1970

The Relationship of the Wechsler Preschool and Primary Scale of Intelligence to the 1960 Revision of the Stanford-Binet Intelligence Scale: Form L-M

James William Futterer
Loyola University Chicago

THE RELATIONSHIP OF THE WECHSLER PRESCHOOL AND PRIMARY SCALE OF INTELLIGENCE TO THE 1960 REVISION OF THE STANFORD-BINET INTELLIGENCE SCALE: FORM L-M

by

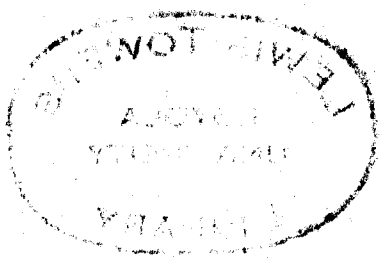
James William Futterer

A Thesis Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

June

1970



ACKNOWLEDGMENTS

Special gratitude is respectfully and gladly offered to Patricia Barger, Ph.D. and to Jeanne Foley, Ph.D. who served as advisors for this thesis and without whose guidance and direction it would not have been possible. Many thanks are also due to the secretarial staff of the Loyola University Child Guidance Center for their time and patience in arranging testing appointments.

Finally, a special thank you is due my wife, Bobbi, for her help in typing this thesis and its many rough drafts as well as for providing her unwavering support and confidence in this endeavor.

LIFE

James William Futterer was born September 11, 1945, in Joliet, Illinois. He graduated from Joliet Catholic High School in June 1963, and received his Bachelor of Arts in Psychology from Lewis College, Lockport, Illinois in June 1967.

The author entered the graduate program in Clinical Psychology at Loyola University of Chicago in September 1967. He served as an assistant to Dr. Roderick W. Pugh for one year. From October 1968 to April 1969 he served his clerkship at the Veterans Administration Westside Hospital, Chicago, Illinois. In October 1969 the author was commissioned an officer in the United States Army Medical Service Corps and assigned to continue his studies toward the degree of Doctor of Philosophy. He is currently engaged in this endeavor at Loyola University of Chicago.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The purpose of this study was to investigate the relationship of the WPPSI to the 1960 S-B. The three specific hypotheses investigated were that the WPPSI and the S-B are significantly related, that differences between WPPSI and S-B IQ scores are positively related to S-B scores in that the higher the S-B score the larger the difference, and that differences between WPPSI and S-B IQ scores are negatively related to chronological age in that the younger the age the larger the difference. The Ss consisted of 50 boys and 50 girls between the ages of 4-5 and 6-5 years with two boys and two girls at each month level. Each S was individually administered the WPPSI and S-B in a counterbalanced design by one of four male examiners.

Person product-moment correlations were computed between the WPPSI and S-B scores and between the S-B scores and the differences between S-B and WPPSI scores. In addition, tests of significance between related means were performed between the mean S-B and WPPSI scores. Finally, Chi-squares were used to analyze the cross-tabulations of chronological age and differences between S-B and WPPSI scores.

The results of the statistical analysis indicated that the WPPSI and S-B were significantly related but not interchangeable so that caution must be used when employing the WPPSI scores for clinical judgment. The results also indicated that differences between WPPSI and S-B scores were related to the S-B scores in that the higher the S-B score the larger the difference. However, the results failed to support the hypothesis that chronological age is related to differences between S-B and WPPSI scores in that the younger the age the larger the difference. It was cautioned that these results are limited in their generalizability by the small geographic region from which the sample was selected and possibly by the age, sex, and experience of the examiners.

CHAPTER I

Introduction

The importance of intelligence in determining an individual's future in society today is increasing on a rate at par with society's increasingly technical nature. Our educational system has recently taken cognisance of the fact that all children's levels of intelligence and learning abilities are not equal. The child with a special intellectual difficulty is no longer thrown into the educational ring with the child of average or better ability. Educators have discovered that if given proper attention and educational procedures geared to their needs, the children with special difficulties can advance to levels never before thought possible. However, in order to be given the advantages available in the special program, the child with intellectual difficulties must be identified at an early age for it has also been found that in order to gain maximum benefits from special education classes the child should enter them when he first begins school or even earlier.

Recently, Wechsler (1967) has developed a new test designed to assess a child's intellectual level of functioning on a preschool basis, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). If this new test proves to be as valuable an assessment instrument at a preschool

level as the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Adult Intelligence Scale (WAIS), it will indeed be a welcome addition to the repertoire of assessment instruments. In his WPPSI manual, Wechsler describes his new test and points to the fact that the WPPSI and WISC are closely related yet different Scales:

> It (the WPPSI) continues the methodological and theoretical approaches to the measurement of mental ability which were the guiding principles in the construction of the Wechsler Intelligence Scale for Children (WISC). Like the WISC, it consists of a battery of subtests, each of which when treated separately may be considered as measuring a different ability, and when combined into a composite score, as a measure of overall or global intellectual capacity (Wechsler, 1967, pp. 1-2).

Wechsler continues,

> The WPPSI is at once an extension of the WISC and a separate scale designed to cope more effectively with the psychometric problems presented in testing the 4- to 6 1/2 year-old child. It consists of eleven tests, six Verbal and five Performance. Eight of the tests provide the same measures as the WISC, and may be seen as continuous with the WISC. Only five of the Verbal tests are used in determining the Verbal Score. Sentences is to be used as a supplementary test . . . Unlike the WISC, the WPPSI is administered with the Verbal and Performance tests intermixed (Wechsler, 1967, p. 7).

The subtests that Wechsler refers to above are as follows: for the Verbal Scale, Information, Vocabulary, Arithmetic, Similarities, Comprehension, and, as a supplementary test, Sentences; for the Performance Scale, Animal House, Picture Completion, Mazes, Geometric Design, and Block Design. All of the above subtests are extensions from the WISC except for the Sentences, Animal House, and Geometric Design. These

three tests have replaced four WISC tests (Digit Span, Picture Arrangement, Object Assembly, and Coding) which could not be carried over for various reasons.

Wechsler also points out that,

> The WPPSI (as do the WISC and WAIS) retains the Intelligence Quotient (IQ) as the most effective measure or way of expressing a child's mental endowment relative to children of his own age. WPPSI IQs, like those provided by the WISC, are deviation IQs, that is, measures of relative position calculated in terms of the degree to which a child's score differs from the mean of his age group, rather than by the historic MA/CA approach (Wechsler, 1967, p. 5).

However, before the new Wechsler test can become a worthwhile addition to the assessor's tools, it must be established as a valid instrument. While Wechsler has taken great pains in his standardization procedures to be comprehensive, this is not sufficient grounds for unqualified acceptance of the WPPSI as a valid instrument. The American Psychological Association recommends, in its Standards for Educational and Psychological Tests and Manuals (1966), methods for determining test validity. A simple procedure recommended for investigating what a test measures is to compare it with other more established tests which measure the same variable as the new test. The general purpose of this research is to make such a comparison between the WPPSI and the more established 1960 Stanford-Binet Form L-M and thus provide new evidence concerning the validity of the WPPSI. A more formal statement of the hypotheses to be specifically investigated follows the Review of Related Literature chapter of this paper.

CHAPTER II

Review of Related Literature

Since its publication, little research has been reported on the WPPSI. That research which has been reported has not been of the highest quality and has dealt with such things as short forms (Silverstein, 1967, 1968a, 1968b), assessing reading achievement (Plant & Southern, 1968), sex differences (Herman, 1968), upward extension of the IQ tables (Silverstein, 1968c), and the development of a screening test (Boyd & Means, 1968). This type of research on the WPPSI seems a bit premature, since before the above mentioned studies become meaningful, information on the validity and reliability of the test must be assembled.

