Mean-Score Differences between the WISC-R and the Stanford-Binet Intelligence Scale: Fourth Edition
Peter N. Prewett and Mark A. Matavich
Assessment for Effective Intervention 1992 17: 195
DOI: 10.1177/153450849201700304

The online version of this article can be found at:
http://aei.sagepub.com/content/17/3/195

Published by:
Hammill Institute on Disabilities

and

SAGE

http://www.sagepublications.com

Additional services and information for Assessment for Effective Intervention can be found at:

Email Alerts: http://aei.sagepub.com/cgi/alerts

Subscriptions: http://aei.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav
Mean-Score Differences between the WISC-R and the Stanford-Binet Intelligence Scale: Fourth Edition

PETER N. PREWETT
MARK A. MATAVICH

Abstract

The mean score differences between the Stanford-Binet Intelligence Scale: Fourth Edition and the WISC-R changed across the ability continuum for 126 referred students. The Stanford-Binet Test Composite was significantly higher than the WISC-R Full Scale score at the lower end of the ability continuum, but was significantly lower than the WISC-R Full Scale score at the higher end. The two tests yielded similar scores for students with WISC-R Full Scale scores ranging from 70 to 89. The WISC-R scores correlated significantly with the Stanford-Binet scores.

Diagnostic criteria of educationally handicapping conditions, such as mental retardation and specific learning disabilities, include scores obtained on norm-referenced tests of intellectual and academic functioning. Because tests designed to measure similar constructs do not always yield similar scores when administered to the same individual (Bracken, 1988), however, different diagnostic conclusions could be reached depending on the test given.

Because diagnostic decisions could be related to the test administered, it is important to know if one test usually gives higher or lower scores than another test that purports to measure the same construct. For intellectual assessments, it is important to know how other intelligence tests compare to the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) because the WISC-R has been the most widely used intelligence scale (Goh, Teslow, & Fuller, 1981; Reschly, Genshaft, & Binder, 1986).

Although the WISC-R has been replaced by its revision, the WISC-III (Wechsler, 1991), knowledge of how the WISC-R relates to other intelligence tests remains important. This is because most previously identified handicapped students were assessed with the WISC-R. Therefore, when comparing the test results of a triennial reevaluation with previous assessments, knowledge of whether a change in test scores is
related to test characteristics (e.g., typical and expected mean score differences) or a relative change in the intellectual functioning of the student is needed.

This study examined the relationship between the WISC-R and the Stanford-Binet Intelligence Scale: Fourth Edition (Stanford-Binet; Thorndike, Hagen, & Sattler, 1986). This study is needed because the available research in this area has generally used small samples of students with similar abilities. Lukens (1990) found the Stanford-Binet Test Composite was significantly higher than the WISC-R Full Scale score when both were administered to mentally retarded students. In contrast, Rothlisberg (1987) found the Stanford-Binet gave significantly lower (by seven points) scores than the WISC-R with normal second grade students \( n = 32 \). In view of the contrasting results of Lukens (1990) and Rothlisberg (1987), research with samples of students with abilities spanning the ability continuum is needed to examine the reason for the discrepant findings. For the present study, it was predicted that the two tests would be highly correlated and that the Stanford-Binet IQ will be significantly lower than the WISC-R IQs.

**Method**

**Subjects**

The subjects were 126 students (91 males, 35 females; 76 Whites, 50 Blacks) attending a large urban school district in central Ohio. The average age of the sample was 9 years, 7 months (range = 6 years, 5 months to 15 years, 10 months). These students had been referred to a school psychologist because of their academic difficulties in the classroom.

**Procedures**

One of two male certified school psychologists administered the Stanford-Binet and WISC-R in counterbalanced order in one session to the subjects. The students were administered the tests as part of a multifactored evaluation to determine special education eligibility.

A series of \( t \) tests compared the five Stanford-Binet scores with the three WISC-R scores. Because the Stanford-Binet and the WISC-R scores have different standard deviations (16 & 15, respectively), the Stanford-Binet scores were transformed to a standard deviation of 15 for the \( t \)-test comparisons. The Bonferroni inequality (Stevens, 1986, p. 7) kept the experimentwise error rate at 1% \( (p < .0007 \) for each of the 15 \( t \)-tests). To determined if the Stanford-Binet Area and Test Composite scores correlated similarly with the WISC-R Verbal and Performance
scores, five $t$ tests for comparing differences between correlations from a dependent sample were calculated (Hinkle, Wiersma, & Jurs, 1979, p. 223).

**Results**

For the total sample, the mean Stanford-Binet and WISC-R scores tended to be similar (Table 1). The Stanford-Binet Test Composite and the WISC-R Full Scale scores differed by a nonsignificant .8 points [$t(125) = 1.00$]. The mean Stanford-Binet Verbal Reasoning score was significantly higher than the mean WISC-R Verbal and Full Scale scores [$t(125) = 5.85$ & $5.10$ respectively]. The Stanford-Binet Test Composite was significantly lower than the mean WISC-R Performance Score [$t(125) = 3.95$]. The discrepancies between the other Stanford-Binet and WISC-R scores were not significant.

