

Verbal-Performance Discrepancy Scores on the WISC-R, the Lorge-Thorndike Intelligence Test, and the Cognitive Abilities Test

John H. Rosenbach
University at Albany
State University of New York

Mary Veitch
Greater Johnstown School District
Johnstown, New York

This study was designed to address an aspect of the validity of the widely used Verbal-Performance (V-P) discrepancy score from the WISC-R, viz., is this discrepancy score test specific (i.e., idiosyncratic to the WISC-R) or does it appear concordantly on other tests with similar verbal/nonverbal formats? Data from two samples of elementary students, all of whom had been referred for school psychological services, were used. For one sample of 70, scores and V-P discrepancies from the WISC-R were compared with scores and Verbal-Non Verbal discrepancies from the Lorge-Thorndike Intelligence test. For a second sample of 55, similar comparisons were made for the WISC-R and the Cognitive Abilities Test. In both samples, a degree of concordance was demonstrated, reflected in positive correlations between discrepancy scores as well as in a reasonably high agreement between extreme discrepancies on the WISC-R (i.e., $V-P = \pm 15$ or more) and discrepancies of similar signs of five or more on the other tests.

The purpose of this study was to provide information regarding aspects of the validity of the Verbal-Performance (V-P) discrepancy score on the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974). Many clinicians have used, and continue to use, this score for a wide range of diagnostic questions, ranging from those related to psychopathology, brain damage, or delinquency, for example, to those of a more educational nature, such as learning disabilities, reading problems, or underachievement. Moreover, school psychologists will often make recommendations to modify instruction based on the V-P discrepancy (e.g., Kaufman, 1976b, 1979).

This construct has generated considerable research across a broad array of topics, including reliability (Berk, 1982; McNemar, 1957), norming (Kaufman, 1976a), brain damage and psychopathology (Guertin, Ladd, Frank, Rabin, &

Hiester, 1971), delinquency (Haynes & Bensch, 1983; Hubble & Groff, 1981; Ollendick, 1979; Prentice & Kelly, 1963), learning disabilities and other educational topics (Berk, 1983; Kaufman, 1976b; 1979), and acting out behavior (Fernald & Wisser, 1967), among others.

Literally hundreds of studies have appeared in the literature; for the period 1965-1970, alone, Guertin, *et al.* (1971) reviewed 260 such articles. Despite the enormity of this literature (or, perhaps because of it), there is little agreement as to the value of the V-P score. The findings are mixed. Some studies report positive results (Berk, 1983; Hubble & Groff, 1981; Haynes & Bensch, 1983; Prentice & Kelly, 1963) while others report findings that are of neither statistical nor practical significance (Fernald & Wisser, 1967; Ollendick, 1979; Vance, Singer, Kitson, & Brenner, 1983). Furthermore, the literature is characterized by a lack

of concern as to what the V-P discrepancy measures (i.e., its construct validity); the construct seems to have attracted interest because extreme scores are simply different or unusual rather than because of any notions as to what underlies the discrepancy.

One fundamental question that appears to have been overlooked is whether the V-P discrepancy score is test-specific, i.e., idiosyncratic to the Wechsler scales, or whether it will appear, for the same individuals, on other instruments which produce similar language-based and non-language-based scores. The present study was designed to investigate this question by comparing WISC-R discrepancies with those from the Lorge-Thorndike Intelligence Test (Thorndike & Hagen, 1966) and the Cognitive Abilities Test (Thorndike & Hagen, 1983). Both are well-known, group-administered measures which produce a verbal score and a nonverbal score.

Method and Procedures

Two sets of data were collected from elementary students enrolled in a small city school district in Upstate New York. All students had been tested by the district's school psychologists on the WISC-R (WI). In one sample of 70, scores were available on the Lorge-Thorndike Intelligence Test (LT); in the other sample of 55, scores on the Cognitive Abilities Test (CAT) were used. The first sample (LT) represented a fairly typical group of general school psychology referrals. Most of the students were in grades 3 through 6, ages 7 through 12. There were 48 boys and 22 girls. The other sample was composed, primarily, of students who were reevaluated after prior referral and screening. Most of these students were in grades 4 through 6, ages 9 through 13. There were 34 boys and 21 girls. (For both samples, the ages and grade levels listed were at the time of the WI administration. In almost all instances, the LT and CAT had been administered within one year of the WI). Means and standard deviations were computed for each test and subtest, and test and subtest scores were intercorrelated within each data set. In addition, each child's discrepancy scores were

calculated by subtracting the performance or nonverbal scores from the verbal scores. The resulting pairs of discrepancies were correlated and also placed in frequency distributions.

