1978

A comparison of the WISC and WISC-R by order of administration

Michael Murphy

A COMPARISON OF THE WISC AND WISC-R BY ORDER OF ADMINISTRATION

A Thesis

Presented to the

Faculty of

California State College,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Psychology

---

by

Michael Murphy

June 1978

A COMPARISON OF THE WISC AND WISC-R BY ORDER OF ADMINISTRATION

A Thesis

Presented to the

Faculty of

California State College,

San Bernardino

---

by

Michael Murphy

June 1978

Approved by:

[redacted]

Chairman

June 7, 1978

Date

[redacted]

[redacted]

ABSTRACT

Forty-two subjects representing three samples of 14 children each, ages 6 through 14, were randomly selected from a private, Catholic school and tested on the 1949 WISC and on the 1974 WISC-R. The three samples corresponded to the major ethnic groups that attended the school: whites, blacks, and chicanos. Half of the children in each group were given the WISC first followed by the WISC-R and half the WISC-R first followed by the WISC. The interval between test administrations averaged 25 days. Data were presented indicating that the subjects in each group given the WISC-R first achieved significantly higher mean Verbal, Performance, and Full Scale IQs on both tests than the WISC-first groups, and highly elevated Performance and Full Scale scores on their second test encounter with the WISC. There was no evidence indicating either test as being superior to the other in terms of culture-fair testing. Coefficients of correlation for the three major scales were generally high regardless of the order of test administration. These results suggest the need to reexamine the question of the overall comparability of the WISC and WISC-R.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to Miss Patricia Faye Raymond, my loyal and dedicated assistant, who not only served as my co-examiner, but also typed this manuscript and helped in numerous, additional ways. My thanks also to Drs. Warren, Khokhlov, and O'hare for serving as members on my advisory committee and to Dr. O'brian of the computer facility for his invaluable assistance in the data analysis. I would also like to extend my warm appreciation to the staff, faculty, and students of St. Anthony's, with special thanks to Mr. Donald Murphy, my brother, who was principal of the school. Finally, I am indebted to Mr. Boris Doggin for his finely honed insights and suggestions for the preparation and writing of this thesis.

INTRODUCTION

Since it was first introduced in 1949, the Wechsler Intelligence Scale for Children (WISC) has gained general recognition as the individual intelligence test of choice for use with a wide range of children (Osborn, 1972). In developing the scale the author considered four basic areas: technical aspects (e.g., scoring, administration, and standardization procedures); sufficient diversity of subtests; some diagnostic potential among the subtests; and correlation with other composite measures of intelligence (Matarazzo, 1972). These principles are also embodied in the newly revised 1974 WISC (WISC-R), which came about as the result of practitioners' comments and criticisms of the "old" scale. However, regarding the last of the above considerations—correlation with other tests of intellect—the manual for the revised WISC reports comparisons of the WISC-R with the Wechsler Preschool and Primary Scale of Intelligence, the Wechsler Adult Intelligence Scale, and the Stanford-Binet (Form L-M, 1972 Norms) but fails to provide any information regarding perhaps the most obvious comparison, the WISC-R with the WISC, the very test the WISC-R was designed to replace (Wechsler, 1974).

Predictably, this "glaring omission" has since provided a compelling rationale for a small flurry of recently published WISC/WISC-R comparisons. To date, nearly all of these studies have reported significant differences between WISC and WISC-R IQs, with the WISC-R

consistantly producing the lower scores. The purpose of the present paper is to examine this ever-growing body of research material and to provide new evidence regarding the comparability of the WISC and WISC-R. Before doing so, however, it is important to briefly consider the tests themselves.

The WISC

The manual for the Wechsler Intelligence Scale for Children (Wechsler, 1949) described the WISC as a logical outgrowth of the Wechsler-Bellevue Intelligence Scales used for adolescents and adults. Most of the items on the WISC were taken directly from the earlier scales with the addition of easier items to permit examination of children as young as five years of age. The WISC consists of twelve individually administered subtests, of which ten are to be used for the derivation of IQs. The subtests are grouped into Verbal and Performance Scales as follows:

VERBAL	PERFORMANCE
1. General Information	6. Picture Completion
2. General Comprehension	7. Picture Arrangement
3. Arithmetic	8. Block Design
4. Similarities	9. Object Assembly
5. Vocabulary	10. Coding (or Mazes)
(Digit Span)	

The optional subtests, Digit Span and Mazes (or Coding), are considered supplementary tests to be given if time permits or as alternate tests when some other test has been invalidated. The optional

subtests were designated as such because they had the lowest correlations with their respective scales. If all subtests are administered, the scores must be prorated before IQs are computed.

Directions for scoring each of the subtests are given in the administrative sections of the manual. Some tests, like Arithmetic and Coding, are completely objective; others, like Vocabulary and Similarities require considerable evaluative judgement by the examiner. Once the tests have been given, a raw score is secured for each. Raw scores are first transmuted into normalized scaled scores within the child's own age group. Tables of scaled scores are provided for every 4-month interval between the ages of 5 and 15 years. Scaled subtest scores are then added and converted into IQs with means of 100 and standard deviations of 15. Verbal, Performance, and Full Scale IQs are all derived in this manner.

The standardization sample for the WISC included 100 boys and 100 girls at each age from 5 through 15 years. All subjects were obtained in schools, except for 55 mental retardates tested in institutions. The development of the standardization sample was carried out in 85 communities located in 11 states. The distribution of subjects conformed closely to the 1940 U.S. census for the nation at large, in terms of geographical area, urban-rural proportion, and parental occupation. Only white children were included, however.

