1986

Stability of the Wechsler Intelligence Scale for Children-Revised with Learning Disabled Children

Martha J. Lally

STABILITY OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-REVISED WITH LEARNING DISABLED CHILDREN

BY

MARTHA J. LALLY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

IN

PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

1986

Abstract

The stability of the Wechsler Intelligence Scale for Children-Revised (WISC-R) over a three-year time span for 60 learning disabled junior high students was investigated. The subjects were classified using an ability/achievement discrepancy formula between the WISC-R and Metropolitan Achievement Test (MAT) and placed in either a resource or full-time special education program. The learning disabled subjects demonstrated stable WISC-R scores with retesting, regardless of special education placement. Differences were found between the special education placements on WISC-R scores, MAT scores, years in special education, and years between testing. Implications for the special education reevaluation process are discussed.

Table of Contents

List of Tables

BACKGROUND

The Wechsler Intelligence Scale for Children-Revised (WISC-R) has universally been regarded as a stable and reliable assessment instrument in measuring children's intelligence. It has emerged as the clear-cut choice in the identification of learning disabled children. Based on the federal definition, a child is diagnosed as learning disabled if there is a significant discrepancy between ability and achievement in one or more of the following areas: reading comprehension, basic reading skills, oral expression, listening comprehension, written expression, mathematics calculation, or mathematics reasoning (United States Office of Education, 1977). This discrepancy is operationalized by administering an IQ test, such as the WISC-R, and an achievement test.

Once a child is diagnosed as learning disabled, educational services need to be provided as stated by PL 94-142 (Education for All Handicapped Children Act, 1975). The law also requires that children involved in special education be given a full psychoeducational assessment at least every three years. Consequently, children receiving special education services are given a WISC-R at least every three years for as long as they receive those services. For some children, this may result in four or even five WISC-R administrations in their record. One would assume that since

the WISC-R has proven reliable, repeated testing with learning disabled children is not necessary. However, the WISC-R may not be as reliable with learning disabled children as normal children. The test-retest stability of the WISC-R for learning disabled children, particularly those receiving special education services, has not been conclusively demonstrated. If the WISC-R is found to be as stable with learning disabled children as it is with the normed population, administering the WISC-R every three years might not be necessary. It can be safely assumed that the student's ability level will not change over the three years, and only the other aspects (e.g., emotional and academic) of the student's functioning need be assessed. The release from this testing might allow school psychologists to provide more extensive services to all students.

Other support for not fully retesting learning disabled children every three years is presented by Galvin and Elliott (1985) who surveyed the practices, results, and importance of school reevaluations of handicapped children. For those children classified as learning disabled, 100 school psychologists estimated low levels of diagnostic change (2%-3%), and placement change (less than 4%) with retesting. Based on the psychologists' estimates of their reevaluation caseloads (15-19) per year, this averages less than one actual change per year. The authors argued that the relative low incidence of changes in diagnosis and placement challenges the

necessity for automatically reevaluating handicapped children with a comprehensive evaluation every three years.

Reliability testing of the WISC-R began with Wechsler (1974) who retested 303 normal children after a one-month interval. This retesting produced on all three scales IQ point gains which were atttributed to practice effects. The test-retest reliability estimates, however, remained high: .93 Verbal, .90 Performance, and .95 Full Scale. No additional stability studies were done by Wechsler as the WISC-R was highly correlated with the original WISC which had been standardized on thousands of children. The test's stability has essentially remained unquestioned except for a few reliability studies. Tuma and Appelbaum (1980) retested forty-five normal children ranging in age from 7.8 years to 15.0 years on the WISC-R after a six month interval. They found significant gains in the Performance and Full Scale scores, and high correlations: .95 Verbal, .89 Performance, and .95 Full Scale. Practice effects were again considered responsible for the IQ point gains.

