kicklighter, R. WAIS versus WISC in adolescents. Journal of Consulting and Clinical Psychology, 1970, 35, 179-82.

Harnischfeger, A., & Wiley, D. Achievement test scores drop. So what. Educational Researcher, 1976, 5(3), 5-12.

Holland, W. R. Language barrier as an educational problem of Spanish speaking children. Exceptional Children, 1960, 27, 42-47.

Holroyd, R. G., & Bickley, J. Comparison of the 1960 and 1972 revisions of the Stanford-Binet L-M. Journal of Youth and Adolescence, 1976, 5, 101-104.

Hughes, R. B., & Lessler, K. A comparison of WISC and Peabody scores of Negro and white rural school children. American Journal of Mental Deficiency, 1965, 69, 877-80.

Illinois Office of Education, Pupil Personnel Services Division. Annual report of Illinois school psychologists, 1974-75. Springfield, Illinois, 1975.

Jackson, G. On the report of the ad hoc committee on educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 88-92.

Jensen, A. R. Test bias and construct validity. Paper presented at the 1975 annual meeting of the American Psychological Association, Chicago, Illinois, September 1975.

Jensen, A. R. Another look at culture fair testing. In J. Hellmuth (Ed.), Disadvantaged child. Vol. 3. New York: Bruner/Mazel, 1970.

Jensen, A. R., & Figueroa, R. A. Forward and backward Digit Span interaction with race and IQ. Journal of Educational Psychology, 1975, 67(6), 882-93.

Kaufman, A. Factor analysis of the WISC-R at 11 age levels between 6-1/2-16-1/2 years. Journal of Consulting and Clinical Psychology, 1950, 14, 99-110.

Kaufman, A. S., and Doppelt, J. E. Analysis of WISC-R standardization data in terms of the stratification variables. Child Development, 1976, 47, 165-71.

Kaufman, A., & Weiner, S. A comparison of WISC-R and WISC for black children aged 7 to 10 years. Paper presented at the annual meeting of the Eastern Psychological Association, New York City, April 1976.

Kirchev, A. A revision that really is--The WISC-R. Psychology in the Schools, 1975, 12(1), 126-28.

Klemt, L., & Petersen, L. An investigation of the continuing education habits and interests of Illinois school psychologists. Unpublished paper, January 1975.

Larrabee, G. J., & Holroyd, R. G. Comparison of WISC and WISC-R using a sample of highly intelligent children. Psychological Reports, 1976, 38, 1077-80.

Larry P. et al. v. Wilson Riles et al. United States District Court, Northern District of California, Case No. C-71-2270 RFP.

Lesiak, W. 1970-71 directory of Michigan school psychologists. Unpublished paper.

Littell, W. M. The WISC--A review of a decade of research. *Psychological Bulletin*, 1960, *57*, 132-62.

Loehlin, J. C., Lindzey, G., & Spahler, J. N. *Race differences in intelligence*. San Francisco: Freeman, 1975.

Mercer, J. *Sociocultural factors in educational labeling*. Paper presented at the NICHD Conference, Niles, Michigan, April 18-20, 1974.

Mercer, J. Sociocultural factors in labeling mental retardates. *Peabody Journal of Education*, 1971, *48*, 188-203.

Novick, M., & Peterson, N. Towards equalizing educational and employment opportunity. *Journal of Educational Measurement*, 1976, *13*, 77-88.

Ortiz, K. K. *A concurrent validation study of a Spanish version of the Pictorial Test of Intelligence and a critical analysis of the use of the WISC with bilingual American children*. Unpublished Master's thesis, California State College, 1968.

Osborne, R. T. Review of the Wechsler Intelligence Scale for Children. In K. Buros (Ed.), *Seventh mental measurements yearbook*, 1971, *1*, 802-803. Highland Park, N.J.: Gryson Press, 1971.

Owens, W. A. Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology*, 1966, *57*, 311-25.

Peterson, N., & Novick, M. An evaluation of some models for culture fair selection. *Journal of Educational Measurement*, 1976, 13, 3-29.

Prillaman, D. Class placement factors for the educable mentally retarded. *Phi Delta Kappan*, September 1973, 53.

Quereshi, M. Practice effects on the WISC subtest scores and IQ estimates. *Journal of Clinical Psychology*, 1968, *68*, 73-82.