Each set of data is described separately, below, as Set 1 (WI and LT) and Set 2 (WI and CAT).

Results

Set 1.

Descriptive statistics for the WI and LT are presented in Table 1. The WI means are consistently higher than those of the LT by about 4 to 6 points. The standard deviations, however, are quite similar, with the exception of the LT Non-Verbal (NV) standard deviation which at 16 exceeds all others by 2 to 4 points.

The intercorrelations for this data set are presented in Table 2. These correlations are typical of those found with similar tests. As expected, the smallest r 's (.57 and .63) are between, respectively, the WI V and the LT NV and the WI P and the LT V. Excluding intratest r 's, the highest r 's are between WI FS and LT Total ($r = .78$), WI FS and LT V ($r = .76$), and WI P and LT NV ($r = .75$).

The distributions of discrepancy scores for WI V-P and LT V-NV are shown in Table 3. This table contains several interesting results. First, the correlation between the discrepancy scores is .44. Second, the majority of paired discrepancies (37 of 70 or 53%) appears between -9 and +9. Third, the percentage of extreme WI discrepancies of 15 or more (26%, based on 6 positive cases and 12 negative) parallels figures reported by Kaufman (1979) and Sattler (1990). The LT extreme discrepancies total 21% (3 positive and 12 negative). Fourth, of the 12 extreme negative discrepancies (WISC-R), 10 show a negative LT discrepancy of at least 5 points. The other 2 fall into the interval -4 to +4. These results appear in the lower left-hand corner of Table 3. Fifth, of the 6 extreme positive WI discrepancies, 2 are concordant with the LT (i.e., discrepancies of +5 or more), while 3 are discordant, with negative scores of 5 or more. Sixth, of 12 LT extreme negative discrepancies, 7 are

Table 1
Means and Standard Deviations of WISC-R and Lorge-Thorndike (n = 70)

Tests		Means	SDs	Low-High
WISC:	Verbal	98.2	14.1	74-147
	Performance	100.6	12.1	66-139
	Full Scale	100.2	12.3	78-147
LT:	Verbal	92.4	13.6	70-132
	Non Verbal	96.8	16.1	68-146
	Total	94.5	13.4	74-138

Table 2
Intercorrelations of Lorge-Thorndike (LT) and WISC-R (WI) (n = 70)

	LT Verbal	LT Non-Verbal	LT Total	WI Verbal	WI Performance	WI Full Scale
LT Verbal	—	.77	.93	.70	.63	.76
LT Non-Verbal		—	.95	.57	.75	.72
LT Total			—	.67	.74	.78
WI Verbal				—	.73	.91
WI Performance					—	.84
WI Full Scale						—

concordant (WI scores of minus 10 or greater), while 5 are discordant (WI scores of plus 5 or more). Seventh, of 3 LT positive discrepancies, 1 falls into the same interval (15 or greater), with the remaining 2 between -4 to +4.

Set 2.
Means and standard deviations for the 55 subjects for whom WI and CAT data were available are shown in Table 4. In all instances, the values are substantially smaller than in Set 1, reflecting the effects of prior screening for this sample of students, who were being reevaluated. As in Set 1, the WI means are consistently higher than those of the CAT; and, as in Set 1, the standard deviation of the CAT NV is higher than all others.

The correlation matrix for the WI and CAT is presented in Table 5. Because of the reduced variance, the correlations

are much lower than those in Set 1. However, as anticipated, the smallest *r*'s are between the WI V and CAT NV (*r* = .32) and the WI P and the CAT V (*r* = .25); the highest *r*'s are for the WI V and CAT V (*r* = .44), WIP and CAT NV (*r* = .54), and WIFS and CAT NV (*r* = .56). The CAT does not provide a total score and the CAT Quantitative (Q), although shown here, was not used in calculating discrepancy scores.

