DOCUMENT RESUME

ED 178 572 TM 009 720

AUTHOR Tabachnick, Barbara Gerson
TITLE Test Scatter on the WISC-R in Learning Disabled Children.
PUB DATE Apr 79
NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences; Elementary Secondary Education; Intelligence Differences; *Intelligence Tests; *Learning Disabilities; Norms; Performance Tests; *Scores; Sex Differences; Test Interpretation; *Test Results; Verbal Tests
IDENTIFIERS *Test Scatter; *Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

Scatter on the revised Wechsler Intelligence Scale for children (WISC-R) was evaluated for 105 learning disabled (LD) children. Scatter was defined as range of scaled scores on: (1) five regular verbal tests; (2) five regular performance tests; and (3) all 10 regular subtests. Pairwise combinations of 11 subtests were also evaluated for deviations in scaled scores. A discrepancy of 3 or more points was defined as significant deviation. A significantly greater range of scaled scores over the 10 subtests was found for the LD sample as compared with the 1976 standardization sample for the WISC-R. On the separate major scales, LD children on the average showed greater range over performance, but not verbal scales. Pair deviations for LD children were generally similar to those of normative children Most notably, all pairs including coding showed greater deviation for LD children. Among LD children, differences in test scatter were associated with age, but not IQ or sex. LD children under 11 years showed a greater range of verbal scale scores than older LD children. All statistically significant differences were of small magnitude, so that test scatter, by itself, would not be recommended as a major diagnostic sign. (Author/GDC)

Test Scatter on the WISC-R
in Learning Disabled Children

Barbara Gerson Tabachnick

California State University, Northridge

and

California Center for Educational Therapy

Running head: Test Scatter on the WISC-R

Test Scatter on the WISC-R in Learning Disabled Children

Diagnosis of pathology on the basis of inter-subtest scatter on the WISC-R has been brought into question by Kaufman (1976). As he rightly points out, amount of scatter can only be assessed in terms of a baseline produced by children without known pathology. He conveniently provides this baseline in the form of an index of scatter: the size of the difference between the highest and lowest scaled scores earned by the child.

By providing means and standard deviations for this index, classified by several demographic variables, Kaufman has given clinicians a potentially useful tool for evaluating WISC-R scatter. A serious question remains, however. Do children with identified pathology produce more test scatter than found by Kaufman in his baseline standardization sample? Unless it can be shown that learning disability, for example, is associated with significantly more (or less) test scatter than can be expected without pathology, the simple finding of deviation from normal is of limited use in diagnosis of learning disability.

For a sample of 41 Hawaiian learning-disabled children studied by Anderson, Kaufman, & Kaufman (1976) the answer was no. Differences in scatter on the WISC-R between LD and Kaufman's (1976) normative children were found to be not statistically significant. A few questions about this study, however, suggest the desirability of replication. First, a sample size of 41 seems small in the light of variability among LD children. Given this limited statistical power, the fact that all three measures of scatter on the LD children differed in the predicted direction (i.e., greater amount of scatter) from normal becomes interesting. For example, if these sample differences were to hold over replication, an increase to a sample size of 100 would produce statistical significance for the measure of total test scatter.

Another question in the Anderson, et al study arises with respect to the representativeness of the sample. Examination of WISC-R results for these Hawaiian LD children suggests that they may not be typical of the larger U.S. LD population. Means in such verbal subtests as Information, Similarities, and Vocabulary seem exceptionally low (all below 7) for populations typically described as LD. Further, the Full Scale IQ mean of 84 and standard deviation of 15 suggests that this sample includes a disproportionate number of children of below average IQ, if not some who could be classified as mentally retarded. Educators who include in the definition of LD the concept of "at least average intellectual potential" would, then, probably consider the Anderson, et al sample to be unrepresentative. Do the findings of Anderson, et al generalize to LD populations of more nearly average overall intellectual functioning? It is not clear that they do. Indeed, it is in these latter populations that subtest scatter might be expected to be more pronounced.