Quereshi, M., & Miller, S. M. The comparability of the WAIS, WISC, and WBII. *Journal of Educational Measurement*, 1970, *72*(2), 105-11.

Reschly, D. J., & Davis, R. A. Comparability of WISC and WISC-R scores. *Journal of Clinical Psychology*, in press.

Reschly, D. J., Sabers, D. L., & Meredith, K. E. *Analysis of different concepts of cultural fairness using WISC-R and MAT scores from four ethnic groups*. Paper presented at the Annual Meeting of the American Educational Research Meeting, San Francisco, California, April 1976.

Rohrs & Haworth. The 1960 Stanford Binet, WISC and Goodenough tests with mentally retarded children. Journal of Mental Deficiency, 1962, 66, 853-89.

Sattler, J. E. The assessment of children's intelligence. West Philadelphia: W. B. Saunders Co., 1974.

Schuey, A. M. The testing of Negro intelligence. 2nd ed. New York: Social Science Press, 1966.

Schwarting, F. G. A comparison of the WISC and WISC-R. Psychology in the Schools, 1976, 12(2), 139-41.

Seaton, F. W. Implications of three definitions of test bias in the validation of an apprentice program selection procedure. Unpublished Master's thesis, Michigan State University, 1975.

Silverstein, A. B. Psychological testing practices in state institutions for the mentally retarded. American Journal of Mental Deficiency, 1963, 68, 440-45.

Simpson, R. Study of the comparability of the WISC and WAIS. Journal of Consulting and Clinical Psychology, 1970, 2, 156-58.

Solway, K., Fruge, E., Hays, J. R., Gryll, S., & Cody, J. Comparison of WISC and WISC-R scores for juvenile delinquents. Paper presented at the Annual Meeting of the Southwestern Psychological Association, Albuquerque, N.M., April 29-30, 1976.

Swerdlik, M. E., & Rice, W. E. WISC minus WISC-R differences for referred children. Unpublished paper, Ingham Intermediate School District, Mason, Michigan, Summer 1975.

Teahan, J. E., & Drews, E. M. A comparison of northern and southern Negro children on the WISC. Journal of Consulting Psychology, 1962, 26, 292.

Telford, C. W., & Sawrey, J. M. The exceptional child. Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Thorndike, R. Mr. Binet's test 70 years later. Educational Researcher, 1975, 4(5), 3-7.

Thorndike, R. Concepts of culture fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Tittle, C. K. Review of Wechsler Intelligence Scale for Children--Revised. Journal of Educational Measurement, 12(2), 140-43.

Tuddenham, R. D. Soldiers' intelligence in World Wars I and II. American Psychologist, 1948, 3, 54-56.

Tyler, L. E. The psychology of human differences. 3rd ed. New York: Appleton-Century-Crofts, 1965.

Webb, A. Longitudinal comparison of WISC and WAIS with EMR Negros. Journal of Clinical Psychology, 1963, 19, 101-102.

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

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

Weise, P. Current use of Binet and Wechsler tests by school psychologists in California. California Journal of Educational Research, 1960, 11, 73-78.

Whatley, R. G., & Plant, W. T. The stability of WISC IQ's for selected children. Journal of Psychology, 1957, 44, 1965-67.

Wheeler, L. R. A comparative study of the intelligence of East Tennessee mountain children. Journal of Educational Psychology, 1942, 33, 321-34.

Williams, R. The problem of the match and mismatch in testing black children. In L. Miller (Ed.), The testing of black students: A symposium. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Young, F. M., & Bright, H. A. Results of testing 81 Negro rural juveniles with the Wechsler Intelligence Scale for Children. Journal Social Psychology, 1954, 39, 219-26.

Zimmerman, I. L. WISC-R minus WISC differences for referred children. Paper presented at the 1975 Annual Meeting of the American Psychological Association, Chicago, Illinois, September 1975.

Zimmerman, I. L., & Woo-Sam, J. A comparison of the 1960 and 1972 norms of the Stanford-Binet for gifted children. Paper presented at the Annual Meeting of the American Psychological Association, Chicago, Illinois, September 1975.

Zimmerman, I. L., & Woo-Sam, J. Research with the Wechsler Intelligence Scale for Children: 1960-1970. Journal of Clinical Psychology, April 1972 (Monograph Supplement 33).