The manual for the WISC reported split-half reliability coefficients for each subtest, as well as for Verbal, Performance, and Full Scale scores. These reliabilities were computed separately within the 7½-, 10½-, and 13½-year samples. Full Scale reliability coefficients for

the three age levels were .92, .95, and .94, respectively. The corresponding reliabilities for the Verbal Scale were .88, .96, and .96; for the Performance Scale, they were .86, .89, and .90. Reliability figures for the subtests were somewhat lower. Most were evenly distributed in the .60's, .70's, and .80's. No discussion of validity was included in the manual for the WISC.

The WISC-Revised

Swerdlik (1977) characterized the differences between the WISC and WISC-R as both obvious and subtle. The author of the tests (Wechsler, 1974) described the revision of the WISC as a synthesis of two somewhat opposing aims: the retention of as much of the 1949 WISC as possible because of its widespread use and acceptance, and the modification or elimination of items felt by some test users to be ambiguous, obsolete, or differentially unfair to particular groups of children. Specifically, five primary changes were made: 1) the WISC-R standardization sample included a proportional number of nonwhite children and is presumably more representative than the WISC; 2) the WISC-R has new administration and scoring criteria; 3) there are major and minor changes in item content; 4) the revised test has a different sequence of subtest administration; and 5) the age range has been changed from 5 through 15 years on the WISC to 6 through 16 on the WISC-R. In terms of percentages, 78% of WISC-R items are taken directly from the WISC, an additional 5.9% are from the WISC with substantial alteration, and 16.1% of WISC-R items are new. Like the WISC, the WISC-R has the same subtest format and still yields a Verbal, Performance, and Full Scale IQ with a mean

of 100 and a standard deviation of 15. In a recent study employing factor analysis, Kaufman and Van Hagen (1977) offered empirical evidence that "structurally" the old and new batteries are alike.

Comparisons of the WISC and WISC-R

Traditionally, researchers attempting to determine the usefulness of new tests have done so by comparing them to older, more established instruments. Since the publication of the WISC-R, there have been several attempts to obtain direct empirical evidence of the systematic differences (i.e., score discrepancies) and similarities (correlations) existing between the original and revised WISCs. One of the early investigations into the relationship between scores on the two tests was reported by Coven (1976), who compared WISC and WISC-R Full Scale IQs (only) for 101 elementary school children with learning difficulties. Fifty-eight of the subjects were attending classes for the educable mentally retarded, seven for the trainable mentally retarded, and 36 were enrolled in classes for children with learning disorders. All subjects were from low socioeconimic backgrounds. For this sample, the WISC-R Full Scale IQ correlated .95 with the WISC. The results of a *t* test indicated a small but significant Full Scale IQ score discrepancy of 2.63 points with the WISC-R producing the lower score. Similar data were obtained when scores were evaluated by sex and race.

Although providing some of the initial data regarding the comparability of WISC and WISC-R IQs, there were major difficulties with the design of the Coven investigation that threaten its generalizability. Such problems included the use of a highly restricted sample of children

represented only by the lower portions of the tests' standardization samples, and the fact that the WISC was always given first. Additionally, there was a two-year interval between the administration of the WISC and WISC-R that provided no control for growth effects, that is, changes that occur in children over time.

In another of the earlier WISC/WISC-R comparisons, Solway, Fruge, Hays, Cody, and Gryll (1976) compared WISC and WISC-R scores obtained from large groups of juvenile delinquents (Ns = 180 and 185) equated for age, sex, race, and grade level. Significant differences were found on six of the ten subtests used and between WISC and WISC-R IQs on the Verbal, Performance, and Full Scales. Again, the WISC-R produced significantly lower scores in all cases except the Arithmetic subtest score. Mean WISC minus WISC-R IQ discrepancies were small to moderate: 3.81 points for Verbal, 6.51 points for Performance, and 5.17 points for Full Scale scores. Difficulties with this study include a limit to the generalizability of results because of the restricted sample and the use of two separate groups of subjects, each of which took only one of the tests. The groups were assumed identical and compared by means of a *t* test. The reported WISC/WISC-R score discrepancies may, therefore, reflect differences in the two groups as well as test differences.

In contrast to the bulk of WISC/WISC-R studies which used subjects of generally low ability, Larrabee and Holroyd (1976) compared scores earned by 38 highly intelligent fifth graders on both the WISC and WISC-R. All of the children attended Polytechnic School in Pasadena, California, a private school with a reputation for academic excellence. The children, 19 males and 19 females, were of upper-middle to upper class backgrounds

with parents mostly in the professional occupations such as psychiatry, law, engineering, and teaching.

Administration of the tests was partially counterbalanced with 24 of the subjects receiving the WISC first while the remaining 14 subjects were given the WISC-R first. The interval between first and second test administrations was ten weeks. Significant WISC/WISC-R differences were reported for Verbal, Performance, and Full Scale IQs, with the WISC scores being higher in all cases. Mean differences between the tests were large: 9.6 points for the Verbal IQ, 8.4 points for the Performance IQ, and 9.4 points for the Full Scale scores. As the authors expected, coefficients of correlation between the two tests proved generally high, rendering one test, for all practical purposes, an alternate form of the other. There were no significant effects for the two orders of administration.