Wechsler has provided some such data in his manual (Wechsler, 1967). Wechsler presents the coefficients of correlation of scaled scores and IQs of three other intelligence tests with the WPPSI for 98 children between 5 and 6 years of age. The coefficients of correlation reported between the WPPSI and the Stanford-Binet Form L-M were as follows: Verbal IQ and S-B, .76; Performance IQ and S-B, .56; and Full Scale IQ and S-B, .75. The coefficients of correlation reported with the Peabody Picture Vocabulary Test (PPVT) were as follows: Verbal IQ and PPVT, .57; Performance IQ and PPVT, .44; and Full Scale IQ and PPVT, .58. The coefficients of correlation reported between the Pictorial Test of Intelligence

(PTI) and the WPPSI were as follows: Verbal IQ and PTI, .53; Performance IQ and PTI, .60; and Full Scale IQ and PTI, .64. The meaning of the reported correlations between the WPPSI and the Stanford-Binet will become clearer when the sample used in this study is more closely examined.

Wechsler reported that the subjects used were 98 children between the ages of 60 and 73 months, enrolled in a single school in the Alum Rock Union Elementary School District, San Jose, California. While providing some evidence on the validity of the WPPSI this study seems to be of limited value in terms of the general validity of the test.

Wechsler failed to provide some information about this sample that is necessary for judging its adequacy, and, at the same time, noted many restrictions on its generalizability. First, Wechsler failed to report a breakdown of the sample by sex and age. In his tables of Scaled Score Equivalents of Raw Scores, Wechsler provided norms at quarter year increments from ages four to six and one half, while in this study the reader was told only that the sample consisted of children between the ages of five and six. In addition to failing to break the age data into quarter years, this study has not even included all age groups for which the test is intended. The exclusion of the younger age range in this study was a serious deficiency, for at such young ages rapid developmental changes may influence test performance and thus change the coefficients of correlation.

A second weakness of this study was the limited population sampled. All subjects were drawn from a single school in California. Yet this test is intended for use throughout the entire United States. This study does not shed sufficient light on the validity of the WPPSI for use with subjects from widely diversified geographic regions.

Two other areas in which Wechsler failed to report any breakdown of data or even to provide any information at all were the areas of race and social-economic status. While he described his standardization population in these terms, Wechsler failed to do so with the sample for this study.

Finally, the limitations of this sample were emphasized by the fact that the WPPSI mean for this sample fell 10 points below the general population mean of 100 and the SD of this sample fell 9 points below the general WPPSI population SD of 21.

While this study is admittedly a limited one, it is at present one of the three concurrent validity studies of the WPPSI that have been reported. Wechsler, however, was not blind to the study's limitations and has urged more studies of this nature with different populations. This, it will be seen later, is the basic purpose of the present research.

One of the studies mentioned at the beginning of this chapter (Plant & Southern, 1968) was an attempt to develop a short form of the

WPPSI for use in predicting reading achievement. This study is related to that report by Wechsler in his manual in that the sample of the Plant and Southern study consisted of 56 of the 98 children used in the study reported by Wechsler. Plant and Southern did shed some light on the population which was used in their study and in the study reported by Wechsler. They reported that the school from which the sample was drawn served an economically blighted area and that the children came from families on welfare or families wherein the breadwinner was an unskilled laborer. This information naturally restricts the generalizability of the results of both studies.

Plant and Southern, having clarified one aspect of the population sampled, have still not provided information on the breakdown of the sample by sex and age group. This failure to clearly define the population drawn from and the sample used is quite disappointing, especially since these are two of the three studies reported to date on the validity of the WPPSI.

The Plant and Southern article dealt mainly with the ability of three measures of intelligence to predict reading achievement as measured by the Word Reading and Paragraph Meaning subtests of the Stanford Achievement Test, Primary I Battery (SAT). The three predictor measures were the (a) Peabody Picture Vocabulary Test (PPVT), (b) Stanford-Binet Intelligence Test, Form L-M (S-B), and (c) Wechsler Preschool and Primary Scale

of Intelligence (WPPSI). The predictor and criterion tests were administered approximately one and one-half years apart.

Correlational analyses were then performed between the predictor and criterion test scores. The PPVT and S-B IQ correlations with the SAT scores were .44 and .43 respectively; the WPPSI Verbal, Performance, and Full Scale IQ correlations with the same criterion measure were .43, .59, and .55. These results seem to indicate that the WPPSI is a somewhat better predictor of reading achievement than is either the PPVT or the S-B, at least for this sample.

An interesting aspect of these findings which may have some implications for the validity of the WPPSI is the fact that the Performance IQ was more highly correlated with reading achievement (.59) than was the Verbal IQ (.43) or the Full Scale IQ (.55). Logic seems to indicate that the verbal section of the WPPSI should correlate best with reading achievement. These results lead one to wonder what has caused this discrepancy. One possible explanation is that the culturally biased sample has produced biased results. Any more definite explanation must come from further research in this area with different populations.

Barclay and Yater (1969) reported the third of the three concurrent validity studies on the WPPSI. The Ss used in this study were described in more detail than those used in the other two studies. The Ss were 50

culturally disadvantaged children who had been participating in a Head Start program for an average of eight months. The children ranged in age from 60 through 71 months with 13 males, 12 females, 11 Caucasions, and 14 Negroes in each group. Barclay and Yater used a counterbalanced order of presentation of the WPPSI and the S-B. Twenty five children received the WPPSI first while the second 25 received the S-B first. The means and standard deviations for the WPPSI Verbal, Performance, and Full Scale IQs were 91.72 and 10.97, 96.32 and 12.09, and 93.38 and 11.57, respectively. The mean S-B IQ was 100.96 with a standard deviation of 12.46. The mean S-B IQ was significantly higher than the WPPSI Verbal, Performance, and Full Scale IQs at the .001 level for all three. No significant differences due to order of administration, sex, race, or age were found for any of the four IQ estimates. Correlation coefficients for the S-B IQ and the WPPSI Verbal, Performance, and Full Scale IQs were .73, .74, and .81, respectively.

Thus while Barclay and Yater described their sample in greater detail, it still is a restricted population in that it consists of culturally disadvantaged children. So far not one validity study of the WPPSI has been reported using a sample which has not been culturally deprived.

Of the four studies which were reported by Silverstein, three dealt with the development of short forms of the WPPSI (Silverstein, 1967,

1968a, 1968b) and the fourth with an upward extrapolation of the WPPSI and WISC IQ tables (Silverstein, 1968c). The pertinent aspect of these studies is that all of the above studies by Silverstein have used Wechsler's standardization data as reported in the WPPSI manual. Similarly, the study by Herman (1968) on sex differences on the WPPSI also used Wechsler's standardization sample as reported in the WPPSI manual.

The point of the foregoing discussion is that to date only two populations of Ss other than Wechsler's standardization sample have been studied in relation to the validity of the WPPSI. This is quite unfortunate, for already the WPPSI is being used as though it were a well established measure of intelligence. For example, Boyd and Means (1968) report a study in which the Block Designs, Geometrical Designs, Comprehension, and Sentences subtests of the WPPSI were used as part of a screening battery for the detection of mental retardation.

Perhaps the WPPSI is so readily accepted because of the extensive standardization procedures and the Wechsler name. Certainly the Wechsler name tends to carry with it some of the respect which is due to his other tests of intelligence, the WAIS and the WISC, and certainly the extensive standardization procedures are a necessary prerequisite for a usable test of intelligence. However, these factors alone are not sufficient justification for the acceptance being shown the WPPSI. As is pointed out in the American Psychological Association's Standards for Educational and

Psychological Tests and Manuals (1966) the test user must be aware of the validity and reliability of the instrument, and lacking such evidence the test designer and publisher should so inform the test user in the test manual. Wechsler has done so in his manual, but apparently the test users are not heeding his warnings.

