DOCUMENT RESUME

ED 381 977 EC 303 925

AUTHOR Smith, Douglas K.; And Others
TITLE WISC-III/WISC-R Relationships in Special Education Re-evaluations.
PUB DATE Mar 94
NOTE 19p.; Paper presented at the Annual Meeting of the National Association of School Psychologists (Seattle, WA, March 1994).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Disabilities; Elementary Secondary Education; *Intelligence Tests; *Psychometrics; Scores; Special Education; *Student Evaluation; Test Reliability; Test Use; Test Validity
IDENTIFIERS Special Education Reevaluations; *Wechsler Intelligence Scale for Children (Revised); *Wechsler Intelligence Scale for Children III

ABSTRACT

This study examined the relationship between scores on the Wechsler Intelligence Scale for Children-III (WISC-III) and the older Wechsler Intelligence Scale for Children-Revised (WISC-R). School psychologists in Wisconsin were asked to provide data on 300 special education re-evaluations completed during the 1992-93 academic year. Pearson product moment correlations ranged from .80 on the Verbal Scale to .85 on the Full Scale. T-tests for related samples were significant for all global scales, with WISC-III scores lower than WISC-R scores (mean difference of 3.65 to 5.69 points). The study concluded that: (1) WISC-III scores are consistent with those on the WISC-R; (2) correlational data for global scales and individual subtests suggest that they are measuring similar constructs; (3) individual cases of large differences in scores (16 points or more) may occur; (4) score differences are most likely to occur on the Performance Scale and least likely to occur on the Verbal Scale; (5) largest mean differences on subtests occurred on Similarities, Picture Arrangement, and Object Assembly subtests; (6) smallest mean differences on subtests occurred on Information, Digit Span, and Coding subtests; and (7) use of the WISC-III as a replacement for the WISC-R is strongly supported. Tables and graphs detailing study findings are attached. Four tables and three figures present the data. (DB)

WISC-III/ WISC-R Relationships in Special Education Re-evaluations

Douglas K. Smith

Donald L. Stovall

Brenda L. Geraghty

School Psychology Program

University of Wisconsin-River Falls

Ames Education Center

River Falls, WI 54022

Paper presented at the Annual Meeting of the National Association of School Psychologists, Seattle, Washington, March 1994.

We offer our sincere thanks to the many school psychologists in Wisconsin who participated in this study. Without their help, this study could not have been completed.

ABSTRACT

WISC-III/WISC-R Relationships in Special Education Re-evaluations

Douglas K. Smith

Donald L. Stovall

Brenda L. Geraghty

University of Wisconsin-River Falls

Since the introduction of the WISC-III, practitioners have raised questions concerning the equivalency of scores on the WISC-III and WISC-R in the re-evaluation of students receiving special education services. There have been numerous reports of discrepancies in scores ranging from 10 to 20 points, which is considerably larger than the expected 5 to 7 point difference described in the WISC-III Manual. Therefore, the purpose of this study was to examine the relationship between WISC-III and WISC-R scores using a sample of students being re-evaluated for possible continuation of special education services. School psychologists in Wisconsin were asked to provide data on re-evaluations completed during the 1992-93 academic year. They were asked to submit data on all re-evaluations (or a random sample, if case load was excessive). A total of 300 cases were submitted. Pearson product moment correlations ranged from .80 ($p < .001$) on the Verbal Scale to .85 ($p < .001$) on the Full Scale. T-tests for related samples were significant ($p < .001$) for all global scales with WISC-III scores lower than WISC-R scores (mean difference of 3.65 to 5.69 points). Subtest analyses and the implications of the study for school psychology practice are discussed.

Since the introduction of the Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991), practitioners have raised questions concerning the equivalency of scores on the WISC-III and its predecessor, the WISC-R, in the re-evaluation of students receiving special education services. There have been numerous reports of discrepancies in scores ranging from 10 to 20 points, which is considerably larger than the expected 5 to 7 point difference described in the WISC-III Manual.