The distribution of discrepancy scores is displayed in Table 6. In general, the pattern of scores resembles that of Set 1 (Table 3), although there are some striking differences. First, with respect to similarities: the correlation of the discrepancies is .35 compared to .44 in Set 1; 12 cases (22%) show a WI discrepancy of 15 or more (Set 1, 26%); 14 cases (25%) show a CAT discrepancy of 15 or more (in Set 1, LT of 21%). Second, with respect to differences: of the 12

Table 3
Distributions of Discrepancy Scores: WISC-R (V-P) and Lorge Thorndike (V-NV) (n = 70)

	WISC-R: V-P							Totals
	-15 & -	-10 to -14	-5 to -9	-4 to +4	5 to 9	10 to 14	+15	
+15				2			1	3
10 to 14					2			2
5 to 9			2		1	1	1	5
-4 to +4	2	1	5	12	4		1	25
-5 to -9	3		5	5	3		1	17
-10 to -14	2	1	1	1			1	6
-15 & -	5	2			4		1	12
Totals	12	4	13	20	14	1	6	70

Table 4
Means and Standard Deviations of WISC-R and Cognitive Abilities Test (n = 55)

Tests		Means	SDs	Low-High
WISC:	Verbal	94.0	8.9	75-114
	Performance	98.5	8.5	80-118
	Full Scale	95.9	7.4	81-117
LT:	Verbal	89.8	9.4	65-113
	Non Verbal	88.8	13.4	59-130
	Total	87.9	7.8	66-104

Table 5
Intercorrelations of WISC-R (WI) and Cognitive Abilities Test (CAT) (n = 55)

	WI Verbal	WI Perform. ^a	WI Full Scale	CAT N-V ^b	CAT Quant. ^c	CAT
WI Verbal	—	.30	.79	.44	.32	.23
WI Perform. ^a		—	.75	.25	.54	.39
WI Full Scale			—	.42	.56	.32
CAT Verbal				—	.55	.52
CAT N-V ^b					—	.55
CAT Quant. ^c						—

^a WI Performance
^b CAT Non-Verbal
^c CAT Quantitative

extreme WI scores, only one is positive, while in Set 1, 6 of 18 are positive: 10 of 14 CAT extreme scores are positive, while only 3 of 15 LT extremes are on the plus side; 2 WI negative discrepancies are discordant, with CAT positive scores of 5 or more, while there are no such cases in Set 1; and 1 WI positive extreme score is paired with a comparable CAT score, whereas in Set 1, there are 3 discordant pairs; and, of only 4 negative CAT extreme discrepancies, none is discordant, compared to 5 discordant LT and WISC-R pairs.

Summary of Findings.

1. The correlations of the WI to LT and CAT scores are substantial, especially given the sharply reduced range in Set 2. These results support the general contention that these group mental abilities tests are measuring much in common with the Wechsler Scale.

2. The correlations of .44 and .33 between the WI and

LT and WI and CAT discrepancy scores, respectively, are quite impressive, given the reduced reliabilities associated with difference scores.

3. The percentage of extreme discrepancies on the WI are consistent with those reported elsewhere that are based on standardization data. The proportion of extreme discrepancies on the LT and CAT are similar to those on the WI.

4. Comparing WI to LT, of 18 extreme discrepancies (WI), 12 (67%) can be considered highly concordant (i.e., a discrepancy of 5 or more in the same direction) against 3 (17%) that are discordant (i.e., a discrepancy of 5 or more in the opposite direction).

5. Comparing WI to CAT, of 12 extreme discrepancies (WI), 5 (42%) are highly concordant, while 2 (17%) are discordant.

Table 6
Distributions of Discrepancy Scores: WISC-R (V-P) and Cognitive Abilities Test (V-NV) (n = 55)

	WISC-R: V-P							Totals
	-15 & -	-10 to -14	-5 to -9	-4 to +4	5 to 9	10 to 14	+15	
+15	1		1	3	3	1	1	10
10 to 14			2					2
5 to 9	1	1	3	1	2	1		9
-4 to +4	5	2	2	3	4	1		17
-5 to -9	4	1	4	1	1			11
-10 to -14		1	1					2
-15 & -		1	1	2				4
Totals	11	6	14	10	10	3	1	55

Discussion

The general concordance of discrepancy scores from the WI with those from the LT and CAT supports the notion that this construct, at the very least, is not test specific. The findings do not, of course, shed any light on what is being measured by a V>P or V<P discrepancy; they simply suggest that a large discrepancy on the WI (defined here as 15 or more points) is apt to be found in the same direction, although not necessarily of the same magnitude, on another test which has a structure similar to the WI. The results presented here should be viewed within the restraints imposed by problems associated with the reliabilities of difference scores, exacerbated by the effects of regression to the mean. These factors would tend to work against finding high degrees of concordance for extreme discrepancy scores, suggesting that the results can be thought of as being on the conservative side.