There has been some research in testing the reliability of the WISC-R with learning disabled children. Covin (1977) investigated the stability with 30 learning disabled nine year olds. They were tested after one day intervals resulting in correlations of .83 Verbal, .84 Performance, and .85 Full Scale. In another study, a six month interval was used in retesting 160 learning disabled children ranging in age from

6.3 years to 12.1 years (Smith & Rogers, 1978). Reliability estimates were: .82 Verbal and Performance, and .79 Full Scale. Using a number of tests measuring intellectual, academic, and affective assessments, it was concluded that test results for learning disabled children were as reliable as for any children.

Stability of the WISC-R with learning disabled children was also investigated by Smith (1978). A seven month retesting period involving 161 learning disabled children resulted in a test-retest profile stability of .94. A retesting time span of 2.2 years was reported in a study (Vance, Blixt, & Ellis, 1981) where forty-five learning disabled and thirty mentally retarded children were retested. The resulting correlations were .80 Verbal, .91 Performance, and .88 Full Scale. The authors concluded that clinicians can be reasonably sure that a subsequent prognosis based on IQ will not change appreciably. This study did not make a distinction between the learning disabled and mentally retarded children, however, and consequently the stability may have occurred with one of the groups and not the other.

Elliott, Piersel, Witt, Argulewicz, Gutkin, and Galvin (1985) examined the long-term stability of WISC-R IQs for handicapped subjects from three racial or ethnic groups (Anglo, black, Mexican-American). A total sample of 382 students ranging in age from 6.1 years to 16.9 years, and categorized as learning disabled, behaviorally impaired,

mentally retarded, or unclassified were reevaluated after three years. The stability coefficients of the subjects (VIQ=.81, PIQ=.78, FSIQ=.85) compared well with those established during the standardization of the WISC-R. Specific findings indicated the Anglo subjects' IQs were significantly more stable than the blacks or Mexican-Americans. This study did not differentiate the learning disabled children from the other handicapped subjects, however, and this important distinction needs to be addressed.

In reviewing the literature on learning disabilities, it has generally been noted that most classroom learning disabled programs are not composed of true learning disabled children; that is, the children do not have a significant discrepancy between their ability and achievement as stated by the federal definition. Shepard, Smith, and Vojir (1983) examined 800 students classified as learning disabled in Colorado. They found that only 42.6% of those students exhibited characteristics consistent with definitions of learning disabled in federal regulations and professional literature. The majority of the remaining 57.4% of children did have learning problems, but were incorrectly labelled learning disabled. Characteristics they showed included: mild mental retardation, emotional disturbances, low achievement due to language interference, low achievement due to other causes than psychological disabilities, and other miscellaneous

handicaps.

In an additional study, fifty fourth grade children identified as learning disabled and forty-nine children not classified, but who scored at or below the 25th percentile on the Iowa Test of Basic Skills, were compared (Ysseldyke, Algozzine, Shinn, & McGuire, 1982). Considerable similarities were found between the groups. An average of 96% of the scores on forty-nine subtests from various tests were within a common range, and the performances of the two groups on many subtests were identical. The authors concluded that as many as 40% of the children may have been misclassified. In a follow-up study, Ysseldyke, Algozzine, and Epps (1983) found: (a) 85% of 248 students identified as normal could be classified as learning disabled according to 17 operational definitions; (b) of the 49 low achieving students who had not been identified as learning disabled by the school, 92% could be classified as learning disabled; and (c) as many as 4% of those identified as learning disabled did not meet any of the operational definitions used in the study.

A large proportion of the misclassified children are in reality slow learners and not learning disabled. Due to the lack of special programs for them, slow learners are often placed in resource and self-contained programs for learning disabled children. Kirk and Elkins (1975) found that approximately 35% of the children enrolled in Child Service Demonstration Centers for learning disabilities had IQs below

ninety.

Further, there appear to be differences between children receiving only resource assistance versus those children placed in a full-time special education program. Using the WISC-R and achievement tests, Olson and Midgett (1984) compared learning disabled children placed in self-contained classrooms with children receiving only resource service. The resource children had an average IQ of 96, while self-contained children had an average IQ of 90. This is a significant difference, and the only identifiable difference between the two groups.