The expected difference in WISC-R/WISC-III scores was based on a study reported in the WISC-III Manual in which 206 children, ages 6 through 16 years were administered the WISC-III and WISC-R in counterbalanced order with a median interval between testings of 21 days. Correlations between the WISC-III and WISC-R for the Verbal IQ (VIQ), Performance IQ (PIQ), and Full Scale IQ (FSIQ) were .90, .81 and .89, respectively. Differences in mean scores were 2.4 points, 7.4 points, and 5.3 points, respectively, for the VIQ, PIQ, and FSIQ with WISC-III scores lower on each scale. An additional study, utilizing a clinical sample of 104 students, is also reported in the WISC-III Manual. This sample ranged in age from 7 to 14 years of age, was predominantly male and included "57% children with various learning and reading disabilities, 35% children with Attention-Deficit Hyperactivity Disorder, and 8% children with depression or anxiety disorders" (Wechsler, 1991, p. 199). Corrected correlations between the WISC-R and WISC-III VIQ, PIQ and FSIQ were .86, .73, and .86, respectively. WISC-III scores were lower (5.4 points for the VIQ, 5.1 points for the PIQ and 5.9 points for the FSIQ). Two other studies involving WISC-R/WISC-III studies are reported in the WISC-III Manual. One study involved 23 children previously identified as gifted. The mean interval between the previously administered WISC-R and the current WISC-III was 1

year, 1 month. Mean scores on the VIQ, PIQ, and FSIQ were 5.8, 1.1, and 4.9 points lower on the WISC-III. A similar study with 28 students diagnosed as mildly mentally retarded is also reported. The mean test interval between the previously administered WISC-R and the current WISC-III was 2 years, 2 months. Mean WISC-III IQ scores were 8.9, 6.8, and 8.9 points lower than the mean WISC-R IQ scores on the VIQ, PIQ, and FSIQ, respectively.

Since publication of the WISC-III, several studies comparing WISC-R/WISC-III performance have been completed. Sevier, Bain, and Hildman (1993) compared WISC-R/WISC-III scores for 35 gifted and talented students and found WISC-III FSIQ scores to be 12.83 points lower than WISC-R scores on average. Klein and Fisher (1993) used a sample of 127 students receiving special education services in northeastern Illinois in their study. Current WISC-IIIs (administered as part of the periodic re-evaluation on the students) were compared to the previous administered WISC-Rs. WISC-III scores were lower than WISC-R scores on the VIQ, PIQ and FSIQ by 0.9, 4.3, and 2.7 points, respectively. Bolen, Aichinger, Hall, and Webster (1993) utilized a sample of 61 learning impaired students, due for re-evaluation of their special education placement. Each student was administered the WISC-III and comparisons were made with the previously administered WISC-R. Substantial differences were found in WISC-R and WISC-III VIQ (5.20 points), PIQ (9.21 points), and FSIQ (7.95 points) with mean WISC-III scores lower on all three scales. Finally, Potter and Wessel (1993) used a sample of 118 students being re-evaluated for continued special education services in the Milwaukee, Wisconsin public schools. The sample included all re-evaluations completed between November 1, 1992 and December 12, 1992 and the current WISC-III scores were compared to the previously obtained WISC-R scores. Differences were found on the VIQ (6.07 points), PIQ (8.33 points), and the FSIQ (7.72 points)

with mean WISC-III scores lower than the WISC-R scores on all three scales.

Purpose of the Study

The studies completed to date have produced a mixed pattern of results in comparing WISC-R/WISC-III performance. Sample sizes have been limited in some studies and narrowly defined clinical samples have been used in others. In addition, many of the studies have utilized samples from an individual school district. Consequently, the generalizability of the results to other populations is limited. Therefore, the purpose of this study was to examine the relationship between WISC-III and WISC-R scores using a statewide sample of students who were being re-evaluated for possible continuation of special education services.

Method

School psychologists in Wisconsin were asked to provide data on re-evaluations completed during the 1992-93 academic year. They were asked to submit data on all re-evaluations they had completed (or a random sample of re-evaluations if necessary due to a large case load). Data collected included WISC-III/WISC-R subtest and global scores, sex of the student, race/ethnicity of the student, community size, student age, initial disability of the student, type of program in which the student was enrolled (resource, self-contained, etc.), and the placement decision following the re-evaluation with the WISC-III. Data were coded and double-checked by school psychology graduate students prior to data analysis.