Although the Stanford-Binet and WISC-R scores were similar for the total sample, the restricted range of the Stanford-Binet scores relative to the WISC-R scores (Table 1) indicated that the mean score differences between the two tests changed across the ability continuum. To examine this more closely, the subjects were divided into four groups based on the range their WISC-R Full Scale scores fell in: (a) 51–69, (b) 70–79, (c) 80–89, and (d) 90–99 (Table 2). Equal sample sizes for the groups (22) were obtained by randomly deleting 11 subjects from the 70–79 range and 16 subjects from the 80–99 range. Because of their small number (11),

**Table 1: Means and Standard Deviations for Stanford-Binet Area and Test Composite Scores and WISC-R Scores**

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>Transformed Scores$^a$</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford-Binet$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR</td>
<td>85.05</td>
<td>10.90</td>
<td></td>
<td>85.97</td>
<td>10.18</td>
</tr>
<tr>
<td>A/VR</td>
<td>81.67</td>
<td>11.41</td>
<td></td>
<td>82.83</td>
<td>10.68</td>
</tr>
<tr>
<td>QR</td>
<td>83.06</td>
<td>10.68</td>
<td></td>
<td>84.17</td>
<td>9.96</td>
</tr>
<tr>
<td>STM</td>
<td>80.02</td>
<td>10.19</td>
<td></td>
<td>81.22</td>
<td>9.53</td>
</tr>
<tr>
<td>TC</td>
<td>79.60</td>
<td>9.97</td>
<td></td>
<td>80.87</td>
<td>9.35</td>
</tr>
<tr>
<td>WISC-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>81.41</td>
<td>13.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>84.83</td>
<td>13.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Scale</td>
<td>81.67</td>
<td>13.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Stanford-Binet Area and Test Composite scores transformed from a standard deviation of 16 to a standard deviation of 15.

$^b$ VR = Verbal Reasoning; A/VR = Abstract/Visual Reasoning; QR = Quantitative Reasoning; STM = Short Term Memory; and, TC = Test Composite.
Table 2: Mean Score Differences between the WISC-R and Stanford-Binet Test Composite Scores at Various Ranges of WISC-R Full-Scale Scores

<table>
<thead>
<tr>
<th>Full Scale Range</th>
<th>WISC-R Mean</th>
<th>Stanford-Binet Mean</th>
<th>Diff.</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>51–69</td>
<td>62.77</td>
<td>71.54</td>
<td>−8.76</td>
<td>9.69*</td>
</tr>
<tr>
<td>70–79</td>
<td>74.59</td>
<td>77.42</td>
<td>−2.83</td>
<td>1.91</td>
</tr>
<tr>
<td>80–89</td>
<td>83.96</td>
<td>80.70</td>
<td>3.26</td>
<td>2.80</td>
</tr>
<tr>
<td>90–99</td>
<td>94.27</td>
<td>87.86</td>
<td>6.42</td>
<td>4.86*</td>
</tr>
</tbody>
</table>

Note. n = 22 for each range.

a Means were calculated using Stanford-Binet Transformed scores (SD = 15).

*p < .001.

subjects with WISC-R Full Scale scores above 100 were not used to form a group.

A two-factor repeated-measures ANOVA (four groups by two tests) was significant for main effect \[ F(3,84) = 129.19, \ p < .0001 \] and the interaction between group and repeated measures \( F = 29.67, \ p < .0001 \). Four \( t \) tests (reported in Table 2) compared the mean score differences between the Stanford-Binet and WISC-R scores for these groups. The mean Stanford-Binet Test Composite was significantly higher than the mean WISC-R Full Scale score for the 51–69 group but was significantly lower than the WISC-R Full Scale score for the 90–99 group (Table 2). The Stanford-Binet Test Composites and WISC-R Full Scale scores were similar for the 70–79 and 80–89 groups (Table 2). A one-factor ANOVA comparing the mean score differences (see Table 2) between the Stanford-Binet and the WISC-R across groups was significant \[ F(3,84) = 33.53, \ p < .0001 \]. Post hoc Scheffe \( F \)-tests found all between group comparisons were significant with the exception of the 80–89 and 90–99 groups. It appears, then, that the finding of a nonsignificant difference between the two tests for the total sample is misleading because the direction and magnitude of the discrepancy between them varies across the ability continuum.