The current study was designed to investigate scatter produced by a sample of learning-disabled children with average intellectual potential, and to compare WISC-R scatter with that produced by Kaufman's (1976) normative population. A second objective of the study was to investigate differences in subtest scatter among LD children as a function of age, sex, and IQ.

METHOD

Subjects

One hundred and five children with identified learning disability were selected from two private non-profit agencies dealing with such children.[1] These children had been diagnosed as LD on the basis of a variety of factors: difficulty in school, performance on an extensive battery of psychodiagnostic tests, and judgment of experienced diagnosticians. WISC-R scatter, while it may have been considered, was not a primary determinant in diagnosis.

All children who were referred to either agency between the time the WISC-R became available and the summer of 1977, and who had been administered the WISC-R as part of the psychodiagnostic battery at the agency, were considered for the sample. Those who deviated from the typical LD syndrome on the basis of mental retardation (IQ < 75), severe emotional disturbance, etc., were excluded from the sample. The final sample consisted of 22 girls and 83 boys. The children ranged in age from 6-9 to 16-5 with a median of 10-10. Full scale WISC-R IQ ranged from 75 to 141.

Procedure

All children in the sample had been administered the WISC-R by a single psychometrician, who routinely tests children for a variety of purposes in addition to identification of learning disability.

Indices of test scatter were computed for each of the 105 children in the manner described by Kaufman (1976). Due to sample size limitations, distributions were not classified on the basis of all of the demographic variables used by Kaufman. Classification categories in the present sample were based on sex, age (under 11, 11 and over), and Full Scale IQ (under 90, 90-109, 110 and over).

Kaufman provided data for the full WISC-R with 12 subtests, as well as a reduced WISC-R based on 10 subtests (excluding Digit Span and Mazes). Since Mazes are routinely omitted in testing our sample of children, data comparable to Kaufman's analysis of 12 subtests were unavailable. We therefore eliminated Digit Span scores from analysis for the index of scatter thus providing a test comparable to Kaufman's "regular" 10 subtest WISC-R.

For the index, range of scaled scores, the difference between high and low (H-L) scaled scores was computed for each child for three Scales: (1) the five regular Verbal tests (Information, Similarities, Arithmetic, Vocabulary, Comprehension); (2) the five regular Performance tests (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding); and (3) the H-L scaled score on all 10 regular subtests. For each of these three scaled score ranges, means and SDs were calculated for the total sample. These three measures provided data for tests of the major hypotheses. The three scaled score ranges were also computed separately for classification categories based on age, sex, and IQ.

Following Kaufman, another aspect of inter-test variability was examined. Pairwise combinations of eleven subtests were evaluated for deviations in scaled scores. (Digit Span data were included here for additional information in comparing our sample with Kaufman's.) A discrepancy of 3 or more points between a pair of tests was defined as significant deviation, and the proportion of children showing such deviation was computed between each pair of WISC-R subtests.

Statistical significance of difference between Kaufman's normative data (N=2200) and our own total learning-disabled sample (N=105) were tested using two-tailed, two sample t tests. Investigation of differences within our LD sample was based on a stepdown analysis (Bock & Haggard, 1968, pp. 112-113), which provides a test of the multivariate hypothesis of no difference on various measures of scatter as a function of age, sex, and IQ groupings.

RESULTS

A profile of the children in terms of WISC-R means and SDs is provided in Table 1. This is a sample characterized by overall average intellectual

Insert Table 1 about here

functioning. The depressed scores noted in the Anderson, et al (1976) data for Information, Similarities, and Vocabulary are not characteristic of the present sample. As a matter of fact, the only suspiciously low score, on the average, for these children is Coding. This is consistent with the classic picture of a "perceptually handicapped" view of learning disabilities.

LD vs. Normative Children

Scaled Score Range. Means and SDs of the range of the scaled scores on 10 subtests of the WISC-R are shown in Table 2, for LD and normative children.

Insert Table 2 about here

For the major hypothesis test, the means for subtest range based on entire samples were compared for LD vs. normative children. Range was significantly greater for the LD group (mean = 7.7) than for the normative population (mean = 7.0), t (2303) = 3.41, $p < .01$ indicating more inter-subtest scatter for LD children, although the magnitude of the difference is small.