The findings of the Larrabee and Holroyd study, though providing much needed data for the upper ability groups, were based on a sample no less restricted than the retarded groups used in most WISC/WISC-R comparisons. Additionally, each of the two examiners gave only one type of test, either the WISC or the WISC-R. Thus, the reported score discrepancies, among the largest to date, may reflect differences in the examiners as well as test differences.

Schwarting (1976) obtained the WISC and WISC-R scores of 58 children randomly selected from a school in Omaha, Nebraska. The school had a grade span of one through eight and the subjects ranged in age from 6-15 years. Practice and growth effects were controled for by a fully counterbalanced order of administration and a test-retest interval of

approximately two months. Significant differences between the WISC and WISC-R were again reported with the WISC-R yielding the lower scores. Mean differences between the two tests were 4.86, 8.74, and 7.49 IQ points for the Verbal, Performance, and Full Scales, respectively. Though suffering many of the same problems besetting other researchers such as small sample size, this study represented a significant break from tradition in that it used randomly selected normal subjects. As Swerdlik (p.268) observed, "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 somewhat modified version of the WISC/WISC-R comparison study is one that attempts an assessment of IQ score differences between the WISC and WISC-R and then correlates these results with some other measure of IQ, academic achievement, or both. One such study was conducted by Hartlage and Steele (1977), who compared WISC and WISC-R scores for 36 seven-year old children, most of whom were black males. The authors reported WISC-R IQ scores slightly lower than scores from the WISC with small mean differences of two, one, and two points for Verbal, Performance, and Full Scale IQs respectively. Limitations of this study included the fact that the WISC was always the first test administered and the use of a 6-month test-retest interval which may not have provided adequate control for growth effects.

In another multiple-test comparison study, Brooks (1977) compared a number of tests including the WISC and WISC-R among 30 children, ages six through ten, referred for psychological evaluation. Although the design of the study incorporated a fully counterbalanced order of test

administration, both the WISC and WISC-R were given over a span of only one or two days, along with the other tests used in the comparison. It is difficult, therefore, to guage possible confounding of results due to practice and fatigue, as well as the transfer of training to the WISC and WISC-R from tests outside the Wechsler series. At any rate, *t* tests performed between the two Wechsler Scales were significant, showing the familiar pattern of moderate to large score discrepancies (VS=4.47, PS=9.27, FS=7.23) with the WISC-R again producing the lower score.

Not all WISC/WISC-R comparison studies have reported significant IQ score differences between the two tests. The exception to the rule of lower WISC-R scores was reported by Gironda (1977), who compared 20 urban educable mentally retarded students' WISC-R scores with their WISC records administered an average of three years previously. The author found no significant differences between any of the corresponding IQs. Though employing a sample of unusually small size, the study does raise serious questions concerning the outcome of WISC/WISC-R comparisons in relation to the length of the test-retest interval and, in turn, practice and growth effects.

Hamm, Wheeler, Mc Callum, Herrin, Hunter, and Catoe (1976) compared scores on the WISC and WISC-R from forty-eight 10- and 13-year old subjects matched for sex, race, and previous assignment to classes for the educable mentally retarded. Design features included a partially counterbalanced order of test administration with the WISC-R administered first to 34 children and the WISC first to 14. To control for growth effects, the test-retest interval averaged 39 days with no interval less than two weeks. The results of *t* tests revealed significantly lower

IQs on the WISC-R for this rural Georgia sample, with mean differences for the Verbal, Performance, and Full Scales of 6.0, 9.4, and 7.5 IQ points, respectively.

Hamm and his associated also compared the WISC/WISC-R Full Scale scoring patterns for the two age levels to determine if differences between the tests vary at different ages. Their results proved nonsignificant, indicating stability in WISC/WISC-R FSIQ discrepancies for the ages sampled. To determine the significance of the practice effect or positive transfer, a separate *t* test was computed to assess mean differences between WISC and WISC-R Full Scale IQ scores for the 14 subjects given the WISC first. Although the authors reported still significantly lower WISC-R scores for these 14 children, an important observation was made. It was noted that even among certain groups of retarded children, the effects of practice may substantially raise the last-given test score when two similar tests are administered.

To further evaluate the practice effect noted by Hamm and his colleagues, Davis (1977) recently reported the results of a matched pair comparison of WISC and WISC-R scores. From a much larger sample of subjects previously given both of the Wechsler Scales for unrelated assessment purposes, this investigator selected corresponding pairs of test records for 54 children. The pairs of records for the 54 subjects were chosen on the basis of Full Scale IQs on the first-given test, either the WISC or WISC-R, which could be matched within three Full Scale IQ points of a first-given complementary test. When WISC and WISC-R scores were compared with respect to the order in which the tests were administered, it was shown that the WISC-R given first sharply

elevates WISC scores, but when the WISC is first-given, the resulting WISC/WISC-R scores are essentially similar. According to Davis, these findings argue against the expectancy that all subtest scaled scores and IQs will necessarily be lower on the WISC-R than on the WISC.