Of interest, particularly from a clinical and educational standpoint, are seven cases in Set 1 who all showed negative scores on the LT (V<P) and positive scores on the WI (V>P). One person has an LT discrepancy of minus 15 or more and a WI of plus 15 or more; four have LTs of minus 15 or more and WIs of plus 5 to 9; and two have WIs of plus 15 or more and LTs from minus 5 to minus 14. On further investigation it turned out that each of these seven students had been referred to the school psychologist for having a "reading problem," which could be considered consistent with their discrepancy patterns. The verbal subtest of the LT demands a certain level of reading skill, whereas the WI is administered orally. This finding underscores the importance of having additional information about an individual when deciding whether to rely on a group measure, as teachers and counselors often do.

How the results of this study are interpreted is very likely to depend on an individual's point of view. A conservative stance would argue that, although some evidence has been offered to demonstrate that the Wechsler

construct of verbal-performance discrepancies is not idiosyncratic to the Wechsler tests, we still know little with respect to what these discrepancies measure. However, the results presented here can be viewed as a basis for continued research which focuses on the nature of the V-P discrepancy. Such a research direction may prove to be more fruitful than that of the past, which has been dominated by a somewhat haphazard search for relationships with clinical symptoms.

As in most studies dealing with this topic, the samples used for this study represent select populations, in this instance, elementary-age students who had been referred for school psychological services. The issue of generalizability, however, is not simple; the selection of a sample depends, in part, on the population to which one wishes to generalize and, also, on the questions which are being addressed. With respect to populations, it can be argued that clinical samples are appropriate when one wishes to generalize to clinical populations. On the other hand, if one is attempting to explore a construct in order to enhance its understanding, especially from a theoretical point of view, then, samples from a general population may also be required.

References

- Berk, R. A. (1982). Verbal-performance IQ discrepancy score: A comment on reliability, abnormality, and validity. *Journal of Clinical Psychology, 38*, 638-640.
- Berk, R. A. (1983). The value of WISC-R profile analysis for the differential diagnosis of learning disabled children. *Journal of Clinical Psychology, 39*, 133-136.
- Fernald, P. S., & Wisser, R. E. (1967). Using WISC verbal-performance discrepancy to predict degree of acting out. *Journal of Clinical Psychology, 23*, 92-93.
- Guertin, W. H., Ladd, C. E., Frank, G. H., Rabin, A. J., & Hiester, D. S. (1971). Research with the Wechsler Intelligence Scales for Adults: 1965-1970. *The Psychological Record, 21*, 289-339.
- Haynes, J. P., & Bensch, M. (1983). Female delinquent

- recidivism and the P>V sign on the WISC-R. *Journal of Clinical Psychology*, 39,141-143.
- Hubble, L. M., & Groff, M. (1981). Magnitude and direction of WISC-R verbal-performance IQ discrepancies among adjudicated male delinquents. *Journal of Youth and Adolescence*, 10, 179-184.
- Kaufman, A. S. (1976a). Verbal-Performance IQ discrepancies on the WISC-R. *Journal of Consulting and Clinical Psychology*, 44, 739-744.
- Kaufman, A. S. (1976b). A new approach to the interpretation of test scatter on the WISC-R. *Journal of Learning Disabilities*, 9, 160-168.
- Kaufman, A. S. (1979). *Intelligent Testing With The WISC-R*. New York: Wiley-Interscience.
- McNemar, Q. (1957). On WAIS difference scores. *Consulting Psychology*, 21, 239-240.
- Ollendick, T. H. (1979). Discrepancies between verbal and performance IQs and subtest scatter on the WISC-R for juvenile delinquents. *Psychology Reports*, 45, 563-568.
- Prentice, N. & Kelly, F. (1963). Intelligence and delinquency: A reconsideration. *Journal of Social Psychology*, 60, 327-337.
- Sattler, J. M. (1990). *Assessment of Children*. San Diego, CA: Jerome M. Sattler, Publisher.
- Thorndike, R. L., & Hagen, E. (1966). *Technical Manual: Lorge-Thorndike Intelligence Test*. New York: Houghton Mifflin.
- Thorndike, R. L., & Hagen, E. (1983). *Cognitive Abilities Test: Examiner's Manual*. Chicago: Riverside Publishing Co.
- Vance, B., Singer, M. G., Kitson, D. L., & Brenner, O. C. (1983). WISC-R profile analysis in differentiating LD from ED children. *Journal of Clinical Psychology*, 39, 125-131.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.

John Rosenbach is Professor of Educational Psychology at the University at Albany, State University of New York, Albany, NY.

Mary Veitch is School Psychologist in the Greater Johnstown School District, Johnstown, NY.