Because of the above mentioned dearth of validity studies on the WPPSI, the present discussion will now turn to the validity studies of the WPPSI's forerunner, the WISC. This approach seems justified in two respects, namely by the similarity in standardization, construction, content, purpose, and theoretical foundation between the two and by the fact that the procedures for obtaining validity data are identical for both tests.

In reviewing studies involving the WISC and the S-B, three main themes will be focused upon. The first theme is that the WISC and the S-B are significantly related but that Ss tend to score higher on the S-B. The second is that the discrepancies in IQ scores between the WISC and the S-B are related to the S-B score in that the discrepancy increases as the S-B IQ increases. The third theme is that discrepancies between WISC and S-B IQs are related to chronological age in that the younger the S the larger the discrepancy. These same trends were examined by Littell (1960) in a review article.

The hypothesis that the WISC and S-B are related significantly with S-B yielding higher scores is supported by several studies which compared the WISC with the 1937 S-B. In one study Gehman and Matyas (1956) tested 60 school children with both the WISC and the S-B Form L and then retested the same children four years later. The Ss consisted of 29 boys and 31 girls with a mean age of 11-1 at the time of the first testing and a mean age of 15-2 at the time of the second testing. The correlations reported between the S-B and WISC Full Scale, Verbal, and Performance IQs for the first testing were .78, .46, and .73, respectively. At the time of the second testing the correlations were .76 with the Full Scale, .64 with the Verbal Scale, and .77 with the Performance Scale.

Mussen, Dean, and Rosenberg (1952) reported a study using 62 Ss, ages 6-0 to 13-1, in which they compared S-B IQs and WISC IQs. All of the children were attending grades 1 to 7 at the Ohio State University School. They reported the following correlations: WISC Full Scale and S-B, .85; WISC Verbal Scale and S-B, .84; and WISC Performance Scale and S-B, .72. The mean for the S-B was 120.3 while the means for the WISC Full, Verbal, and Performance Scales were 106.6, 109.0, and 107.7. The SD for the S-B was 27.0 while the SDs for the WISC Full, Verbal, and Performance Scales were 16.7, 18.0, and 14.7, in that order. The mean S-B IQ was significantly higher than all of the three mean WISC IQs.

These differences were particularly marked in this investigation because the population was a highly selected one, and the mean S-B IQ was considerably elevated by several scores over 154, the upper limit of the WISC norms.

Triggs and Cartee (1953) reported a study using 46 five-year-old children. The WISC and the S-B Form M were administered and correlations were computed between them. The correlations were found to be .58 between the S-B and the WISC Verbal Scale, .48 between the S-B and the WISC Performance Scale, and .62 between the S-B and the WISC Full Scale. Tests of significance were also computed between the mean score on the S-B (124.11) and the mean scores on the WISC Full Scale (107.56), Verbal Scale (103.39), and Performance Scale (111.07). The results of the tests of significance were all significant beyond the .01 level. A closer look at the data revealed that 93% of the Ss scored higher on the S-B than on the WISC Full Scale, 80.4% on the S-B than on the WISC Verbal Scale, and 91.3% on the S-B than on the WISC Performance Scale. Fifty-two per cent of the Ss scored higher on the S-B by 21 or more IQ points than on the WISC Full Scale and 78.3% of the Ss scored higher on the S-B by 11 points or more than on the WISC Full Scale. The above percentages are certainly impressive as are the results of the tests of significance. They indicate that at this age level the IQs of the S-B and WISC are not interchangeable.

Krugman, Justman, Wrightstone, and Krugman (1951) reported a well designed study in which 332 children drawn from 18 schools in the city of New York were tested both with the WISC and the 1937 S-B Form L. The data were then compared at each year level between ages 5 and 15 as well as for the entire group. Product-moment correlations were computed between S-B IQs and those obtained on the WISC Full Scale, Verbal Scale, and Performance Scale at each age level and for the total group. For the total group the correlations were: S-B and Full Scale, .82; S-B and Verbal Scale, .74; and for the S-B and Performance Scale, .64. The same correlations for the different age levels ranged from .73 to .92 for the S-B and WISC Full Scale, .64 to .88 for the S-B and WISC Verbal Scale, and from .48 to .79 for the S-B and WISC Performance Scale.

Krugman et al. then computed the mean IQs and the significance of mean differences of S-B and WISC Full Scale, Verbal, and Performance IQs at each age level and for the total group. The mean IQ for the total group was 108.45 on the Binet, 101.23 for the WISC Full Scale, 103.37 for the WISC Verbal Scale, and 98.28 for the WISC Performance Scale. The differences for the total group between the S-B and the WISC Full, Verbal, and Performance Scales were 7.22, 5.08, and 10.17, in that order. For the different age levels, the mean differences between the S-B and WISC Full Scale IQs ranged from 11.18 at age 5 to

their direct bearing on the second and third trends.

In considering the second trend, that discrepancies between the WISC and the S-B IQs are related to the S-B scores in that the discrepancies increase as the S-B IQs increase, two groups of studies will be examined. The first group consists of those studies which have used Ss of an average or above IQ level while the second group consists of those studies which have used Ss of a below average IQ level. This approach is necessary since few studies have actually used samples which have ranged from above average to below average intelligence levels.

In the first group of studies, a study by Weider, Noller, and Schramm (1951) used a group of 106 white children, ranging in age from 5-0 to 11-11 years. All of the Ss were administered both the WISC and Form L of the S-B. The correlation for all Ss between the WISC Full Scale and the S-B was .89; for the Verbal Scale and S-B, .89; and for the Performance Scale and S-B, .77. The means obtained for the S-B and the WISC Verbal, Performance, and Full Scales were 93.1, 91.1, 89.9, and 90.0, respectively. For the six Ss who scored highest on the S-B, the discrepancy between the S-B and WISC Full Scale IQs averaged 17 IQ points while for the six Ss who scored lowest on the S-B the discrepancy averaged 7 IQ points in favor of the S-B.

The study mentioned earlier by Krugman et al. (1951) which used

332 children and compared the WISC and Form L of the S-B also reported a definite tendency for greater differences between S-B and WISC IQs to be associated with higher S-B IQs. Correlations between the S-B IQs and the differences between S-B and WISC IQs were significant not only for the total group but also at almost every age level.

Harlow, Price, Tatham, and Davidson (1957) also compared 90 white Ss at three age levels and found that, in general, brighter children at all age levels tested higher on the S-B than on the WISC, with only a slight tendency for duller Ss to test lower on the WISC.

Triggs and Cartee (1953) who used 46 children found that 93% of the Ss scored higher on the S-B than the WISC Full Scale and that the largest discrepancies occurred at the above average and superior ranges of intelligence.

On the other hand, Holland (1953) who compared IQ scores for 52 children on the 1937 S-B and the WISC failed to find discrepancies between WISC and S-B scores that were related to the level of intelligence for his average intelligence level sample but did obtain support for the first trend. The 52 Ss ranged in age from 5-0 to 13-0 years, with 31 being boys and 21 girls. The correlations of the S-B with WISC IQs were .87, .88, and .73 for the Full, Verbal, and Performance Scales. The mean IQ for the S-B was 113.8 and for the WISC Full, Verbal, and

Performance Scales the mean IQs were 112.1, 111.2, and 111.3, respectively.

An article by Frandsen and Higginson (1951) which bears upon the hypothesis that level of intelligence is related to discrepancies between S-B and WISC IQs failed to support the hypothesis. Here 54 unselected, fourth-grade children, ranging in age from 9-1 to 10-3, were tested with both the 1937 S-B and the WISC. The intercorrelations between the S-B and the WISC Full, Verbal, and Performance Scales were, respectively, .80, .71, and .63. The means for the group were as follows: S-B, 105.8; WISC Full Scale, 102.4; WISC Verbal Scale, 100.9; and WISC Performance Scale, 103.5. The SDs for the group in the same order were 11.15, 11.15, 12.25, and 11.20. These results indicate that the WISC Full Scale and the S-B are measuring to a considerable extent the same factor or factors. The authors concluded that IQ norms from the S-B and WISC are comparable at the average level, and very probably fairly comparable within the range of one or two SDs above and below the means.