The greatest value of the Davis investigation was to identify for other researchers the operation of differential practice effects dependent on order of test administration (sequence effects). Implicit in these findings was the suggestion that in some cases scores on the revised test may only appear lower if first- and second-given WISC-Rs are compared to first-given WISCs plus sharply elevated second-given WISC scores. Davis reported that these elevated second-given WISC IQs result from first exposure to the WISC-R, which represents a substantially greater learning experience than the WISC. Based on his findings, Davis concluded:

> ... the new structure of the WISC-R, particularly the instructions for administering the test, has changed the nature of the instrument so that, in contrast with the WISC, it now consists largely of a series of learning situations, and that it is not possible to obtain direct empirical evidence of systematic differences between standard scores on the two tests. (p.163)

The following WISC/WISC-R comparison, Davis notwithstanding, sought to obtain direct empirical evidence of systematic differences between scores on the two tests by incorporating a number of specific design features. Most important among these was the use of a built-in order factor (Kirk, 1968) to evaluate the influence of order of administration

and the possible sequence effects that may result. Also, because previous research with the two scales has primarily relied on special and thus restricted samples, an attempt was made to test groups of normal children of average intellectual ability. An additional consideration, one that has received relative neglect in the past, involves a comparison of the scales among children of differing ethnic backgrounds to determine if the construction of the revised test has made it less sensitive than the WISC to the ethnic differences of minority children. Thus, for the following research, some important questions to answer were: How do the WISC and WISC-R compare for normal children in light of sequence effects? Does the order in which the two tests are administered influence scores on the first-given tests? On second-given tests? Is there evidence to suggest that the revised test is more culturally fair than its predecessor, the WISC?

METHOD

Subjects

Forty-two children representing three independant samples of 14 children each were randomly selected from the student body of St. Anthony's, a private, multi-ethnic Catholic school in San Bernardino, California. The samples corresponded to the three major ethnic groups that attended the school: whites, blacks, and chicanos. Each child was identified as belonging to a particular ethnic group on the basis of three criteria: physical characteristics, school records, and the ethnic identity of the parents, parent, or guardian. To avoid confounding of cultural factors, no children of known mixed heredity or interracial family situations were used.

The subjects ranged in age from 6 yrs- 4 mos. to 14 yrs.- 8 mos. The average age for each of the three samples at the administration of the first test was: whites, 10-0 (8 males and 6 females); blacks, 11-11 (4 males and 10 females); and chicanos, 10-4 (12 males and 2 females).

The school itself has a grade span of K through 8 and is located in an economically depressed urban-residential setting which carries the designation ESEA Title I target area. Fifty-three percent of the student body scored below the national norm for reading and/or arithmetic on the SRA Achievement Series administered in the fall of 1976. The major occupations among the parents of attending students were in the semiskilled or unskilled areas. The school served many single

parent families. Approximately two thirds of the St. Anthony's student body was of the Catholic faith.

Procedure

Half of the children within each ethnic group were randomly allotted to one of the two orders of test administration, WISC followed by WISC-R and WISC-R followed by WISC, and half to the other. To control for growth effects, a test-retest interval between first and second test administrations was imposed averaging 25 days with a range of from 17 to 35 days. All children received both tests in quiet, comfortable quarters located in the school's convent. The particular testing room assigned to each child was held constant over both test administrations to control for the effects a changed environment may produce on test scores. Standard administration procedures were used according to the manuals for each test.

Two examiners administered all tests. They were the author, a trained white male experienced with both tests, and a white female elementary school teacher and graduate psychology student with special training in administration procedures for each test. The method for the assignment of examiners to subjects was as follows: within each ethnic group, examiner one was arbitrarily assigned to test three of the subjects from the WISC-first order and four subjects from the WISC-R-first order. Examiner two, on the other hand, observed the remaining four subjects from the WISC-first group and the three from the WISC-R-first group. This pattern was repeated within each of the three samples. The same examiner administered both tests to a

particular child, and each examiner observed the same number of children within ethnic groups. The 84 WISC and WISC-R protocols were scored and rescored by the author after all identifying information was removed from the front of each record and placed elsewhere. Though an occasional unique response might bring to mind the identity of a particular subject, scoring was generally accomplished without knowledge of an individual's sex or race.

RESULTS

WISC and WISC-R Verbal, Performance, and Full Scale IQs were computed for all subjects from the 10 regular subtests used for the derivation of IQs. The optional subtests Digit Span and Mazes were excluded from the comparison. Table 1 reports the mean IQs and standard deviations obtained for each of the three groups on the WISC and WISC-R by order of administration and by test independent of order of administration (orders combined). A three-factor analysis of variance design was employed to assess differences between these means for ethnic groups, tests, and the two orders of test administration.

TABLE 1

Means and Standard Deviations for Verbal, Performance, and Full Scale WISC and WISC-R IQs for Groups by Order of Administration

Ethnic Group	Order of Administration		Verbal Scale WISC	Verbal Scale WISC-R	Performance Scale WISC	Performance Scale WISC-R	Full Scale WISC	Full Scale WISC-R
Whites	WISC first	M	101.4	98.9	103.1	106.7	102.4	102.6
		SD	11.7	11.2	16.0	16.2	13.3	13.9
	WISC-R first	M	112.1	112.4	133.3	109.1	124.4	112.3
		SD	10.2	12.4	12.6	9.5	10.1	10.8
	Combined	M	106.8	105.6	118.2	107.9	113.4	107.4
		SD	11.9	13.4	20.9	12.8	16.1	13.0
Blacks	WISC first	M	95.1	99.3	89.6	95.1	91.9	97.0
		SD	9.8	7.9	9.2	5.3	9.5	5.7
	WISC-R first	M	106.9	103.8	112.9	96.9	110.6	100.4
		SD	10.5	10.4	13.7	16.3	8.8	12.0
	Combined	M	101.0	101.6	101.2	96.0	101.2	98.7
		SD	11.5	9.2	16.5	11.6	13.1	9.2
Chicanos	WISC first	M	97.9	97.1	104.3	106.4	101.0	101.4
		SD	19.0	9.7	14.4	14.6	15.8	11.9
	WISC-R first	M	106.6	107.6	127.0	108.3	118.1	108.9
		SD	14.0	11.7	8.4	11.4	12.2	12.3
	Combined	M	102.2	102.4	115.6	107.4	109.6	105.1
		SD	16.7	11.6	16.4	12.6	16.2	12.3