A total of 300 cases was submitted. Students ranged in age from 6 years, 1 month to 14 years, 8 months at initial testing with the WISC-R and from 7 years, 0 months to 17 years, 9 months at the time of re-evaluation with the WISC-III. The sample was composed of 195 males and 88 females (gender was not indicated for 17 students). Services for students with learning disabilities

were being provided for 61% of the sample. Nine percent of the sample was from communities of less than 2,000 residents, 26% from communities of 2,000 to 5,000, and 41% from communities over 5,000 (community size was not indicated for 23% of the sample).

Results and Discussion

Mean standard scores on the WISC-R global scales ranged from 90.54 (VIQ) to 96.81 (PIQ). For the WISC-III, the range was from 86.89 (VIQ) to 91.12 (PIQ). The mean difference in scores was 3.65 points on the VIQ, 5.69 points on the PIQ, and 5.03 points on the FSIQ, with scores on the WISC-III being lower than the WISC-R. These differences in mean global scale scores were analyzed by t-tests for related samples and produced statistically significant results for the VIQ ($t = -6.95$, $p < .001$); for the PIQ ($t = -10.82$, $p < .001$); and for the FSIQ ($t = -11.00$, $p < .001$). The mean difference in scores, however, is similar to the differences reported in the WISC-III manual. These results are presented in Table 1.

Insert Table 1 about here

Mean WISC-R subtest scores ranged from 6.72 (Digit Span) to 10.29 (Picture Arrangement), while mean WISC-III subtest scores ranged from 6.61 (Arithmetic) to 8.94 (Picture Completion). The mean difference in subtest scores ranged from -1.67 to +.39. WISC-III subtest means were lower than WISC-R subtest means for all subtests except Information and Digit Span. The differences in mean subtest scores were analyzed by t-tests for related samples and produced statistically significant results ($p < .05$) for these subtests: Similarities, Arithmetic, Vocabulary, Comprehension, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and

Coding. The WISC-III mean scores were lower than the WISC-R mean scores for all subtests except Digit Span. These results are presented in Table 2.

Insert Table 2 about here

Pearson product moment correlations were calculated for the global scales of the two tests and produced these results: $r = .80$ ($p < .001$) for VIQ; $r = .83$ ($p < .001$) for PIQ; and $r = .85$ ($p < .001$) for FSIQ. These results suggest that the global scales of the two instruments are highly related and that the difference in mean scores, although statistically significant, have limited practical significance. It should be noted that the mean FSIQ and mean VIQ on the WISC-III are less than 90 (87.96 and 86.89, respectively) while the corresponding WISC-R scores were above 90 (92.99 and 90.54, respectively). For states with rigid criteria for placement in learning disabilities programs that require an IQ score of 90 or above, this difference in scores could be problematic.

To further analyze the results of the study, Pearson product moment correlations were calculated on the mean subtest scores of the WISC-R and WISC-III. Significant correlations ($p < .001$) ranging from .42 (Comprehension) to .69 (Block Design) were obtained for all subtests except Mazes ($r = .09$, NS). The mean verbal subtest correlation was .56, while the mean performance subtest correlation was .58 (excluding Mazes due to limited sample size). These results suggest that the subtests of the two instruments are highly related. The differences in mean scores, although statistically significant in many cases, have limited practical significance. The complete correlational results are presented in Table 3.

Insert Table 3 about here

Difference scores (WISC-III score minus WISC-R score) were calculated for the VIQ, PIQ, and FSIQ. These scores ranged from -29 points to +21 points (mean of -3.6 points and standard deviation of 8.96) for the VIQ, from -34 points to +26 points (mean of -5.69 and standard deviation of 9.00) for the PIQ, and from -29 points to +21 points (mean of -5.03 and standard deviation of 7.78) for the FSIQ. For all three global scales, the standard deviations and range in difference scores are relatively large. Consequently, for some students large differences between WISC-III and WISC-R scores will result. The frequency of these differences are examined later in this paper. The difference score results have been plotted against a normal curve and are presented in Figures 1, 2, and 3.

Insert Figures 1, 2, 3 about here

The final analysis involved calculating the frequency of specific score differences on the global scales. Twelve students (4% of the sample) demonstrated no difference in FSIQ on the WISC-R and WISC-III, while 216 students (75% of the sample) produced lower WISC-III scores than WISC-R scores and 61 students (21% of the sample) produced higher WISC-III scores than WISC-R scores. Similar results were obtained for the Verbal and Performance Scales. These data are presented in Tables 4 and 5.