Of all the studies reported, the study by Frandsen and Higginson (1951) is the only one which found such comparable IQs between the WISC and the S-B. The authors reported no discrepancies due to intelligence, age, or any other factor. Perhaps this can be explained by the fact that, compared to the other studies, the age range of Ss in

the Frandsen and Higginson study is quite limited. Their sample was also more homogeneous than other samples studied in which SDs in the 14.0 to 16.0 range were generally obtained. Another point which makes this study difficult to compare with other such studies is that the authors failed to specify which form of the 1937 S-B was used. Most other studies have used Form L, however, Frandsen and Higginson may have used Form L, Form M, or both intermixed.

Both of the studies pertinent to the second hypothesis which used the 1960 S-B tended to support it. Estes et al. (1961) found that the difference in IQ scores between the 1937 S-B and the WISC was significant at the .00002 level of confidence, with the S-B having the higher mean IQ; and that the difference in IQ scores between the 1960 S-B and WISC was significant at the .002 level of confidence, with the 1960 S-B again having the higher mean IQ. This was true with two groups of subjects in the Superior range of intelligence while their were no significant differences in mean IQs between groups within the average range of intelligence. As Himelstein (1966) pointed out the direction of the discrepancy is to be anticipated in view of the larger standard deviation of the Binet scale. In their study, Barclay and Carolan (1966) also found that the 1960 revision of the Binet yields, as did the 1937 revision, higher scores among children of normal mental ability than does the WISC ($p < .05$).

Turning now from average and above average samples to studies using below average intelligence level Ss, one finds further support for the hypothesis that level of intelligence does influence discrepancies between the S-B and the WISC scores.

One of the studies reported in this group was by Sloan and Schneider (1951) who used 40 mental defectives ranging in age from 9-1 to 15-5. They found correlations between the 1937 S-B and WISC Full, Verbal, and Performance Scales of .76, .75, and .64, respectively. The mean IQs obtained were: on the S-B, 56.3; on the WISC Full Scale, 58.3; on the WISC Verbal Scale, 59.7; and on the WISC Performance Scale, 64.6. The SD on the Binet was 4.8 while those for the WISC in the same order as the means were 9.5, 6.2, and 12.7. The correlations reported in this study were comparable to those reported in the previous studies with average and above average intelligence level samples. The mean IQs reported, however, in addition to being generally lower, were quite different from those reported with the average level of intelligence samples in that the WISC means were higher than the S-B mean, significantly so for the Verbal and Performance Scale means ($p < .001$).

Stacey and Levin (1951) in a similar study using 44 morons reported similar results. In this study the correlation between the 1937 S-B

and the WISC Full Scale was .60 and between the S-B and the WISC Verbal Scale .56. No coefficient of correlation was reported between the S-B and WISC Performance Scale. Here again the authors reported the means and standard deviations for the different scales. The mean IQs were: for the S-B, 62.5; for the WISC Full Scale, 61.2; for the WISC Verbal Scale, 63.6; and for the WISC Performance Scale, 65.4. In the same order the SDs were 6.78, 5.48, 6.16, and 8.37. The correlations were similar to those reported with average level of intelligence samples, but the means were different. Again the WISC means were consistently higher than the S-B mean although these differences were not significant.

Nale (1951) reported a study using 104 mental defectives in which he compared only the 1937 S-B and the WISC Full Scale IQs. Nale found a correlation of .91 between the S-B and the WISC Full Scale IQs, with the mean S-B score being 55.38 and the mean WISC Full Scale score being 57.97. This difference was found to be significant at the .001 level of confidence.

Sandercock and Butler (1962) followed with a study using 90 mental defectives ranging in age from 10 to 16. They found correlations of ,76 between the 1937 S-B and WISC Full Scale IQs, .80 between S-B and WISC Verbal Scale IQs, and .66 between S-B and WISC Performance Scale

IQs. The mean IQs for this sample were 58.5 for the S-B, 59.0 for the WISC Full Scale, 62.8 for the WISC Verbal Scale, and 62.6 for the WISC Performance Scale. Here significant differences were found between the S-B and the WISC Verbal and Performance Scale mean IQs ($p < .001$).

A study by Vanderhorst, Sloan, and Bensberg (1953) using 38 defectives again produced results similar to those reported above. While no correlations were reported, the mean IQ scores were reported. They were 59.3 for the 1937 S-B, 62.18 for the WISC Full Scale, 61.74 for the WISC Verbal Scale, and 70.05 for the WISC Performance Scale. A significant difference at the .001 level was found only between the mean S-B score and the mean WISC Performance Scale score.

The above studies lend support for the hypothesis that level of intelligence influenced discrepancies between 1937 Stanford-Binet and WISC IQs in that when significant discrepancies were found between the tests with samples of average or above average intelligence, the S-B was found to be the higher of the two scores; with mental defectives, however, it has been seen that when significant discrepancies were found the WISC scores were higher than the 1937 S-B scores.

Rohrs and Haworth (1962), using a sample of 46 institutionalized mental defectives, with a previous average IQ of 61.12 as measured by

the 1937 Stanford-Binet, compared the 1960 Stanford-Binet with the WISC. They obtained the following correlations: .69 between the 1960 S-B and WISC Full Scale, .72 between 1960 S-B and WISC Verbal Scale, and .50 between the 1960 S-B and WISC Performance Scale. The mean IQs and standard deviations for the scales were: 56.91 and 6.38 for the 1960 S-B, 52.76 and 9.70 for the WISC Full Scale, 56.43 and 9.05 for the WISC Verbal Scale, and 57.54 and 11.07 for the WISC Performance Scale. The only significant difference found was between the S-B and the WISC Full Scale, with the S-B mean being 4.15 above the WISC mean. The above correlations seem to closely approximate those found by other investigators using a mental defective sample and the 1937 S-B. A surprising result of the above study is the significantly lower mean score on the WISC Full Scale than on the 1960 S-B. The previous studies with the 1937 Binet have generally indicated a trend in the opposite direction.

Rohrs and Haworth explained the discrepancy between their results and the results obtained by other investigators using the 1937 Binet in the following manner. The sample scored a mean 1960 S-B IQ significantly lower than the mean of their previous IQs. Since the previous IQs were obtained mainly on the 1937 S-B, the question can be raised as to whether the lower scoring is due to a difference between the 1960 and 1937 S-Bs. At first such a difference would seem to exist.

However, when one considers the fact that the present WISC mean IQ is also considerably below the previous mean IQ the difference seems to be one of intellectual deterioration rather than a difference in 1960 Binet vs. 1937 Binet test structure. The authors concluded that this deterioration might be due to the constricting effects of institutional living, or possibly due to an earlier deceleration of intellectual growth in these Ss than in the public in general. They also concluded that the higher 1960 S-B scores as compared to the WISC scores were probably due to peculiarities of their sample that are not known.

Still, the Rohrs and Haworth study did lend some support to the second hypothesis for the mean WISC Performance Scale IQ was actually higher than the mean S-B IQ though not significantly.

Thus it can be seen that the studies which compared the 1960 S-B and the WISC did, in general, support the hypothesis that level of intelligence does influence the size and direction of discrepancies between the 1960 S-B and WISC IQ scores. A progression of decreasing differences can be seen even in these studies in which the 1960 S-B and the WISC were compared: Estes et al. (1961) found a significant difference ($p < .002$) between the 1960 S-B and the WISC Full Scale mean IQs with two superior level groups but not with average intelligence level groups. Barclay and Carolan (1962) found that the mean 1960

S-B IQ was always higher than the mean WISC IQs with Ss of average intellectual ability. Rohrs and Haworth (1962) then found with mental defectives that the mean WISC Performance Scale IQ actually exceeded the mean S-B IQ though not significantly. In any event, sufficient grounds for further experimentation in this area are present, and intelligence level should be considered in any comparison between the Wechsler Scales and the S-B.