The results of the analysis of variance of Verbal IQs are summarized in Table 2. These data indicate significant differences in Verbal IQs between the two orders of test administration. Subjects in each ethnic group assigned to the WISC-R-first order of administration obtained significantly higher Verbal scores on both tests than subjects to whom the WISC was first administered. WISC-R-first subjects averaged 10.4 VIQ points higher on the WISC and 9.5 VIQ points higher on the WISC-R than the WISC-first subjects. The discrepancy between first-given tests averaged 9.9 VIQ points, with the WISC-R yielding the higher score.

The results of the analysis showed no significant variation in Verbal IQs among the three ethnic samples. Mean WISC and WISC-R scores for the three groups averaged 106.2, 101.3, and 102.3 VIQ points for whites, blacks, and chicanos, respectively. When first- and second-

TABLE 2

Analysis of Variance of Verbal Scores

Source	SS	df	MS	F
1 Mean	895693.8	1	895693.8	
2 A (race)	380.0952	2	190.0476	.7932
3 C (order)	2080.048	1	2080.048	8.6814 **
4 B (test type)	.4285714	1	.4285714	.0103
5 AC	57.52381	2	28.76190	.1200
6 AB	11.14286	2	5.571429	.1344
7 CB	3.857143	1	3.857143	.0930
8 S(AC)	8625.571	36	239.5992	
9 ACB	104.8571	2	52.42857	1.2644
10 SB(AC)	1492.714	36	41.46429	

** $p < .01$

given WISC IQs were compared to first- and second-given WISC-R IQs, there were no significant differences between test types on the Verbal Scales. Interactions of race and test type, race and order of administration, and test type and order of administration also proved nonsignificant.

TABLE 3

Analysis of Variance of Performance Scores

Source	SS	df	MS	F
1 Mean	974814.3	1	974814.3	
2 A (race)	3527.167	2	1763.583	5.9991 **
3 C (order)	3936.012	1	3936.012	13.3889 **
4 B (test type)	1320.107	1	1320.107	42.2703 **
5 AC	70.88095	2	35.44048	.1206
6 AB	91.35714	2	45.67857	1.4626
7 CB	2870.012	1	2870.012	91.8987 **
8 S(AC)	10583.14	36	293.9762	
9 ACB	49.73810	2	24.86905	.7963
10 SB(AC)	1124.286	36	31.23016	

** $p < .01$

Table 3 reports the results of the analysis of variance of Performance IQs. These data indicate that mean differences in Performance scores between the two orders of test administration were again significant, with higher scores on both tests for the WISC-R-first order. Subjects in each group given the WISC-R first averaged 25.4 PIQ points higher on the WISC and 2 PIQ points higher on the WISC-R than the WISC-first subjects. The mean discrepancy between first-given tests was 6.4 PIQ points, with the WISC-R producing the higher score.

Differences in Performance scores among the ethnic samples were also significant. Mean WISC and WISC-R PIQs for the three groups averaged 113, 98.6, and 111.5 PIQ points for whites, blacks, and chicanos, respectively. An overall comparison of tests from both orders of administration showed significant mean differences between test types on the Performance Scales, with the WISC-R averaging 7.9 PIQ points lower than the WISC. The results also indicated a significant interaction of test type and order of administration on the Performance Scales, which is illustrated in Figure 1b. There were no significant interactions of race and test type or race and order of administration.

The analysis of variance of Full Scale IQs is presented in Table 4. These results indicated significantly higher mean scores on both tests for subjects assigned to the WISC-R-first order of administration. Subjects given the WISC-R first averaged 19.3 points higher on the WISC and 6.9 points higher on the WISC-R than WISC-first subjects. The mean discrepancy between first-given tests was 8.8 FSIQ points, with the WISC-R yielding the higher score.

Full Scale score differences among the three ethnic groups were significant. Mean WISC and WISC-R FSIQs for each group averaged 110.4, 100, and 107.4 for whites, blacks, and chicanos, respectively. The overall comparison of FSIQs from both orders of administration showed significant differences between test types with mean WISC-R scores an average of 4.31 points lower than WISC FSIQs. There was also a significant interaction of test type and order of administration, which is shown if Figure 1c. There were no significant interactions of race and test type or race and order of administration on the Full Scales.

TABLE 4

Analysis of Variance of Full Scale Scores

Source	SS	df	MS	F
1 Mean	942340.6	1	942340.6	
2 A (race)	1620.167	2	810.0833	3.3467 *
3 C (order)	3588.107	1	3588.107	14.8235 **
4 B (test type)	390.0119	1	390.0119	13.6542 **
5 AC	86.64286	2	43.32143	.1790
6 AB	43.02381	2	21.51190	.7531
7 CB	810.9643	1	810.9643	28.3916 **
8 S(AC)	8714.000	36	242.0556	
9 ACB	27.21429	2	13.60714	.4764
10 SB(AC)	1028.286	36	28.56349	

* $p < .05$ ** $p < .01$

Figure 1 illustrates the effects of order of administration on the Verbal Scales (a) and the combined order and interaction effects on the Performance (b) and Full Scales (c). For the sake of simplicity, these results are presented across all subjects. Close inspection of Table 1, however, will verify the general uniformity of the pattern with each of the ethnic samples.