Insert Tables 4 and 5 about here

Chi-square analyses were conducted on the data provided in Table 4 and produced significant ($p < .001$) chi-square values of 150.27, 175.07 and 238.79 for the Verbal, Performance and Full Scales, respectively. An analysis of the data in Table 4 indicates that a student was more likely to obtain a WISC-III score lower than the previous WISC-R score than a WISC-III score that was equal to or higher than the WISC-R score on all three scales (Verbal, Performance and Full Scale).

To further analyze the score differences, additional chi-square analyses were conducted on the data provided in Table 5. Significant ($p < .001$) chi-square values of 76.85, 55.42, and 93.42 were obtained for the Verbal, Performance and Full Scale score differences for students obtaining WISC-III scores lower than the previous WISC-R scores. Thus, the differences were more likely to be in the 1 to 10 point range than in the 11 to 15 point range or higher. Although some students did exhibit differences in scores of over 20 points, these cases were unusual and represented only 5% of the sample of 189 students. Similar analyses were conducted for the students obtaining WISC-III higher than the previous WISC-R results and significant ($p < .001$) chi-square values of 74.78, 46.66, and 72.24 were obtained for the Verbal, Performance and Full Scale score differences. Thus, the differences were more likely to be in the one to five point range than in the 11 to 15 point range or higher.

While this study utilized a larger sample than many of the other studies reviewed earlier, there are limitations in the present study. The sample was primarily white and from rural or suburban areas; large, metropolitan areas are not well represented in the sample. While every effort was made to eliminate a selection bias in the cases included in the study, there is no way to fully eliminate this variable. Finally, this study, as well as others, have depended on previously administered WISC-Rs. Thus, practice effects

cannot be ruled out. The alternative approach, administering both the WISC-III and WISC-R in counterbalanced order, would have greatly reduced the sample size and was not feasible. Additional research, utilizing this methodology, is strongly encouraged, however.

Summary and Conclusions

Using a statewide sample of 300 students being re-evaluated for continued special education placement, the results of this study suggest modest differences in WISC-III/WISC-R scores. Although one fourth of the sample had equal or higher WISC-III scores as compared to WISC-R scores, the "typical" student displayed lower WISC-III scores. Mean differences in scores ranged from 3.65 on the Verbal Scale to 5.69 points on the Performance Scale. The overall FSIQ difference was 5.03 points, which is consistent with studies reported in the WISC-III Manual. Chi-square analyses indicated that the difference in scores was most likely in the 1 to 10 point range. Only 5% of the sample had score differences of 21 or more points.

On the basis of this study and the review of studies completed to date analyzing WISC-III/WISC-R relationships, it may be concluded that:

1. The WISC-III produces scores that are consistent with those generated by the WISC-R. While the WISC-III scores are typically lower than WISC-R scores (usually by 5 to 8 points), they are consistent with expectations based on differences in the age of the standardization samples (1970s vs 1990s).
2. The correlational data for WISC-R/WISC-III global scales and individual subtests suggest that they are measuring similar constructs.
3. In individual cases large differences in scores (16 points or more) between previously administered WISC-Rs and current WISC-IIIs may occur. Such differences, however, occurred in less than 10% of the sample.
4. Differences in scores are most likely to occur on the Performance Scale

and least likely to occur on the Verbal Scale.

5. Largest mean differences on subtests (more than one point) occurred on Similarities, Picture Arrangement, and Object Assembly.

6. Smallest mean differences on subtests (less than one-half point) occurred on Information, Digit Span, and Coding).

7. Use of the WISC-III as a replacement for the WISC-R is strongly supported by the results of this study as well as others that are reviewed in this paper.

References

Bolen, L.M., Aichinger, K. S., Hall, C. W., & Webster, R. E. (1993, April). Comparison of the performance of learning impaired children on the WISC-R and WISC-III. Paper presented at Annual Meeting of National Association of School Psychologists, Washington, DC.

Klein, E. S., & Fisher, G. S. (1993, April). WISC-R and WISC-III: A comparison of IQ, subtest, and factor/index scores. Paper presented at Annual Meeting of National Association of School Psychologists, Washington, DC.