The Stanford-Binet Area and Test Composite scores correlated significantly with the WISC-R Verbal, Performance, and Full Scale scores \( p < .01 \) (Table 3). The Stanford-Binet Verbal Reasoning Area correlated significantly higher with the WISC-R Verbal than Performance score \( t(123) = 6.13, \ p < .01 \). The Stanford-Binet Abstract/Visual Reasoning, Quantitative Reasoning, Short Term Memory, and Test Composite scores correlated similarly with the WISC-R Verbal and Performance scores \( t(123) = 2.61, \ .19, \ 1.09, \ \text{and} \ 1.66, \text{respectively} \).
Table 3: Pearson Correlations between Stanford-Binet Area and Test Composite Scores and WISC-R Scores (N=126)

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>WISC-R Performance</th>
<th>Full Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford-Binet³</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR</td>
<td>75</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td>A/VR</td>
<td>45</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>QR</td>
<td>48</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>STM</td>
<td>43</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>TC</td>
<td>69</td>
<td>62</td>
<td>73</td>
</tr>
</tbody>
</table>

Note. Decimals omitted and all correlations significant (p < .01).
³ VR = Verbal Reasoning; A/VR = Abstract/Visual reasoning; QR = Quantitative Reasoning; STM = Short Term memory; and, TC = Test Composite.

Discussion

The relationship between the WISC-R and Stanford-Binet varied across the ability continuum. The Stanford-Binet yielded significantly higher scores than the WISC-R at the lower end of the ability continuum but yielded significantly lower scores than the WISC-R at the higher end. The two tests provided similar scores, however, when the WISC-R Full Scale scores ranged from 70–89.

The current findings differ with studies reported in the Stanford-Binet manual (Thorndike et al., 1986). These studies show the Stanford-Binet Test Composite and WISC-R Full Scale scores differed by 2.0 points with learning disabled (p. 75), 1.3 points with mentally retarded (p. 80), and 2.0 points with nonexceptional students (p. 62). However, consistent with the current findings, Lukens (1990) found the Stanford-Binet Test Composite was significantly higher than the WISC-R Full Scale score with mentally retarded students and Rothlisberg (1987) found the Stanford-Binet Test Composite was significantly lower (by 7 points) than the WISC-R Full Scale score with normal second-grade students.

This study's findings may be useful when comparing Stanford-Binet scores with those obtained on a previously administered WISC-R. For instance, a learning disabled student who previously received an average range score on a WISC-R, will likely obtain a significantly lower score if reevaluated with the Stanford-Binet. A significantly lower Stanford-Binet than WISC-R score in this case might best be interpreted as an expected finding and not necessarily as a change in the student's relative level of cognitive functioning.

The results of this study also have implications for diagnostic practices. It appears that individuals similar to the ones in the present study would be more likely to receive a diagnosis of mental retardation or learning disabled when the WISC-R is administered instead of the Stanford-Binet.
This is because individuals who obtain a WISC-R Full Scale score in the mentally retarded range (<70) would likely obtain a significantly higher Stanford-Binet Test Composite that might fall outside this range. For instance, 22 subjects in the present study obtained a WISC-R Full Scale score below 70. Of these 22 subjects, only 8 obtained a Stanford-Binet Test Composite of less than 70. On the other hand, the significantly higher WISC-R Full Scale score than Stanford-Binet Test Composite found with the average ability group would translate to these individuals having a larger ability/achievement discrepancy score with the WISC-R than with the Stanford-Binet. When these results are viewed with Prewett and Giannuli's (1991), who found that although the reading subtests of four achievement batteries were highly correlated, they did not always yield similar scores, the issue of psychoeducational diagnoses being dependent on the test or test battery administered is illustrated.

The unusual relationship between the Stanford-Binet and the WISC-R found in this study appears to be related to the Stanford-Binet's restricted, relative to the WISC-R, range of scores. The reason the Stanford-Binet scores were relatively restricted is unclear. Future research should explore reasons why the Stanford-Binet showed less variability than the WISC-R with students experiencing academic difficulties. This type of research might examine the item difficulty gradients of the Stanford-Binet subtests. For example, a sudden shift in item difficulty, where a group of relatively easy items precedes a group of significantly more difficult ones, would result in a restricted range of scores.

Part of the reason for the discrepant scores between the Stanford-Binet and the WISC-R may be related to differences in the constructs measured by the two intelligence tests. Although the Stanford-Binet Test Composite and the WISC-R Full Scale scores correlated highly, their 53% shared variance shows that they have as much uncommon as common variance. Also, although the 56% and 33% shared variance between the verbal (WISC-R Verbal & Stanford-Binet Verbal Reasoning) and nonverbal (WISC-R Performance & Stanford-Binet Abstract/Visual Reasoning) scales, respectively, of the two tests was significant, the large portion of unshared variance suggests that the constructs assessed by the similar scales are not interchangeable. A joint factor analysis of the subtests comprising the Stanford-Binet and the WISC-R would help determine what is similar and what is dissimilar in the constructs assessed by these instruments.

Peter N. Prewett is Assistant Professor of Psychology, West Florida University. Mark A. Matavich is with the Columbus, Ohio Public Schools.
References