It should be noted that the data failed the test for homogeneity of variance. That is, the range of scaled scores for LD children showed significantly greater variance ($SD^2 = 7.84$) than did that of normative children ($SD^2 = 4.41$), $F\ (104, 2199) = 1.78$, $p < .01$. Because of the enormous number of degrees of freedom, this failure does not invalidate the t test. The finding, however, is interesting in its own right in that it suggests that LD children can occasionally show exceptionally narrow range of scaled scores, although on the average their range is greater than for normal children.

For the WISC-R broken down into Verbal and Performance subtests, means and SDs of the range of scaled scores are shown in Table 3.

Insert Table 3 about here

As additional hypothesis tests[2], differences between sample means for subtest range for LD vs. normative children were compared individually for Verbal and Performance scales. For both Verbal and Performance comparisons, homogeneity of variance was supported. For Verbal scale, no statistically significant difference between mean ranges was found, $t\ (2303) = 1.05$, $p > .05$, indicating no more scatter among Verbal tests for LD children than for normal children. For the Performance scale, however, LD children did show significantly more scatter as measured by scaled score range (mean = 6.1) than did normal children (mean = 5.5), $t\ (2303) = 2.60$, $p < .01$, although again the magnitude of the difference was small.

<u>Pair Deviation.</u> As another index of scatter, deviation between pairs of subtests on the WISC-R was examined. Significant difference was defined as a three or more point difference between the pair of subtests. Table 4 shows

the percent of the 105 cases which have significant differences between each

Insert Table 4 about here

pair of subtests on the WISC-R. Data for Digit Span are included, but not for Mazes.

For Kaufman's (1976) normative sample, the average percent of deviating cases over all subtest ~~pairs was~~ approximately 44%. Among Verbal subtests, the range of percent of cases was 29-48%. For Performance subtests, the range was 35-51%, and for pairs spanning Verbal and Performance, the range was 40-51%. As can be seen from Table 4, our ranges are comparable with only a few exceptions. For Block Design vs. Vocabulary, our 55% of cases seems clearly outside the normative range, as are our Digit Span vs. Vocabulary at 55% and Digit Span vs. Picture Arrangement at 61%. Further, all of the pairs involving Coding are out of normal range, with only differences involving comparisons with Arithmetic and Picture Arrangement even close to the values typical of normative children. This corroborates the finding, illustrated in Table 1, that for this LD sample, discrepant performance on the Coding subtest is symptomatic.

Differences Among LD Children.

A stepdown analysis was performed to assess WISC-R scatter as a function of age, sex, and IQ groupings within the sample of LD children. The three measures of scatter used were scaled score ranges on (1) 10 subtests of the WISC-R, (2) five Performance subtests, and (3) five Verbal subtests, respectively, as shown in Table 2.

A stepdown test requires an a priori ordering of variables, and allocation of Type I error rate. A hierarchical analysis is then used to test a multivariate hypothesis with a total Type I error rate under a specified value, in this case 5%. This procedure prevents the finding of several "significant" effects among correlated variables which reflect similar behavior. Each measure of scatter is evaluated in turn for added significance, over and above higher priority measures. In this analysis, the finding of a single significant stepdown F allows rejection of the hypothesis of no differences among groups, with groups in this case formed by classification into young vs. older, male vs. female, and three levels of IQ.[3] Results of the analysis appear in Table 5.

Insert Table 5 about here

As shown in Table 5, differences in test scatter were not significantly associated with either IQ or sex differentiation. Age group, however, was significantly associated with test scatter. LD children who were less than 11 years old produced a greater range of Verbal Scale scores over the 5 subtests (adjusted verbal range mean = 5.4) than those who were 11 years or older (adjusted verbal range mean = 4.5). The difference is small, however. As evaluated by η^2, only 2% of the variance in Verbal Scale scores is accounted for by variation in age group, after adjusting for effects of sex, IQ group, Full Scale IQ range, and Performance Scale IQ range.