The third theme, that differences between WISC and S-B IQs increase with younger Ss has received some support from studies which have compared the 1937 S-B and the WISC.

In the study by Krugman et al. (1951) which used the 332 children, it was found that the mean differences between the S-B and WISC Full Scale IQs ranged from 11.18 at age 5 to only 3.75 at age 14-15; for the S-B and WISC Verbal Scale IQs from 9.34 at age 5 to 2.00 at age 14-15; and for the S-B and WISC Performance IQs from 12.15 at age 5 to 4.00 at age 14-15. As was mentioned earlier, in all cases the Binet scores were higher than the WISC scores and at only one age level (14-15 years) was the difference not significant. As Krugman et al. pointed out, the obtained significant differences in mean IQs progressed from fairly small differences at the older age levels to large differences at the younger age levels.

Triggs and Cartee (1953) in a study using 46 five-year-old children reported the mean score of the S-B as 124.11 and the mean scores of the WISC as 107.56 for the Full Scale, as 103.39 for the Verbal Scale, and as 111.07 for the Performance Scale. These results lend support to the findings of large discrepancies at early ages by Krugman et al. (1951).

Pastovic and Guthrie (1951) reported a study in which the WISC and the 1937 S-B Form L were compared at two different age levels. Fifty children were tested at 5-6 and 50 at 7-6 years. The order of administration was alternated to control for practice effects. At age 7-6 the correlations between the S-B and the WISC Full, Verbal, and Performance Scales were .88, .71, and .82. At age 5-6 the correlations were .71, .57, and .63. The mean IQ scores for the 7-6 group were 115.08 for the S-B, 108.56 for the WISC Verbal Scale, 112.68 for the WISC Performance Scale, and 111.50 for the WISC Full Scale. At age 5-6 the mean IQs, in the same order, were 113.20, 101.58, 104.24, and 103.16. All differences at both age levels were significant at the .01 level except for the WISC Performance and S-B comparison at the 7-6 age. This study is perhaps one of the most revealing so far for the present study. That is, in this study one group of Ss fell within the age range which was used in comparing the 1960 S-B and the WPPSI. The study by Pastovic and Guthrie points out the increasing discrepancy between the WISC

and S-B IQs at the lower age ranges, with the S-B being the higher scoring of the two tests.

Some studies using the 1937 S-B have failed to support the age hypothesis. Weider, Noller, and Schramm (1951) divided their 106 Ss into two groups, with Group A consisting of 44 Ss who were 5-0 to 7-11 years of age and Group B consisting of 62 Ss who were 8-0 to 11-11 years of age. No significant mean differences were found between the groups for any of the Scales though the Ss tended to score higher on the S-B.

Holland (1953), who used 52 Ss ranging in age from 5-0 through 13-0 years, also divided the Ss into two groups. Mean IQs were computed separately for Ss less than 10 years-old and for Ss 10 years and older. The mean S-B IQ for 23 Ss 10 years or older was 115.3. The mean WISC IQs were 113.2 for the Verbal Scale, 112.9 for the Performance Scale, and 113.6 for the Full Scale. The results of the tests of significance indicated that none of the mean WISC IQs were significantly different from the mean S-B IQ. The mean S-B IQ for the 29 Ss less than 10 years-old was 112.3. The mean WISC IQs were 109.1 for the Verbal Scale, 109.6 for the Performance Scale, and 110.6 for the Full Scale.

These results at first glance seem to indicate that age may not

be a factor influencing discrepancies between WISC and 1937 Binet IQs. Perhaps, though, this is a function of the age ranges used for the comparison. That is, the age ranges of the groups may have been large enough to hide true differences for the Weider, Noller and Schramm (1951) and the Holland (1953) studies. A significant trend was found by Krugman et al. (1951) who compared their Ss at each age level and by Pastovic and Guthrie (1951) who compared Ss differing more widely in age.

Estes et al. (1961) who used both the 1937 S-B and the 1960 S-B and 82 Ss ranging from grade one through eight failed to support the age hypothesis with either the 1937 S-B or the 1960 S-B and WISC discrepancies. Again, though, this may have been due to the large age spans used in forming the groups.

Barclay and Carolan (1966) compared two age levels of children using the 1960 S-B and the WISC. As Ss they used 104 children divided by age and race into four groups of 26. Two groups were composed of Negro children at the 7 and 12 year levels, and two groups were composed of white children at the same age levels. Differences between Binet and WISC IQ scores for 7 and 12-year-olds were significant at the .05 level for whites and at the .01 level for Negroes. Barclay and Carolan interpreted these findings as indicating that the 1960

revision of the Binet, as did the 1937 S-B, yielded higher scores at younger age levels among children of normal abilities than did the WISC. A disappointing aspect of the Barclay and Carolan study is their failure to report means and standard deviations for the sample. In fact the reader is left with no idea of how large the actual differences were in terms of IQ points. This limits the interpretation and application of these results.

Based on the above studies and the great similarity between the WISC and the WPPSI, certain relationships can be anticipated between the 1960 Stanford-Binet Form L-M and the Wechsler Preschool and Primary Scale of Intelligence. The purpose of the present study is to investigate these relationships. Specifically, the following hypotheses are proposed: (a) The S-B and the WPPSI IQ scores are significantly related; (b) Differences found between S-B IQs and WPPSI IQs are positively related to intelligence, i.e., the higher the 1960 S-B score, the larger the discrepancy between the scores; (c) Differences found between the S-B IQs and WPPSI IQs are negatively related to the chronological age of the subject, i.e., the younger the subject the greater is the difference between the two scores in favor of the S-B.

CHAPTER III

Method

Subjects

The subjects consisted of 50 boys and 50 girls between the ages of 4-5 and 6-5. The Ss were chosen so that two boys and two girls were included at each month level between 4-5 and 6-5 years. Each S was placed in a particular year and month category according to his age in days, months, and years on the date of the first testing. A child was placed in a specific category if on the date of the first testing he was within 15 days of that month and year category. For example, if a child was 5 years, 2 months, and 15 days old on the date of first testing, he was placed in the age group 5-2. On the other hand, if a child was 5 years, 2 months, and 16 days old, he was placed in the age group 5-3.

The subjects were obtained from the current kindergarten classes of three parochial schools located on the far north side of Chicago, Illinois, and from the lists of applicants for future entry to these same schools. All Ss were Caucasion and came from middle and upper-middle class families.

Measures

The measures used were the 1960 Stanford-Binet Form L-M and the

Wechsler Preschool and Primary Scale of Intelligence. The Matching Familiar Figures Test was administered in connection with another experiment and was not included in the data analysis of this experiment.

Procedure

In order to investigate the present hypotheses, the Ss were divided into four age groups: Group I consisted of the 28 Ss between 4-5 and 4-11, Group II consisted of the 24 Ss between 5-0 and 5-5, Group III consisted of the 24 Ss between 5-6 and 5-11, and Group IV consisted of the 24 Ss between 6-0 and 6-5 years of age. Each S was then individually administered the WPPSI and 1960 Stanford-Binet. The tests were all administered in a counterbalanced design to control for practice effects. That is, half of the Ss received the WPPSI first and half of the Ss the S-B first.

The tests were administered by four, second-year graduate students in clinical psychology who had completed both a practicum course in intelligence testing and a clinical clerkship in a setting requiring the administration and scoring of intelligence tests. Each of the examiners administered all three measures to approximately one quarter of the Ss. The same examiner also administered all three measures to control for experimenter effects.