Potter, A., & Wessel, J. (1993, October). Analysis of WISC-III data from an urban population of referred children. Paper presented at Fall Meeting of Wisconsin School Psychologists, Oshkosh, WI.

Sevier, R., Bain, S. K., & Hildman, L. K. (1993, April). Comparison of WISC-R and WISC-III scores in gifted children. Paper presented at Annual Meeting of National Association of School Psychologists, Washington, DC.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children-III: Manual. San Antonio, TX: The Psychological Corporation.

Table 1

Comparison of WISC-III and WISC-R IQ Scores

Scale	n		WISC-III	WISC-R	Mean Difference	t	p
Verbal Scale	291	M	86.89	90.54	-3.65	-6.95	<.001
		SD	(13.74)	(14.47)			
Performance Scale	293	M	91.12	96.81	-5.69	-10.82	<.001
		SD	(15.54)	(15.24)			
Full Scale	289	M	87.96	92.99	-5.03	-11.00	<.001
		SD	(14.34)	(14.38)			

Table 2

Comparison of WISC-III and WISC-R Subtest Scores

Subtests	n		WISC-III	WISC-R	Mean Difference	t	p
Information	266	M	7.82	7.43	+.09	+.56	NS
		SD	2.88	2.94			
Similarities	266	M	8.39	9.80	-1.41	-8.15	<.001
		SD	3.13	3.27			
Arithmetic	265	M	6.63	7.41	-.78	-5.20	<.001
		SD	2.68	2.58			
Vocabulary	265	M	7.26	8.25	-.99	-6.84	<.001
		SD	2.84	2.86			
Comprehension	260	M	8.14	9.09	-.95	-4.77	<.001
		SD	3.10	2.85			
(Digit Span)	174	M	7.05	6.66	+.39	+1.98	.05
		SD	2.43	2.60			
Picture Completion	269	M	9.03	9.98	-.95	-5.61	<.001
		SD	3.08	2.80			
Picture Arrangement	265	M	8.63	10.30	-1.67	-7.78	<.001
		SD	3.80	3.28			
Block Design	269	M	8.29	9.04	-.75	-4.75	<.001
		SD	3.43	3.08			
Object Assembly	260	M	8.49	9.72	-1.23	-6.57	<.001
		SD	3.43	3.29			
Coding	261	M	8.10	8.55	-.45	-2.41	.02
		SD	3.17	2.84			
Mazes	23	M	8.96	10.00	-1.04	-1.19	NS
		SD	4.11	3.25			

Table 3

WISC-III and WISC-R Correlations

Global Scales	n	Pearson Correlation	p
Verbal Scale	291	.80	<.001
Performance	293	.83	<.001
Full Scale	289	.85	<.001
Subtests			
Information	266	.63	<.001
Similarities	266	.61	<.001
Arithmetic	265	.56	<.001
Vocabulary	265	.66	<.001
Comprehension	260	.42	<.001
(Digit Span)	174	.48	<.001
Picture Completion	269	.56	<.001
Picture Arrangement	265	.53	<.001
Block Design	269	.69	<.001
Object Assembly	260	.60	<.001
Coding	261	.50	<.001
(Mazes)	23	.09	NS

Table 4

Summary of Frequency of Score Differences on Global Scales of WISC-III and WISC-R

	Equal Scores	WISC III < WISC-R	WISC III > WISC-R
Verbal Scale	16 (5%)	187 (64%)	89 (31%)
Performance Scale	14 (5%)	222 (75%)	59 (20%)
Full Scale	12 (4%)	216 (75%)	61 (21%)

Note. % refers to percentage of cases for each scale which is 293 for Verbal Scale, 295 for Performance Scale and 286 for Full Scale

Table 5

Frequency of Score Differences on Global Scales of WISC-III and WISC-R

WISC-III Scores Lower Than WISC-R Scores

	1-5 points	6-10 points	11-15 points	16-20 points	21 or more
Verbal Scale	63 (23%)	67 (24%)	33 (12%)	11 (4%)	13 (5%)
Performance Scale	67 (24%)	69 (25%)	44 (16%)	26 (9%)	13 (5%)
Full Scale	80 (29%)	69 (25%)	43 (15%)	17 (6%)	7 (3%)