DISCUSSION

Do LD children produce more inter-subtest scatter on the WISC-R than the "normal" standardization sample? The answer for our sample is yes, but not much. Looking at the total range of scaled scores, the difference between LD and standardization children is reliable and consistent across breakdown by age, sex, and IQ groupings. The difference is being produced by LD children showing more scatter within Performance subtests, and between Verbal and Performance tests, rather than within Verbal subtests. On the average, these are children who are inconsistent in mastery of skills measured by Performance subtests, and furthermore tend to have particular difficulty in Coding.

Can one, then, diagnose a learning disability on the basis of subtest scatter alone? Clearly the answer is no. The overlap in subtest scatter between LD and normative children is substantial and, as mentioned previously, some LD children in our sample are characterized by exceptionally low scatter. By itself, then, range of scaled scores is of little diagnostic value. As one of several inputs to a diagnosis, however, range of scaled scores may add diagnostic information that is not provided by individual scores themselves. Utility of this supplementary information may also be enhanced by knowledge of the child's demographic characteristics. For example, greater verbal subtest scatter in our LD sample is associated with younger (CA < 11) children. Since Kaufman (1976) reported no such differences associated with age in the standardization sample, it may be that subtest scatter is particularly useful diagnostically for younger children. In any event, the use of subtest scatter should be explored as a bit of data in statistical approaches to diagnosis, say as an input to a discriminant function analysis. Here the addition of these derived scatter scores to those directly produced on diagnostic instruments may reliably enhance classification into diagnostic categories.

As suggested by Tabachnick and Tabachnick (1976) however, the real LD problem is probably not in differentiating these children from "normals", but in forming categories relevant to treatment within the LD diagnosis. Is it possible that LD children who show a high degree of WISC-R scatter are more amenable to certain types of educational therapy? For example, it may be that the classic "perceptually handicapped" child is the one associated with greatest WISC-R scatter on Performance Scale and is meaningfully helped through a program of supplemental visual-perceptual retraining. At the same time another child who is characterized by difficulties in the more strictly cognitive domain may not show exceptionally high scatter and may not be best helped by the same remedial program. It is in this realm that consideration of WISC-R scatter--indeed consideration of all results of psychoeducational evaluation--may be especially fruitful.

Future research, then, could profitably be addressed to investigation of differences between LD and normal children on the basis of a number of variables including WISC-R scatter, and to differences in WISC-R scatter among LD children with different kinds of learning difficulties.

REFERENCES

Anderson, M., Kaufman, A. S., & Kaufman, N. L. Use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 1976, 13(4), 381-386.

Kaufman, A. S. A new approach to the interpretation of test scatter on the WISC-R. Journal of Learning Disabilities, 1976, 9(3), 160-168.

Tabachnick, B. G., & Tabachnick, K. So perceptual tests won't solve all our problems: A reply to Larsen, Rogers, and Sowell. Journal of Learning Disabilities, 1976, 9(7), 455-456.

Footnotes

1. Thanks are due to Kenneth Tabachnick of the California Center for Educational Therapy in Woodland Hills, Ca. and to Robert Roy Houghton of the Hillside Developmental Learning Center in La Canada, Ca., for graciously providing access to the data used in this study. Research assistance was provided by Daniel Wick, Sherri Teffeteller, and William Gott, under funding available through the first-named Center above, and the Federal work-study program. Computing assistance was provided through the CSUN Computer Center.
2. Lack of full data for the normative children precluded a more appropriate multivariate test of the hypothesis of no difference between LD and normative children.
3. Analysis was run through SPSS ANOVA. Interactions among grouping variables were ignored after preliminary univariate analyses revealed no significant interactions. Each effect was tested by first adjusting for effects of the other two grouping variables. Then scatter variables were tested by setting higher priority variables, if any, as covariates.

Table 1

Means and SDs on the WISC-R

for LD Sample (N=105).