Because of the time required to administer both the WPPSI and the

S-B, and because of the ages of the Ss, the tests were administered on two separate occasions. The average time between the first and second testing was approximately 10 days, with a range from 3 to 20 days. In administering the tests all procedures and instructions contained in the respective manuals were observed in the strictest possible manner. The actual testing took place at the Loyola University Child Guidance Clinic to insure uniform testing conditions for all Ss.

CHAPTER IV

Results

In order to test the first hypothesis, namely that the WPPSI and the 1960 S-B are significantly related, Pearson product-moment correlations were computed between the 1960 S-B IQ scores and the WPPSI Full Scale, Verbal, and Performance IQ scores for the four age groups, for the total sample, and for the total sample by sex. These correlation coefficients, all of which proved to be significantly different from zero at the .001 level, are presented in Table 1. In general, the correlations between the S-B and the WPPSI Full Scale and Verbal Scale are higher than those obtained between the S-B and the WPPSI Performance Scale. Also, the correlations for the males between the S-B and all three of the WPPSI Scales are higher than those for the females between the same scales.

As a second step in the consideration of the first hypothesis, the mean IQs and standard deviations obtained by the Ss at each age level and by the different sexes were computed for the S-B and the WPPSI Full Scale IQs. These data and the results of the tests of significance for the mean differences are presented in Table 2. All but one of the obtained differences between the mean S-B and the mean WPPSI IQs at the

TABLE 1

Correlation between Stanford-Binet and WPPSI Verbal, Performance, and Full Scales

Group	N	S-B vs WPPSI Verbal	S-B vs WPPSI Performance	S-B vs WPPSI Full Scale
4-5 to 4-11	28	.77	.55	.77
5-0 to 5-5	24	.93	.74	.89
5-6 to 5-11	24	.86	.65	.85
6-0 to 6-5	24	.80	.52	.86
Males	50	.87	.69	.87
Females	50	.82	.56	.79
Total	100	.84	.61	.82

Note.—All correlations were significant beyond the .001 level.

TABLE 2

Mean IQs, Standard Deviations, and Tests of Significance of Mean Differences of S-B and WPPSI Full Scale IQs at Each Age Level, by Sex, and for Total Group

Group	N	Mean	SD	t
4-5 to 4-11				
Binet	28	113.50	17.67	4.18 ***
WPPSI	28	104.61	15.01	
5-0 to 5-5				
Binet	24	114.67	18.73	3.95 ***
WPPSI	24	107.50	18.90	
5-6 to 5-11				
Binet	24	114.21	13.61	5.62 ***
WPPSI	24	105.79	12.67	
6-0 to 6-5				
Binet	24	107.29	11.50	1.53
WPPSI	24	105.46	9.40	
Male				
Binet	50	111.88	15.09	4.87 ***
WPPSI	50	106.74	14.44	
Female				
Binet	50	113.04	16.59	5.69 ***
WPPSI	50	104.84	14.16	
Total				
Binet	100	112.46	15.79	7.40 ***
WPPSI	100	105.79	14.26	

*** $p < .001$.

different age levels were significant at the .001 level. The results of the test of significance for the 6-0 to 6-5 age group were not significant while the results were significant at the .001 level for males, females, and the total sample.

The results of the tests of significance for the mean differences between the S-B and the WPPSI Verbal IQs are presented in Table 3 along with the respective means and standard deviations at each age level and by sex. Again all but one of the obtained differences between the mean S-B and WPPSI IQs at the different age levels were significant at the .001 level. The 6-0 to 6-5 age group's results were significant at the .01 level. These differences were also significant at the .001 level for males, females, and the total sample.

Table 4 presents the results of the tests of significance for the mean differences between the S-B and the WPPSI Performance IQs as well as the means and standard deviations at each age level and by sex. The results for age levels 4-5 to 4-11 and 5-6 to 5-11 were significant at the .01 level and those for the 5-0 to 5-5 age level were significant at the .05 level, while those for the 6-0 to 6-5 group were not significant. The results for the males were significant at the .05 level and those for the females and the total sample were significant at the .001 level.

That Ss tended to score higher on the S-B than on the WPPSI

TABLE 3

Mean IQs, Standard Deviations, and Tests of Significance of Mean Differences of S-B and WPPSI Verbal IQs at Each Age Level, by Sex, and for Total Group

Group	N	Mean	SD	t
4-5 to 4-11				
Binet	28	113.50	17.67	4.49 ***
WPPSI	28	103.64	16.25	
5-0 to 5-5				
Binet	24	114.67	18.73	6.35 ***
WPPSI	24	105.63	17.77	
5-6 to 5-11				
Binet	24	114.21	13.61	7.09 ***
WPPSI	24	104.25	12.40	
6-0 to 6-5				
Binet	24	107.29	11.50	3.56 **
WPPSI	24	102.21	10.48	
Male				
Binet	50	111.88	15.09	6.96 ***
WPPSI	50	104.44	14.62	
Female				
Binet	50	113.04	16.59	7.16 ***
WPPSI	50	103.40	14.36	
Total				
Binet	100	112.46	15.79	9.90 ***
WPPSI	100	103.92	14.43	

** $p < .01$.
*** $p < .001$.

TABLE 4

Mean IQs, Standard Deviations, and Tests of Significance of Mean Differences of S-B and WPPSI Performance IQs at Each Age Level, by Sex, and for Total Group

Group	N	Mean	SD	t
4-5 to 4-11				
Binet	28	113.50	17.67	3.02 **
WPPSI	28	104.64	14.58	
5-0 to 5-5				
Binet	24	114.67	18.73	2.38 *
WPPSI	24	108.13	18.82	
5-6 to 5-11				
Binet	24	114.21	13.61	3.36 **
WPPSI	24	106.50	13.27	
6-0 to 6-5				
Binet	24	107.29	11.50	0.38
WPPSI	24	108.17	11.81	
Male				
Binet	50	111.88	15.09	2.33 *
WPPSI	50	108.02	14.46	
Female				
Binet	50	113.04	16.59	3.56 ***
WPPSI	50	105.52	14.96	
Total				
Binet	100	112.46	15.79	4.22 ***
WPPSI	100	106.77	14.69	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

scales is shown by the results presented in Table 5. In addition, it can be seen that over 50% of the Ss scored more than five IQ points higher on the S-B than on the WPPSI Verbal and Full Scales and that over 40% of the Ss scored more than five IQ points higher on the S-B than the WPPSI Performance Scale. At the same time only 15% of the Ss scored higher on the WPPSI Verbal Scale by more than five IQ points, only 21% of the Ss scored higher on the WPPSI Performance Scale by more than five IQ points, and only 4% of the Ss scored higher on the WPPSI Full Scale by more than five IQ points. For all of the comparisons, the largest percentage of No Difference was 6% for the WPPSI Full Scale vs the S-B comparison.

In order to investigate the second hypothesis that differences found between S-B IQs and WPPSI IQs are positively related to the S-B in that the higher the S-B IQ the larger the discrepancy between the scores, correlations between S-B IQs and the differences between S-B IQs and WPPSI Full Scale, Verbal, and Performance Scale IQs were computed and are reported in Table 6. While all correlations were significant for the 4-5 to 4-11 age level and for all Ss, the findings were less consistent for the other age levels.