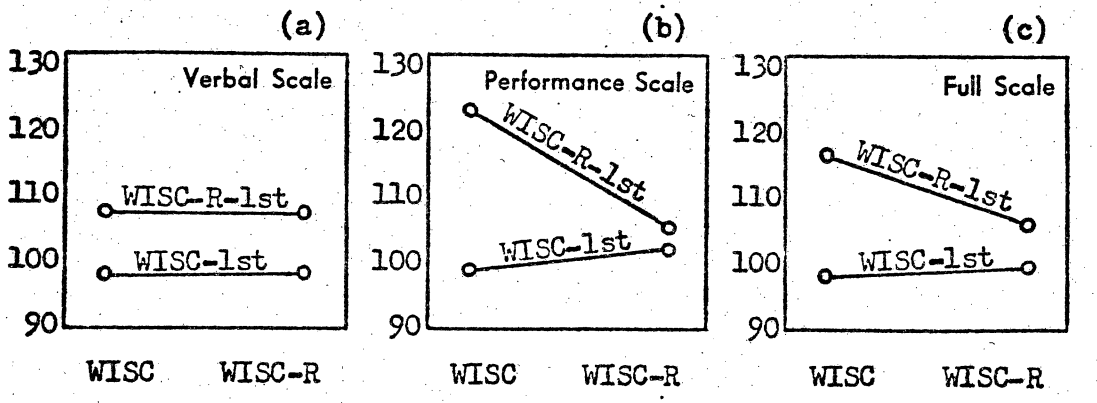

Figure 1. Overall Order Effects for Verbal IQ and Combined Order and Interaction Effects for Performance and Full Scale IQs.

Previous investigations of the WISC and WISC-R have often employed coefficients of correlation as an additional means of comparison of the two scales. To provide comparable data, Table 5 reports Pearson correlations of WISC and WISC-R subtest scale scores and IQs by order of administration. These data indicate generally high coefficients of correlation for the Verbal, Performance, and Full Scales regardless of the order of test administration.

TABLE 5.

Pearson Correlations Between WISC and WISC-R IQs and Scaled Scores for All Subjects by Order of Administration

Subtest and Scale	WISC-first	WISC-R-first
Information	.795	.740
Similarities	.744	.653
Arithmetic	.519	.561
Vocabulary	.616	.684
Comprehension	.694	.524
Picture Completion	.728	.804
Picture Arrangement	.201	.423
Block Design	.729	.728
Object Assembly	.792	.588
Coding	.696	.771
Verbal IQ	.751	.699
Performance IQ	.813	.864
Full Scale IQ	.827	.811

DISCUSSION

The present study compared the WISC and the WISC-R among samples of normal children to determine if the order in which the instruments were administered had an influence on test scores. The results indicate that order of administration plays a significant role in the assignment of IQ scores to normal children. Data were presented showing that when the two orders of test administration are combined, a situation analogous to counterbalancing, the familiar pattern of generally lower WISC-R scores obtains. In this case, both mean Performance and Full Scale WISC-R IQs were significantly lower than complementary WISC IQs for each group by an average of 7.9 and 4.3 IQ points, respectively. However, by order of administration, those subjects in each group randomly assigned to the WISC-R-first order achieved significantly higher mean IQs on both tests than did the three WISC-first groups. In other words, first given WISC-R Verbal, Performance, and Full Scale IQs were actually higher than first given complementary WISC IQs in three independent samples. Thus, despite past reports to the contrary, these results provide the first indication that for most children the WISC-R will yield significantly higher IQs than would otherwise have been obtained through the use of the WISC.

Hannon and Kicklighter have previously stated (1970, p.182) "The precise effects of order of administration are difficult to determine ..." The present case was no exception. Not only did the results of

this study support an hypothesis of order effects clearly favoring WISC-R-first subjects, there were significant interactions of test type and order of administration on the Performance and Full Scales (figure 1). In addition to significantly higher scores on both tests, WISC-R-first subjects also achieved markedly elevated Performance and Full Scale IQs on their second test encounter with the WISC. These unexpected findings suggest a relationship between IQ scores on the WISC and WISC-R that is far more complex than heretofore reported.

In attempting to explain the obtained results, it seems reasonable to focus on some of the many subtle differences existing between the two scales. Based on his results, Davis argued that certain provisions in the instructions for the newer test tend to promote learning, much of which will be consolidated during reminiscence and demonstrated when similar items are presented later on the WISC. The present findings, however, suggest that extra learning resulting from first exposure to the WISC-R is immediately consolidated and demonstrated on both tests, whereas first-given WISCs tend to promote a much less effective learning-set that will also influence scores on both tests. Of particular relevance here are the general scoring rules for the revised test only (Wechsler, 1974, p.60). According to these instructions, an examiner may repeat items to which the child said "I don't know" if the child gives correct responses to more difficult items on the same subtest. Similarily, should the child refuse an item by saying "I can't do it", or if he discontinues an item before the time limit is up, the examiner may "gently urge" the child to proceed. It

is possible that when a child does respond correctly to an item that he initially preceived as being above his level, the intrinsic satisfaction and resultant examiner praise may provide enough reward to motivate the child towards more vigorous effort on following items and also to give answers of which he is unsure but which may be correct nonetheless. Carry-over effects on the WISC, which places a high premium on spontaneity, might then result in elevated scores when that test is given second. The underlying implication is that the learning experience represented by the initial test, either the WISC or WISC-R, tends to foster an approach to test taking that will influence the scores on both the first- and second-given tests. These tentative findings raise serious questions concerning the interpretability of not only the previous investigations of the WISC and WISC-R, but all test comparison studies where sequence effects may produce uncontrolled distortions in the final outcome. Caution must be exercised, though, in generalizing the present results to all children represented by the WISC and WISC-R standardization samples because the nature of the sequence effects may be highly variable in other samples of differing ability, geographical location, socioeconomics, test-retest intervals, and the like.