The intent of this study is to determine the stability of the WISC-R after a three year period with a group of learning disabled students. Previous investigations have shown the WISC-R is stable for the general population. Consequently, for the purposes of the study, it is assumed that the normal population (as tested by Wechsler) would not show a three-year mean change due to the concept of a stable intelligence quotient. Therefore, the primary hypothesis of this study is that children identified as learning disabled and receiving special education services, will demonstrate similar stable WISC-R scores over a three year time period. Stability will be defined as the consistency of the two WISC-R Scale scores obtained by the same individual.

A secondary purpose of this study is to examine the relationship between WISC-R stability and whether the learning

disabled students receive resource services or are in a self-contained program. Additionally, Metropolitan Achievement Test (MAT) scores will be analyzed to determine if a relationship exists between special education placement and MAT changes. Any differences occurring between the two groups can only be speculated on at this time due to the lack of research in these areas, and placement policies of the school system.

METHOD

Subjects

A record review was done on all seventh, eighth, and ninth grade students classified as learning disabled in a large junior high school (enrollment during the two years of the study-1500) in a suburban school system in a New England state. Of those classified, only students who were receiving or who had received in the past special education services, and who also had more than one WISC-R in their reoord were eligible. Dual diagnoses of behaviorally disordered/learning disabled and physically handicapped/learning disabled were excluded due to an inability to determine primary handicapping condition. The criteria for inclusion in this study were met by 135 subjects of whom 73% were male and 27% female.

These subjects were then screened to separate out only those who demonstrated a learning disability based on an achivement/ability discrepancy formula. To distinguish this new group of subjects from the initial 135 subjects classified as learning disabled, the group meeting the ability/achievement discrepancy will be called "learning disabled", while the original group will be referred to as "all classified". Sixty subjects were identified as learning disabled, and then were further divided into two groups depending on the type of special education service they received. Thirty subjects received only resource services,

and thirty subjects had been placed in the more intensive self-contained program.

Table 1 presents the means and standard deviations for subject's age, years between testing, years in special education, and T-score point discrepancy. A T-score point discrepancy is defined as the difference between a subject's ability (measured by the WISC-R) and achievement (measured by the MAT) where both test scores are converted into T-score points. To investigate whether the two learning disabled groups (resource and self-contained) were different on these four variables, four t-tests were performed.

Table 1

Means and Standard Deviations for Age, T-score Point Discrepancy, Years Between Testing, and Years in Special Education for Resource and Self-Contained Groups.

	Age		T-score Points		Years Btwn. Testing		Years In Spec. Ed.	
Group	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Resource (N=30)	14.03	0.96	11.60	7.90	2.92	0.79	4.07	1.70
Self-Cont. (N=30)	14.01	1.20	11.78	9.80	3.53	1.08	5.63	1.45
Total L.D. (N=60)	14.02	1.08	11.69	8.87	3.23	0.99	4.85	1.75

The mean age of subjects was 14.02 years with no significant difference between groups. Also, no difference occurred between groups on average T-score discrepancy. There

was a significant difference in years between testing, $t(58) = 2.5$, $p<.01$. The self-contained group's retesting interval (M=3.53 years) was longer than the resource group's retesting interval (M=2.92 years). Additionally, there was a significant difference in the time the two groups spent in special education, $t(58) = 3.8$, $p<.01$. The self-contained group spent more years (M=5.63 years) than the resource group (M=4.07 years).

Instruments

The instruments used in this study were the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the Metropolitan Achievement Test (MAT). The WISC-R is a test of general intelligence consisting of twelve subtests, two of which are used only as supplementary tests. Six of the subtests form the verbal scale (Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span), and the other six form the performance scale (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes). The globality, utility, and reliability of the WISC-R has already been discussed.