WISC-R Test	Mean	SD
Verbal		
Information	10.2	3.2
Similarities	10.6	3.1
Arithmetic	9.3	2.8
Vocabulary	11.0	3.0
Comprehension	10.2	2.9
Digit Span (N=98)	8.5	2.6
Performance		
Picture Completion	10.2	2.7
Picture Arrangement	10.3	2.8
Block Design	10.4	3.2
Object Assembly	10.5	2.6
Coding	7.9	2.8
Verbal IQ	101.3	14.9
Performance IQ	98.7	12.3
Full Scale IQ	100.2	13.1

Table 2

Means and SDs of Child's Scaled-Score Range on Regular WISC-R Subtests for LD Sample and Normative Population

	Scaled Score Range					
	Learning Disabled			Normative*		
	N	Mean	SD	N	Mean	SD
ENTIRE SAMPLE	105	7.7	2.4	2200	7.0	2.1
AGE GROUP						
Under 11	55	8.2	2.4	1000	7.0	2.1
11 and over	50	7.2	2.2	1200	7.0	2.2
SEX						
Males	83	7.9	2.5	1100	7.1	2.2
Females	22	7.1	2.0	1100	7.0	2.1
WISC-R Full Scale IQ						
110 and above	25	8.6	2.7	574	7.3	2.0**
90-109	57	7.5	2.2	1089	7.0	2.1
89 and below	23	7.3	2.3	537	6.8	2.2**

*From Kaufman (1976)

**Estimated

Table 3

Means and SDs of Child's Scaled-Score Ranges on

Five Verbal and Five Performance Subtests

for LD Sample and Normative Population

		Scaled Score Range			
	N	Verbal		Performance	
Learning Disabled		Mean	SD	Mean	SD
ENTIRE SAMPLE	105	4.7	2.0	6.1	2.5
AGE GROUP					
Under 11	55	5.3	2.1	6.4	2.6
11 and over	50	4.0	1.6	5.7	2.4
SEX					
Males	83	4.7	1.9	6.4	2.5
Females	22	4.7	2.1	5.0	2.3
WISC-R Full Scale IQ					
110 and above	25	4.5	1.9	7.4	3.0
90-109	57	4.7	1.9	6.0	2.3
89 and below	23	4.8	2.1	5.3	2.2
Normative*					
ENTIRE POPULATION	2200	4.5	1.9	5.5	2.3

*From Kaufman (1976)

Table 4

Percent of Cases Having Significant Differences

Between Each Pair of Tests*

	I	S	A	V	C	DS	PC	PA	BD	OA	Cd
Verbal											
Information											
Similarities	35										
Arithmetic	37	42									
Vocabulary	39	32	44								
Comprehension	40	35	45	32							
Digit Span	47	49	42	55	45						
Performance											
Picture Comp.	46	38	48	35	40	48					
Picture Arrange.	45	50	46	47	48	61	45				
Block Design	53	43	55	45	37	51	35	41			
Object Assembly	42	37	47	44	44	51	31	50	36		
Coding	61	64	52	63	60	52	59	53	60	65	

*N=105, except for pairs with Digit Span where N=98.

Table 5

Stepdown Analysis of Three Measures of Scatter by

Age, Sex, and IQ Level

Measure of Scatter	Grouping Variable	Univariate F[a]	df	Priority	Stepdown F	df	Critical F[b]	α
TOTAL	AGE	4.63	1/103	1	0.20	1/100	5.29	.025
RANGE	SEX	2.06	1/103		1.22	1/100	5.29	.025
	IQ	2.37	2/102		2.03	2/100	3.93	.025
PERFORMANCE	AGE	2.03	1/103	2	3.34	1/99	7.08	.01
RANGE	SEX	6.12	1/103		3.59	1/99	7.08	.01
	IQ	3.05	2/102		4.53	2/99	4.98	.01
VERBAL	AGE	13.71	1/103	3	7.16**	1/98	7.08	.01
RANGE	SEX	0.01	1/103		0.27	1/98	7.08	.01
	IQ	0.08	2/102		0.15	2/98	4.98	.01

**P < .01

[a] Univariate F does not reflect adjustment for effects of other grouping variables.

[b] Conservative, based on 60 df for error.

20