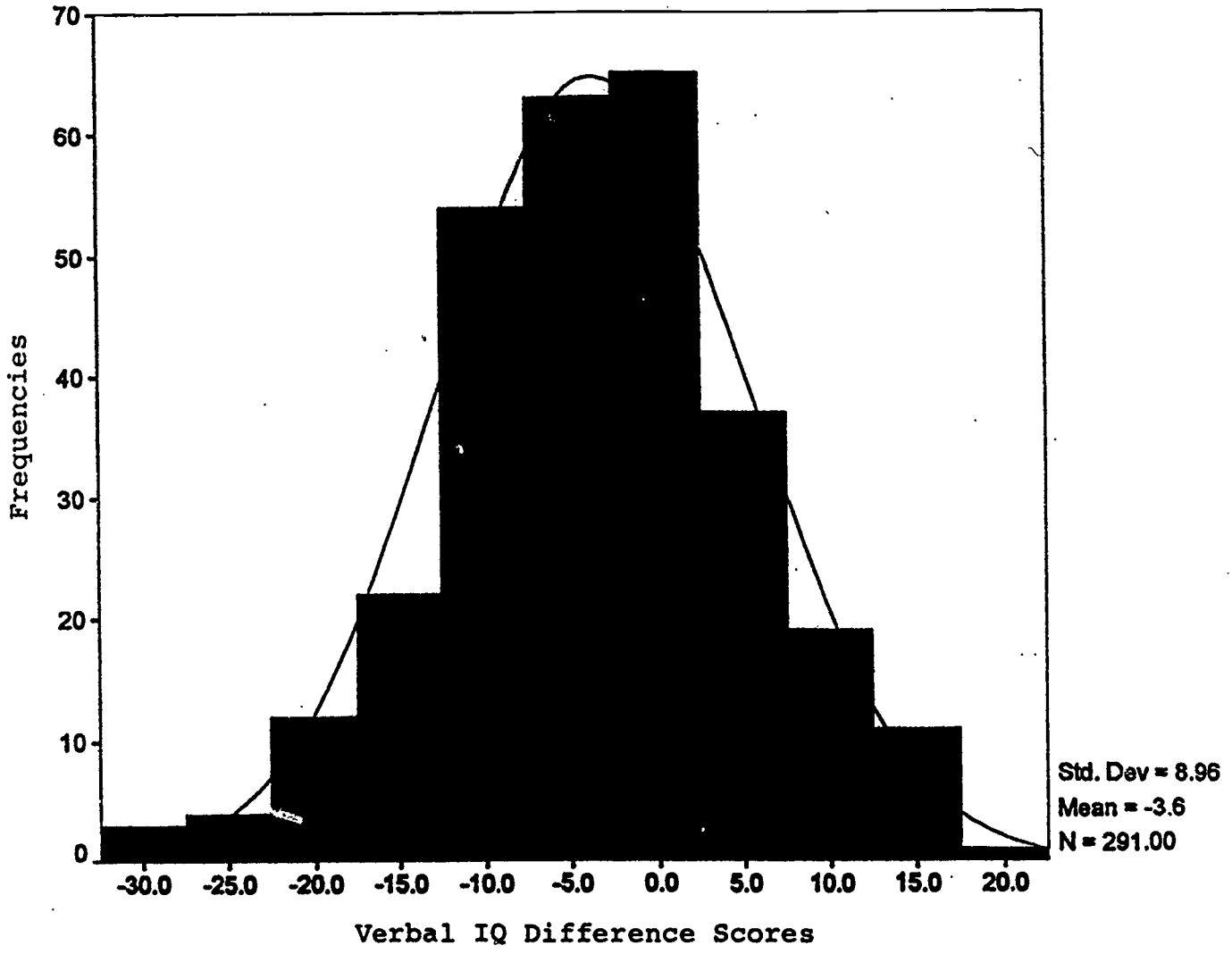
WISC-III Scores Higher Than WISC-R Scores

	1-5 points	6-10 points	11-15 points	16-20 points	21 or more
Verbal Scale	46 (17%)	25 (9%)	12 (4%)	4 (1%)	1(<1%)
Performance Scale	30 (11%)	17 (6%)	9 (3%)	3 (1%)	1(<1%)
Full Scale	39 (14%)	13 (5%)	8 (3%)	--	1(<1%)