The MAT is a nationally standardized achievement test consisting of five levels: Primary I, Primary II, Elementary, Intermediate, and Advanced. The reliability estimates of these five levels range from .88 to .96 (Kuder-Richardson), and the procedures for establishing content validity were

thorough and sophistioated (Anastasi, 1982). Similar to the WISC-R, the MAT is an academic instrument that was not specifically standardized on learning disabled children, and yet is frequently used in psychoeducational evaluations. To test its stability over time, Smith and Rogers (1978) included the MAT in a study examining the reliabilities of several tests of intelligence, academio skill, and affective assessment. A group of learning disabled children with a mean age of 9.9 years was retested after a six month interval on alternate forms of the MAT. The sample size ranged from 58 to 76 subjects. Coefficients of temporal stability ranged from .55 to .77 on the Primary I and from .56 to .73 on the Primary II battery. The authors felt that the values seemed fairly robust due to the long test-retest interval, and the use of alternate test forms. They concluded that the MAT exhibited respectable psychometric characteristics when used with learning disabled children. Zingale, Smith, and Dokecki (1980) also investigated the temporal stability of the MAT. Eighty-two children, ranging in age from 6.2 years to 13.2 years, identified by their school system as having learning problems, and receiving special instruction in a resource room, were retested after a one-month interval. The same form of the Primary 1, Primary II, and Elementary level was administered on both occasions. The level administered depended on the child's current level of academic achievement as estimated by the resource room teacher. Coefficients of

temporal stabilty in Total Reading ranged from .82 (Primary I) to .97 (Elementary). Total Math coefficients were somewhat less reliable ranging from .79 (Elementary) to .90 (Primary I). The authors concluded that the MAT is reliable when administered to learning disabled children (especially the Total Reading test), and that its continued use with such populations is justified. Both tests were individually administered by certified school psychologists (WISC-R) and trained special education teachers (MAT).

Procedure

After the subjects' special education records were originally reviewed (resulting in the sample of 135 subjects who met the initial criteria), further information from the records was gathered. Demographic and educational information collected included: current grade placement, birthdate, initial and subsequent testing dates, WISC-R IQ scores, MAT scores, and type of special education services provided. Based on the test scores, the subjects were further screened to separate out only those students who demonstrated a learning disability in either math computation or reading comprehension; that is, a significant difference between cognitive ability and performance. These two areas were chosen since they are the areas routinely tested in this school system. The discrepancy was based on scores obtained on the WISC-R and MAT. The initial test scores obtained on

both tests were converted into T-scores to provide comparable scores. If the T-score for the WISC-R was eight points or more larger than the MAT score (in either math or reading), the subject was considered learning disabled. This procedure is based on research conducted by Hanna, Dyck, and Holen (1979).

There are several methods currently used to identify children as learning disabled besides an ability/achievement discrepancy. Examples of other deviations from the norm that have classified children as learning disabled include: Language delays, verbal-performance discrepancy, central processing deficit, brain damage, attentional problems, perceptual-motor impairment, and maturational lag (Bryan & Bryan, 1978). The ability/achievement discrepancy model was used in this study because it is stated in the federal regulations, and routinely used in the school system.

Of the 135 subjects who were checked for this ability/achievement discrepancy, only 60 met the criterion for identification as learning disabled with an average reading T-score discrepancy of 11.03 points and an average math T-score discrepancy of 12.52 points. These 60 subjects were then separated into two groups depending on the type of special education service they received, self-contained or resource. The self-contained program provided the students with basic academic instruction in a special education setting in one or more academic areas. Subjects who had received any

self-contained services at all in their educational history were classified as self-contained for the purposes of this study regardless of present placement. The resource program is a part-time supplemental service designed to reinforce classroom curriculum, and to provide activities designed to develop specific learning skills pertinent to the disability. Each group contained thirty subjects.

RESULTS

WISC-R

The first statistical analysis tested the hypothesis that the learning disabled subjects would show stable WISC-R scores. Means and standard deviations for the Full, Verbal, and Performance Scale scores of the WISC-R for the original test and three-year retest, differences in means, and correlations for all classified and the learning disabled sub-groups are presented in Table 2.