Table 7 shows the percentages of children at given S-B IQ levels showing given differences between S-B and WPPSI Full Scale IQs. While

TABLE 5

Percentage of Children Showing Given IQ Differences between S-B and WPPSI Verbal, Performance and Full Scales

IQ Point Difference	WPPSI Verbal	WPPSI Performance	WPPSI Full Scale
Higher on S-B			
1-5	23	15	19
6-10	20	10	22
11-15	17	9	18
16-20	11	13	10
21 or more	7	11	4
Total	88	58	73
Higher on WPPSI			
1-5	3	16	17
6-10	2	13	2
11-15	2	6	1
16-20	1	3	1
21 or more	-	-	-
Total	8	38	21
No Difference	4	4	6

TABLE 6

Correlations between S-B IQs and Differences between S-B IQs and WPPSI Verbal, Performance, and Full Scale IQs

Group	N	S-B and Diff between S-B and WPPSI Verbal	S-B and Diff between S-B and WPPSI Performance	S-B and Diff between S-B and WPPSI Full Scale
4-5 to 4-11	28	.45*	.63**	.57**
5-0 to 5-5	24	.36	.46*	.26
5-6 to 5-11	24	.42	.32	.40*
6-0 to 6-5	24	.14	.10	.22
Total	100	.40**	.44**	.42**

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

TABLE 7

Percentage of Children at Given S-B IQ Level Showing Given Differences between S-B and WPPSI Full Scale IQs

	Higher on S-B					Higher on WPPSI Full Scale					
S-B IQ	1-5	6-10	11-15	16-20	21 or more	1-5	6-10	11-15	16-20	21 or more	No Diff
130 or more	1	3	2	4	3						
120-129	3	4	5	2		2					2
110-119	3	9	4	2		4	2				
100-109	6	4	4		1	5			1		4
90-99	5	1	3	1		4		1			
80-89	1					2					
80 or below		1		1							

again there is a tendency for the size of the discrepancy to increase as the S-B IQ score increases, it is limited by the few S-B IQ scores below 90 found in the sample. It is also apparent from examining Table 7 how few of the Ss scored higher on the WPPSI than on the S-B at all IQ levels.

To test the third hypothesis that differences found between S-B IQ scores and WPPSI Verbal IQ scores are related to the chronological age of the S in that the younger the S the greater the discrepancy between the two scores in favor of the S-B, the cross-tabulation of chronological age and difference between S-B and WPPSI Verbal IQs was computed and is presented in Table 8. The obtained Chi-square was significant at the .05 level for a one-tailed test.

Table 9 presents evidence on the cross-tabulation of chronological age and differences between S-B and WPPSI Performance Scale IQs, while Table 10 presents evidence on the cross-tabulation of chronological age and differences between S-B and WPPSI Full Scale IQs. In neither case was the obtained Chi-square significant.

TABLE 8

Cross-Tabulation of Chronological Age and

Difference between S-B and WPPSI

Verbal Scale IQs

Age Level	-1 and over	0 to 4	5 to 9	10 and over	Total
4-5 to 4-11	2	8	3	15	28
5-0 to 5-5	1	7	3	13	24
5-6 to 5-11	1	4	7	12	24
6-0 to 6-5	5	6	8	5	24
Total	9	25	21	45	100

$X^2 = 14.90$

$p < .05$.

TABLE 9

Cross-Tabulation of Chronological Age and

Difference between S-B and WPPSI

Performance Scale IQs

Age Level	-1 and over	0 to 4	5 to 9	10 and over	Total
4-5 to 4-11	8	4	3	13	28
5-0 to 5-5	9	4	4	7	24
5-6 to 5-11	7	3	4	10	24
6-0 to 6-5	13	4	1	6	24
Total	37	15	12	36	100

$X^2 = 7.36$

TABLE 10

Cross-Tabulation of Chronological Age and

Difference between S-B and WPPSI

Full Scale IQs

Age Level	-1 and over	0 to 4	5 to 9	10 and over	Total
4-5 to 4-11	5	3	5	15	28
5-0 to 5-5	5	5	5	9	24
5-6 to 5-11	3	5	5	11	24
6-0 to 6-5	8	6	8	2	24
Total	21	19	23	37	100

$X^2 = 14.035$

CHAPTER V

Discussion

The first hypothesis, that the S-B and the WPPSI IQ scores are significantly related, was strongly supported by the correlations presented in Table 1. In general, these correlations were slightly higher than those found by other investigators who have compared the S-B and the WPPSI. While the present study obtained total sample correlations of .84, .61, and .82 between the S-B and the WPPSI Verbal, Performance, and Full Scale IQs respectively, Wechsler (1967) reported correlations of .76, .56, and .75 and Barclay and Yater (1969) reported correlations of .73, .74, and .81. The correlations found in this study are also comparable to those found in studies comparing the S-B and the WISC. That the correlations found in the present study are slightly higher may be due to the fact that the present sample consisted of white, middle-class children who obtained a higher mean IQ than the culturally deprived samples of Wechsler (1967) and Barclay and Yater (1969).

While S-B and WPPSI IQ scores are related significantly, the results of the tests of significance between the S-B and WPPSI Verbal, Performance, and Full Scale means indicated that the S-B

and WPPSI IQ scores are not directly interchangeable. Only at the 6-0 to 6-5 age level for the Full Scale and Performance Scale comparisons with the S-B were the results not significant. Examination of the means revealed that the Ss consistently scored higher on the S-B than on the WPPSI. These findings are in accord with those of Wechsler (1967) and Barclay and Yater (1969) as well as with the studies of the WISC and S-B.

The results presented in Table 5 indicate that the magnitude of the differences between the S-B and the WPPSI scores is one which might interfere with accurate clinical judgment. The fact that one third of the children scored higher on the S-B by 11 or more IQ points does not allow the clinician to reliably place these children in one of the intelligence classifications proposed by Wechsler. These results again are consistent with those obtained by investigators who have compared the S-B with the WISC (e.g., Krugman et al., 1951).

The second hypothesis, that differences between S-B and WPPSI IQ scores are related positively to the S-B scores in that the higher the S-B score the larger the difference, was partially supported by the significant correlations for the total sample presented in Table 6. While this relationship did hold true for the total sample and the 4-5 to 4-11 and the 5-6 to 5-11 age levels, it was not found to be

significant for the 5-0 to 5-5 and 6-0 to 6-5 age levels. This failure to obtaine full support for this hypothesis at all age levels may have been due to the limited number of S-B IQ scores below 90. This restriction of IQ range is evident upon examining Table 7. Here again while there was a tendency for the size of the discrepancy between S-B and WPPSI IQs to increase as the S-B IQ increased, the analysis was limited by the low number of S-B IQs below 90. In fact, the children with S-B IQs below 90 all fell into the 4-5 to 4-11 and 5-6 to 5-11 age levels, those same age levels for which significant results were obtained.

The size of the correlations reported in Table 6 for the total sample (S-B differences with WPPSI Verbal, .39; with WPPSI Performance, .43; and with WPPSI Full Scale, .42) indicated that this relationship was accounting for approximately 16% of the variance between the scores. In similar comparisons between the WISC and the S-B it was found that this relationship accounted for approximately 25% to 36% of the variance between the scores.

The third hypothesis, that differences found between S-B IQ scores and WPPSI IQ scores are negatively related to the chronological age of the subject in that the younger the subject the greater the difference between the two scores in favor of the S-B, was only partially supported

by the results presented in Tables 8, 9, and 10. The proposed relationship held for the differences between the S-B and Verbal Scale IQs but not for the differences between the S-B and the Performance and Full Scales. The failure of the relationship to reach significance for the WPPSI Full Scale can be better understood by examining the actual Chi-square value obtained. The Chi-square value obtained in the cross-tabulation which involved the Full Scale was 14.04 while the value necessary for significance was 14.68. This, coupled with the fact that the Full Scale is a composite score based on the Verbal and Performance Scales may have kept the Full Scale from reaching significance. As for the Performance Scale and S-B comparison with chronological age not reaching significance, it must be remembered that there is a heavy loading of performance items on the S-B until about age five. From age five on there is a much heavier loading of verbal items on the S-B. This qualitative change in the Binet scale may have served to mask the proposed relationship between age and size of the difference between S-B and Performance Scale IQs.