Note. % refers to percentage of cases for each scale which is 275 for Verbal Scale, 279 for Performance Scale and 277 for Full Scale

Figure 1

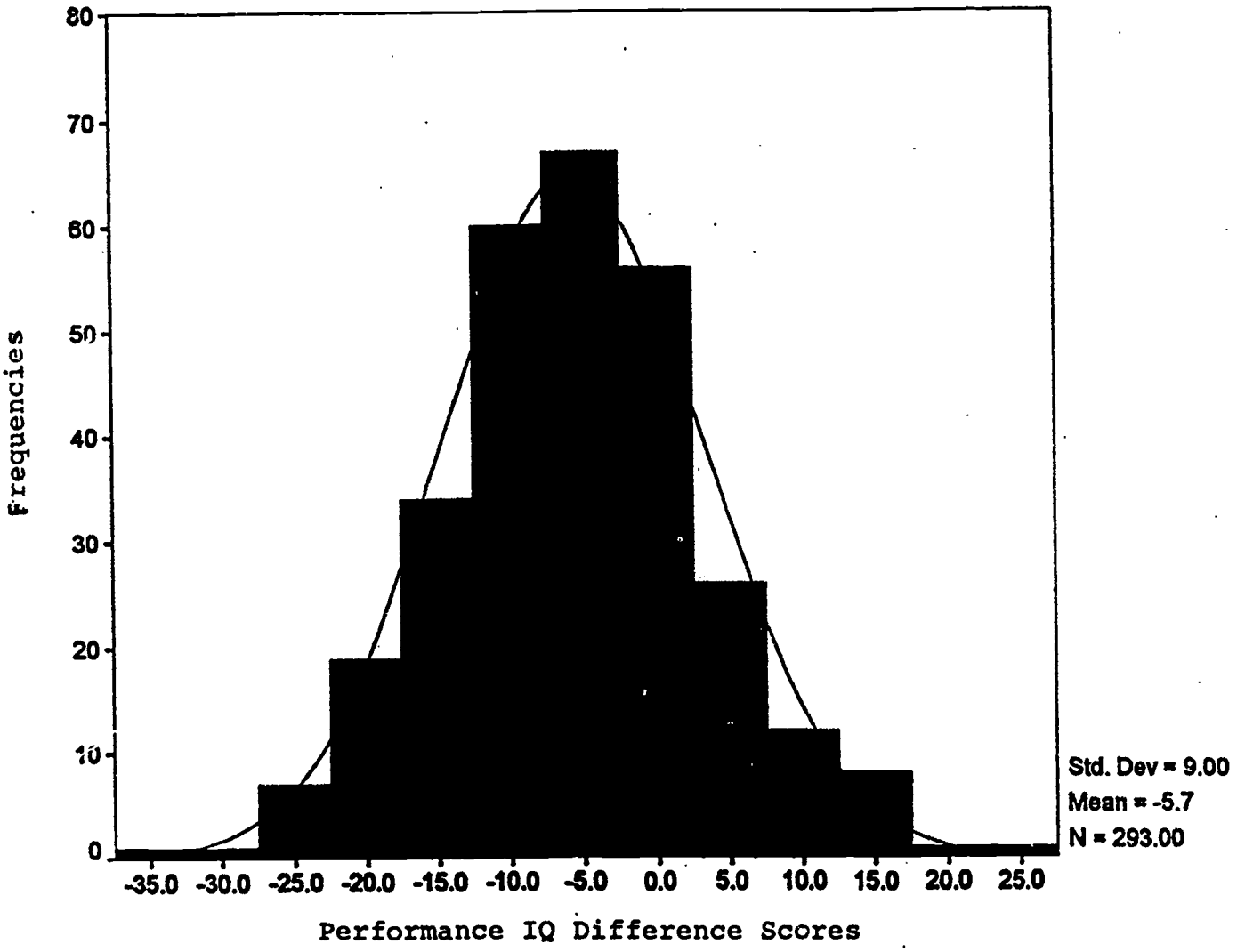
Graph of Verbal IQ Difference Scores



Note. Frequencies on Vertical axis and Verbal IQ Difference Scores on horizontal axis. Normal curve plotted in background.