Twelve one-way repeated measures ANOVA, (Winer, 1971) indicated no significant Full, Verbal, or Performance Scale IQ mean differences for any of the learning disabled groups. However, for all classified (N=135), statistically significant IQ mean increases were found for the Full Scale score ($F(1,134)=12.18$, $p<.001$), and the Performance Scale score ($F(1,134)=21.65$, $p<.001$). All ANOVAs proved to have non-significant Fmax values. Also in Table 2, the Pearson r Correlation Coefficient for each group's test and three-year retest score is presented. All correlations were found to be significant at the .001 level.

Table 2

Means, Standard Deviations, and Differences in Means of WISC-R Test and Retest Full, Verbal, and Performance Scale Scores; and Correlations for All Classified and the Learning Disabled Sub-groups.

	Test		3 year Retest			
Group	Mean	S.D.	Mean	S.D.	Difference In Means	r
All (N=135)						
Full	96.34	12.58	98.87	12.42	+2.53*	.78*
Verbal	97.28	13.42	98.29	12.72	+1.01	.73*
Performance	96.40	13.32	100.16	13.74	+3.76*	.76*
Total L.D. (N=60)						
Full	101.88	12.83	102.95	13.68	+1.07	.82*
Verbal	101.08	12.62	101.87	13.18	+0.79	.73*
Performance	102.63	18.61	104.03	19.59	+1.40	.89*
Self-Contained (N=30)						
Full	95.87	11.55	96.17	12.93	+0.30	.83*
Verbal	94.90	11.26	94.67	10.76	-0.23	.69*
Performance	98.00	14.05	98.87	16.68	+0.87	.82*
Resource (N=30)						
Full	107.90	10.87	109.73	10.45	+1.83	.68*
Verbal	107.27	10.49	109.07	11.06	+1.80	.58*
Performance	107.27	12.63	109.20	11.59	+1.93	.76*

* = $p < .001$

The stability of subtest scores for the learning disabled groups was also examined. Table 3 presents the mean differences for eleven subtest scores between the first and second WISC-R, and correlations between the two tests for all classified and the learning disabled sub-groups. Information was available for only 49 of the learning disabled subjects. Thus, for this analysis, the number of subjects in the self-contained group was 22, and in the resource group 27. The number of subjects in all classified decreased to 116.

Table 3

Mean Differences for Eleven Subtest Scores Between the First and Second WISC-R. and Correlations for All Classified and Learning Disabled Sub-groups.

	All (116)		Total L.D. (60)		Self-Cont. (30)		Resource (30)	
Subtest	Dif.	r	Dif.	r	Dif.	r	Dif.	r
Information	+.19	.62	-.06	.59	-.57	.53	+.33	.58
Similarites	+.50	.55	+.37	.41	+.05	.29#	+.63	.39
Arithmetic	-.08	.41	+.39	.38	+.09	.57	+.63	.19#
Vocabulary	-.43*	.64	-.96**	.67	-.91*	.60	-1.00*	.69
Comprehension	+.49	.49	+.12	.45	-.59	.32#	+.70	.41
Digit Span	-.08	.55	-.55	.74	-.86	.79	-.20	.66
Pict. Comp.	+1.01	.57	+1.18	.57	+1.18	.45	+1.18	.62
Pict. Arran.	+1.07	.47	+.13	.37	+.16	.05#	+.12	.53
Block Design	+.18	.59	+.19	.63	+.23	.74	+.15	.46
Object Assem.	+.26	.61	+.14	.60	+.82	.67	-.42	.58
Coding	+.09	.59	-.39	.60	-1.27	.68	+.33	.58

Dif. = Mean difference betweeen tests
# = Non-significant correlation
* = $p < .05$
** = $p < .01$

Eleven one-way repeated measures ANOVAs found changes in only one subtest which occurred on the Vocabulary test. All four groups' mean vocabulary score decreased significantly on retesting; "All Classified" ($F(1,115)=4.45$, $p < .05$); "Total

Learning Disabled" ($F(1,48)=12.03$, $p<.01$); "Learning Disabled Self-Contained" ($F(1,21)=4.45$, $p<.05$); and "Learning Disabled Resource" ($F(1,26)=7.47$, $p<.05$). Except where noted, all correlations were found to be significantly greater than chance alone.