A second aspect of the present study was to compare the WISC and WISC-R among children of differing ethnic backgrounds in an attempt to determine if one of the scales might assess minority children more favorably than the other. A comparison of this nature seemed especially warranted in the case of the WISC and WISC-R for two reasons. First,

the revised test, unlike the original WISC, was deliberately constructed for multi-racial assessment. Secondly, and perhaps more important, is the increasing popular disenchantment over the use of the 1949 WISC with certain groups of children. A good case in point was the nationally publicised federal class action law suit* and its threatened government ban of the WISC and other tests on the grounds that the use of racially discriminatory assessment tools violates federal law.

The results of the present study suggest that neither test provides differentially more favorable scores for minority children. The data presented in Tables 2, 3, and 4 reveal no significant interactions of race and type of test on any of the three major scales, indicating that relative score differences between ethnic groups remained substantially intact from one test to the other. Thus, despite the inclusion of nonwhite children in the standardization sample, the elimination of items of questionable cultural parity, and the use of obviously non-white human figures in many items, there is no evidence that the WISC-R is any more (or less) culturally fair than its predecessor, the WISC. It is important to add that these findings should not be viewed as necessarily reflecting negatively on the WISC-R, because an understanding of the constituents of culture-fair testing is far from complete.

---

* Larry, P., Et. Al., Plaintiffs, vs. Wilson Riles, Superintendent of Public Instruction for the State of California, Et. Al., Defendants, No. C-71-2270-RFP, United States District Court Northern District of California, San Francisco. For those interested in background information on this important legal controversy, see: The New York Times, Oct. 12, 1977, p.14; San Francisco Chronicle, Oct. 12, 1977, p.17; The San Diego Union, Oct. 12, 1977, p.A-6; San Bernardino Sun-Telegram, Oct. 13, 1977, p.A-6, and Oct. 23, 1977, p.A-4. Also see Appendix B of this paper.

In connection with the previously mentioned legal battle, for example, two items that are commonly drawn from the WISC by its opponents to exemplify so called "culturally biased" test questions are "What is the thing to do if a fellow (girl) much smaller than yourself starts to fight with you?" and "Why is it better to pay bills by check than by cash?" In the present samples, however, both of the minority groups achieved a higher combined raw score point total for these two questions than did the white group.

Some researchers (e.g. Sewell, 1977) have suggested the use of a particular assessment tool among certain minority group children because of IQ estimates that are higher and, therefore, perhaps more appropriate. For those following this line of reasoning, the present data favor the use of the WISC-R for minority populations, not because it is more fair than the WISC, but because first-given WISC-R scores were higher for all groups than first-given WISCs.

The most striking finding of the present study was the impact of the order of test administration. The results generally support Davis' conclusion that the order of administration has a significant effect on the differences between scores on the WISC and WISC-R. But unlike the Davis study in which subjects were matched on the basis of first-given tests, the present data revealed moderate to large descrepancies between first-given WISCs and WISC-Rs, with the WISC-R producing the higher scores. There is little doubt that past investigators who have reported lower scores for the WISC-R have collectively persuaded many psychologist-practitioners into believing that the WISC-R consistantly

produces lower scores. After reviewing some of the published and many of the unpublished WISC/WISC-R comparison studies, Swerdlik (p.268) concluded "Significantly different scores resulting from the WISC and 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." Moreover, one investigator (Schwarting, 1976), after reporting that the WISC-R yields significantly lower scores, offered the practitioner regression equations to predict WISC-R IQs from WISC scores. The results of the present study, however, clearly indicate the need to reexamine the question of the overall comparability of the WISC and WISC-R in light of the generally neglected problem of sequence effects.

APPENDIX A

Parental Consent Form

ST. ANTHONY SCHOOL
1510 West 16th Street
San Bernardino, California

Dear Parents:

As you recall, an article appeared in the February 22nd issue of the Tuesday Times regarding a thesis project to be conducted by graduate students from the California State College at San Bernardino. The aim of the study is simply to determine which of two widely used children's intelligence tests is the better. In order to make this determination, we need to administer both tests to a good number of students from St. Anthony's. The tests, the WISC and the WISC-R, require no writing on the part of the children, and are found by most children (and adults for that matter) to be interesting, challenging, and enjoyable to take. Would you please help us in our effort by allowing your child to undertake approximately two hours of testing during regular school hours, between now and the end of the school term?

Yes - My child_____

in grade____ may participate in the project.

I understand that no names will be used and that test results will be used solely for research purposes.

_____

Parent Signature

Sincerely,

Donald Murphy

Donald Murphy
Principal

APPENDIX B

Legal Documents

November 9, 1977

Michael Murphy
610 W. 40th Street
San Bernardino, Ca. 92407

Dear Mr. Murphy:

Enclosed please find 2 documents, one relating to the name of the case and case number, the other relating the names and addresses of the attorneys involved.