Figure 2

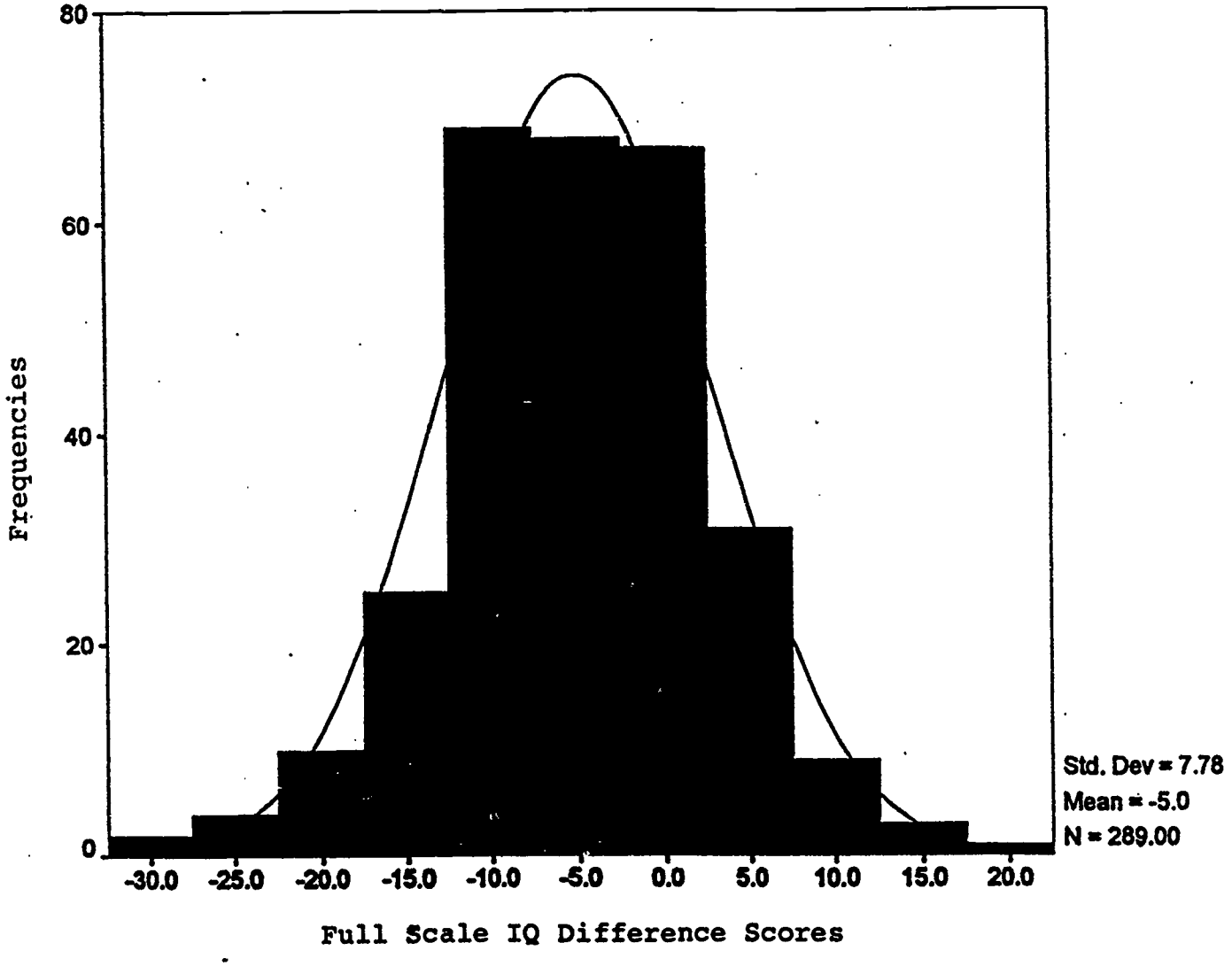
Graph of Performance IQ Difference Scores



Note. Frequencies on vertical axis and Performance IQ Difference Scores on horizontal axis. Normal curve plotted in background.

Figure 3

Graph of Full Scale IQ Difference Scores



Note. Frequencies on vertical axis and Full Scale IQ Difference Scores on horizontal axis. Normal Curve plotted in background.