To determine whether the type of special education the learning disabled subjects received related to their WISC-R retesting scores, three 2X2 Analyses of Variance with repeated measures on one factor were utilized. Summaries for the Full, Verbal, and Performance Scale IQ scores with special education service (Resource and Self-Contained) as one variable, and WISC-R score (first and second administration) as the second variable are presented in Table 4.

Table 4

2X2 Analysis of Variance Between Special Education Groups for each of the WISC-R Scales, and Two Test Administrations.

Source	df	Mean Square	F
Full Scale			
Between Subjects			
Group (G)	1	4915.20	20.44**
Error	58	240.41	
Within Subjects			
Test-Retest (T)	1	34.13	1.05
GXT	1	17.63	0.54
Error	58	32.59	
Verbal Scale			
Between Subjects			
Group (G)	1	5373.41	26.72**
Error	58	201.10	
Within Subjects			
Test-Retest (T)	1	18.41	0.41
GXT	1	31.01	0.70
Error	58	44.50	
Performance Scale			
Between Subjects			
Group (G)	1	2881.20	8.09*
Error	58	356.04	
Within Subjects			
Test-Retest (T)	1	58.80	1.40
GXT	1	8.53	0.20
Error	58	41.86	

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

For each of the summaries, the only significant effect was between services. Those learning disabled subjects receiving only resource services had significantly higher Full, Verbal, and Performance Scale IQ scores than those subjects receiving more intensive services in a self-contained placement.

MAT

To test whether composite MAT scores increased significantly upon retesting for learning disabled subjects as predicted, a repeated measures ANOVA was utilized. A total of 53 MAT pairs (one each in math and reading) was available for the anaysis. Means and standard deviations for the original MAT test and three-year retest, and differences in means for the learning disabled groups are presented in Table 5. Also in Table 5, the Pearson r Correlation Coefficient for each group's test and retest mean are shown. Only the total learning disabled and resource groups demonstrated significant correlations.

Table 5

Means and Standard Deviations for the Original MAT Test and Three-year Retest, and Correlations for the Learning Disabled Groups.

	Test		3 Year Retest		
Group	Mean	S.D.	Mean	S.D	r
Total L.D. (N=106)	22.09	21.58	29.19	21.40	.42*
Self-Cont. (N=53)	17.36	21.69	19.89	19.53	.24
Resource (N=53)	26.83	20.60	38.49	19.36	.51*

* = $p < .001$

To determine whether the subjects made any academic progress during the three years based on their special education placement, a 2X2 Analysis of Variance with repeated measures on one factor was performed on the MAT reading and math scores. The ANOVA summary with special education service (Resource and Self-contained) as one variable, and MAT score (first and second administration) as the second variable is presented in Table 6.

Table 6

Analysis of Variance Between Special Education Groups for Two MAT Administrations.

Source	df	Mean Square	F
Between Subjects			
Group (G)	1	104444.08	18.43***
Error	104	566.69	
Within Subjects			
Test-Retest (T)	1	2667.49	10.31**
GXT	1	1104.98	4.27*
Error	104	258.79	

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

There was a significant interaction between the two groups and two test scores ($F(1,104)=4.27$, $p < .05$). A simple effects analysis indicated that the resource group's test-retest change was significantly greater than the self-contained group's ($F(1,104)=13.92$, $p < .001$). Other results included a significantly higher MAT mean score for the resource group than for the self-contained group ($F(1,104)=18.43$, $p < .001$). Further, the three-year retesting resulted in significantly higher MAT mean scores than the initial testing ($F(1,104)=10.31$, $p < .01$).