The witnesses to date that have testified in this case are as follows: Jane Ross Mercer, Gerald West, Darryl Lester (one of the plaintiffs in the action) Lucille Lester, Gloria Johnson Powell, M.D., Leon J. Kamin, Asa Grant Hilliard, III and George Wilson Albee. There are 2000 pages of transcript to date. The cost would be $.25 per page for a copy of the transcript. The trial is expected to last into December.

I would suggest if you have any questions regarding different contentions in this trial you contact one of the attorneys involved.

Very truly yours,

Roberta L. Rogers
Official Reporter

THE UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

BEFORE THE HONORABLE ROBERT F. PECKHAM, CHIEF JUDGE

LARRY P., ET AL., PLAINTIFFS, VS. WILSON RILES, SUPERINTENDENT OF PUBLIC INSTRUCTION FOR THE STATE OF CALIFORNIA, ET AL., DEFENDANTS.	)	NO. C-71-2270-RFP

REPORTERS' DAILY TRANSCRIPT

REPORTED BY:

ROBERTA L. ROGERS
LAWRENCE WHITE

A P P E A R A N C E S:

FOR PLAINTIFFS:

ARMANDO M. MENOCAL, III, ESQ.
ERICA BLACK GRUBB, ESQ.
WILLIAM G. HARRIS, ESQ.
PUBLIC ADVOCATES, INC.
433 TURK STREET
SAN FRANCISCO, CA. 94102

LOWELL JOHNSTON
NAACP LEGAL DEFENSE AND EDUCATION FUND, INC.
12 GEARY STREET
SAN FRANCISCO, CA. 94108

STEPHEN S. DUNHAM, ESQ.
PALMER BROWN MADDEN, ESQ.
SAMUEL R. MILLER, ESQ.
MORRISON & FOERSTER
ONE MARKET PLAZA
SUITE 4000
SAN FRANCISCO, CA. 94105

FOR STATE DEFENDANTS:

EVELLE J. YOUNGER. ATTORNEY GENERAL
BY: JOANNE CONDAS, ESQ.
DEPUTY ATTORNEYS GENERAL
350 MCALLISTER STREET, RM 6000
SAN FRANCISCO, CA. 94102

FOR UNITED STATES:

JAMES L. BROWNING, JR.
UNITED STATES ATTORNEY
BY: GEORGE CHRISTOPHER STOLL
ASSISTANT UNITED STATES ATTORNEY
CIVIL DIVISION
450 GOLDEN GATE AVENUE
SAN FRANCISCO, CA. 94102

LOUIS M. THRASHER, ESQ.
ARTHUR E. PEABODY, JR., ESQ.
LEONARD RIESER, ESQ.
LUCY L. THOMSON, ESQ.
OFFICE OF SPECIAL LITIGATION
CIVIL RIGHTS DIVISION
U. S. DEPARTMENT OF JUSTICE
WASHINGTON, D.C. 20530

REFERENCES

BROOKS, C. R. WISC, WISC-R, S-B L&M, WRAT: Relationships and trends among children ages six to ten referred for psychological evaluation. Psychology in the Schools, 1977, 14, 30-33.

COVEN, T. M. Comparability of WISC and WISC-R Full Scale IQs for elementary school children with learning difficulties. Psychological Reports, 1976, 39, 280.

DAVIS, E. E. Matched pair comparison of WISC and WISC-R scores. Psychology in the Schools, 1977, 14, 161-66.

GIRONDA, R. J. A comparison of WISC and WISC-R results of urban educable mentally retarded students. Psychology in the Schools, 1977, 14, 271-75.

HAMM, H., WHEELER, J., Mc CALLUM, 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, 4-8.

HANNON, J. E., & KICKLIGHTER, R. WAIS versus WISC in adolecents. Journal of Consulting and Clinical Psychology, 1970, 35, 179-82.

HARTLAGE, L. C., & STEELE, C. T. WISC and WISC-R correlates of academic achievement. Psychology in the Schools, 1977, 14, 15-18.

KAUFMAN, A. S., & VAN HAGEN, J. Investigation of the WISC-R for use with retarded children: Correlation with the 1972 Stanford-Binet and comparison of WISC and WISC-R profiles. Psychology in the Schools, 1977, 14, 10-14.

KIRK, R. E. Experimental Design: Procedures for the Behavorial Sciences. Belmont, Ca.: Brooks/Cole Pub. Co., 1968, p.317.

LARRABEE, G. J., & HOLROYD, R. G. Comparison of WISC and WISC-R using a sample of highly intelligent children. Psychological Reports, 1976, 38, 1071-74.

MATARAZZO, J. D. Wechsler's Measurement and Appraisal of Adult Intelligence. Baltimore: Williams & Wilkins Co., 1972.

OSBORN, R. T. Review of the Wechsler Intelligence Scale for Children In K. Buros (Ed.), Seventh Mental Measurements Yearbook. Highland Park, N.J.: Gryphon Press, 1972, 1, item 432, 802-3.

SCHWARTING, F. G. A comparison of WISC and WISC-R. Pscyhology in the Schools, 1976, 13, 139-41.

SEWELL, T. E. A comparison of the WPPSI and Stanford-Binet Intelligence Scale (1972) among lower SES black children. Psychology in the Schools, 1977, 14, 158-61.

SOLWAY, F. G., FRUGE, J., HAYS, R., CODY, J., & GRYLE, S. A comparison of the WISC and WISC-R in a juvenile delinquent population. Journal of Psychology, 1976, 94, 101-6.

SWERDLIK, M. E. The question of the comparability of the WISC and WISC-R: Review of the research and implications for school psychologists. Psychology in the Schools, 1977, 14, 260-70.

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

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