An alternative explanation of the failure of the WPPSI Full Scale and Performance Scale comparisons to reach significance is that the relationship does not hold between the S-B and the WPPSI as it does between the S-B and the WISC. An even more attractive explanation is that the age spread covered by the WPPSI is too small to reveal such a

relationship if it does exist. This explanation seems likely when one considers that the age spread covered by the WISC is approximately nine years while that covered by the WPPSI is only two and one-half years, and when one considers that in most studies which have found the relationship between chronological age and IQ differences between the S-B and the WISC the Ss compared have differed by five or more years in age. In the present sample the ages ranged from 4-5 to 6-5 years, a spread of two years.

Based on the results of this study and the foregoing discussion, the following conclusions seem warranted:

(a) That S-B and WPPSI Full, Verbal, and Performance IQ scores are significantly related but not interchangeable, and that caution must be used in applying WPPSI IQs clinically.

(b) That there is a positive relationship between S-B IQs and the difference between S-B and WPPSI IQs with the size of the difference increasing as the S-B IQ score increases.

(c) That there is no conclusive evidence supporting the hypothesis that differences between S-B and WPPSI IQ scores are related to chronological age in that the younger the S the larger the difference.

Here it must be recognized that these conclusions are limited in

their generalizability by the fact that all of the Ss were Caucasian, middle-class children drawn from the north side of Chicago and by the fact that most of the Ss were attending kindergarten or a nursery school.

Another factor limiting the generalizability of these results is the fact that all of the examiners were young, male graduate students. Perhaps different results might have been obtained if the tests had been administered by examiners who were older, female, or more experienced. In any event, many more studies using a wide variety of Ss and examiners are necessary before the true generalizability of these results will be known.

CHAPTER VI

Summary

The purpose of this study was to investigate the relationship of the WPPSI to the 1960 S-B. The three specific hypotheses investigated were that the WPPSI and the S-B are significantly related, that differences between WPPSI and S-B IQ scores are positively related to S-B scores in that the higher the S-B score the larger the difference, and that differences between WPPSI and S-B IQ scores are negatively related to chronological age in that the younger the age the larger the difference. The Ss consisted of 50 boys and 50 girls between the ages of 4-5 and 6-5 years with two boys and two girls at each month level. Each S was individually administered the WPPSI and S-B in a counterbalanced design by one of four male examiners.

Pearson product-moment correlations were computed between the WPPSI and S-B scores and between the S-B scores and the differences between S-B and WPPSI scores. In addition, tests of significance between related means were performed between the mean S-B and WPPSI scores. Finally, Chi-squares were used to analyze the cross-tabulations of chronological age and differences between S-B and WPPSI scores.

The results of the statistical analysis indicated that the WPPSI and S-B were significantly related but not interchangeable so that caution must be used when employing the WPPSI scores for clinical judgment. The results also indicated that differences between WPPSI and S-B scores were related to the S-B scores in that the higher the S-B score the larger the difference. However, the results failed to support the hypothesis that chronological age is related to differences between S-B and WPPSI scores in that the younger the age the larger the difference. It was cautioned that these results are limited in their generalizability by the small geographic region from which the sample was selected and possibly by the age, sex, and experience of the examiners.

REFERENCES

American Psychological Association. Standards for educational and psychological tests and manuals. Washington, D.C.: American Psychological Association, Inc., 1966.

Barclay, A., & Carolan, P. 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.

Barclay, A., & Yater, A. Comparative study of the Wechsler Preschool and Primary Scale of Intelligence and the Stanford-Binet Intelligence Scale, Form L-M, among culturally deprived children. Journal of Consulting and Clinical Psychology, 1969, 33, 257.

Boyd, V., & Means, J. A psychological screening test form for children suspected of mental retardation. Psychological Reports, 1968, 23, 213-214.

Estes, B., Curtin, M., DeBurger, R., & Deny, C. Relationships between the 1960 Stanford-Binet, 1937 Stanford-Binet, WISC, Raven, and Draw-A-Man. Journal of Consulting Psychology, 1961, 25, 388-391.

Frandsen, A., & Higginson, J. The Stanford-Binet and the Wechsler Intelligence Scale for Children. Journal of Consulting Psychology, 1951, 15, 236-238.

Gehman, I., & Matyas, R. Stability of the WISC and Binet tests. Journal of Consulting Psychology, 1956, 20, 150-152.

Harlow, J., Price, A., Tatham, L., & Davidson, J. Preliminary study of comparison between Wechsler Intelligence Scale for Children and Form L of revised Stanford-Binet Scale at three age levels. Journal of Clinical Psychology, 1957, 13, 72-73.

Herman, D. A study of sex differences on the Wechsler Preschool and Primary Scale of Intelligence. Proceedings, 76th Annual Convention, American Psychological Association, Inc., 1968.

Himelstein, P. Research with the Stanford-Binet, Form L-M: The first five years. Psychological Bulletin, 1966, 65, 156-164.

Holland, G. A comparison of the WISC and Stanford-Binet IQs of normal children. Journal of Consulting Psychology, 1953, 17, 147-152.

Krugman, K., Justman, J., Wrightstone, J., & Krugman, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. Journal of Consulting Psychology, 1951, 15, 475-483.

Littell, W. The Wechsler Intelligence Scale for Children - Review of a decade of research. Psychological Bulletin, 1960, 57, 132-156.

Mussen, P., Dean, S., & Rosenberg, M. Some further evidence on the validity of the WISC. Journal of Consulting Psychology, 1952, 16, 410-411.

Nale, S. The children's Wechsler and the Binet on 104 mental defectives at the Polk State School. American Journal of Mental Deficiency, 1951, 56, 419-423.

Pastovic, J., & Guthrie, G. Some evidence on the validity of the WISC. Journal of Consulting Psychology, 1951, 15, 385-386.

Plant, W., & Southern, M. First grade reading achievement predicted from WPPSI and other scores obtained 18 months earlier. Proceedings, 76th Annual Convention, American Psychological Association, Inc., 1968.

Rohrs, F., & Haworth, M. The 1960 Stanford-Binet, WISC, and Goodenough Tests with mentally retarded children. American Journal of Mental Deficiency, 1962, 66, 853-859.

Sandercock, M., & Butler, A. An analysis of the performance of mental defectives on the Wechsler Intelligence Scale for Children. American Journal of Mental Deficiency, 1952, 57, 100-105.

Silverstein, A. A short short form of Wechsler's scales for screening purposes. Psychological Reports, 1967, 21, 842.

Silverstein, A. Validity of WPPSI short forms. Journal of Consulting and Clinical Psychology, 1968, 32, 229-230. (a)

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

Silverstein, A. WISC and WPPSI IQs for the gifted. Psychological Reports, 1968, 22, 1168. (c)

Sloan, V., & Schneider, B. A study of the Wechsler Intelligence Scale for Children with mental defectives. American Journal of Mental Deficiency, 1951, 55, 573-575.

Stacey, C., & Levin, J. Correlational analysis of scores of subnormal subjects on the Stanford-Binet and the Wechsler Intelligence Scale for Children. American Journal of Mental Deficiency, 1951, 55, 590-597.

Triggs, F., & Cartee, J. Pre-school pupil performance on the Stanford-Binet and the Wechsler Intelligence Scale for Children. Journal of Clinical Psychology, 1953, 9, 27-29.

Vanderhorst, L., Sloan, W., & Bensberg, G. Performance of mental defectives on the Wechsler-Bellevue and the WISC. American Journal of Mental Deficiency, 1953, 57, 481-483.

Wechsler, D. Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: The Psychological Corp., 1967.

Weider, A., Noller, P., & Schramm, T. The Wechsler Intelligence Scale for Children and the revised Stanford-Binet. Journal of Consulting Psychology, 1951, 15, 330-333.

APPROVAL SHEET

The thesis submitted by James W. Futterer has been read and approved by the director of the thesis. Furthermore, the final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

20 May 1970
Date

Patricia M Barger Ph.D.
Signature of Advisor