DISCUSSION

This section is separated into two areas: Test stability and group differences. The first part discusses the primary hypothesis of the study; that is, children identified as learning disabled will demonstrate stable WISC-R scores over three years. The second part attempts to explain the differences found between the resource and self-contained groups. This part will be more speculative because of the lack of research in this area. Additionally, the classification process of this school system is probably different from other school systems because a consistent and effective method for identifying children as learning disabled has not been developed.

Test Stability

The total special education population originally classified as learning disabled (all classified) showed significant increases for the WISC-R Full and Performance Scales with retesting. These results are contrary to the assumed stability of the test for the normal population, and probably due to the extreme heterogeneity of this population. More importantly, those subjects who met the criterion for learning disabled demonstrated non-significant test-retest differences on all three WISC-R scales. This study, therefore, found that subjects properly diagnosed as learning

disabled and receiving special education services have statistically stable WISC-R test-retest scores, similar to the population used to norm the WISC-R.

Generally, the test-retest correlation coefficients found in this study were lower than expected. Wechsler found reliability coefficients ranging from .90 to .95, and Anastasi (1982) cites desirable reliability coefficients in the .80s and .90s. The lower correlations appear to be due to the inconsistent achievement of learning disabled children both individually and in groups, and the longer time between testing than most test-retest reliability studies. Bloom (1964) reviewed the literature on longitudinal studies of general intelligence and found that stability is greater for shorter time periods than for longer time periods.

The single subtest change for all 135 subjects occurred on the Vocabulary subtest where there was a significant decrease. This appears to be due to the significant decrease on Vocabulary retesting for those subjects found to have a learning disability, regardless of special education placement. Vocabulary has been found to be the most reliable subtest (r=.86) in the scale (Sattler, 1982). This was supported in the study, especially for all 135 subjects, as it had the highest reliability (r=.64) for the subtests. It was also consistently one of the most reliable for the learning disabled groups. The reliability estimates of all the subtests were low, but expected due to the lower reliabilities

of the three WISC-R scales.

Unlike the stable WISC-R scores, the MAT scores increased significantly for the total group of learning disabled children. This increase appears to be due primarily to the almost 12 points gain found for the students receiving only resource services. Although the self-contained group increased by 2 MAT points with retesting, this increase was not significant. It was similar to the stable MAT scores found with learning disabled children in previous studies (Smith & Rogers, 1978; and Zingale, Smith, & Dokecki, 1980).

As was hypothesized, the WISC-R is as reliable for learning disabled children as it is for normal children. This study found stable WISC-R scores can be assumed if proper classification of students as learning disabled is made upon entrance into special education. Thus, if a learning disabled student has received two WISC-R assessments which fall within the standard error of the measurement of each other, the probability that a third WISC-R administration will yield significantly different information is very low. Consequently, systematic WISC-R reevaluation is probably not valuable with learning disabled children.

This conclusion corresponds with that of Galvin and Elliott (1985) who argued that the relative low incidence of changes in diagnosis and placement challenges the necessity for automatically reevaluating handicapped children with a comprehensive evaluation every three years. Therefore, stable

WISC-R scores over time, and infrequent diagnostic and placement changes, indicate that repeated WISC-R testing with learning disabled children is not necessary.

Triennial evaluations certainly are needed to ensure that children are appropriately placed, and receive the educational services that will most benefit them. Repeated cognitive testing using the WISC-R, however, is unnecessary. Other areas of the child's functioning (e.g., educational, social, emotional, behavioral) can elicit more information for future educational services and career preparation than another intelligence test.

Alternatives to routinely administering the full WISC-R every three years have been proposed. Elliott et al. (1985) recommended using the short-form of the WISC-R (consisting of Vocabulary, Arithmetic, Block Design, and Picture Completion). Galvin and Elliott (1985) suggest two alternatives: (a) rereferal and (b) annual review of the IEP. The rereferal alternative would use the existing referal process for handicapped students. Teachers, parents, and others could refer handicapped children whenever the child's progress or behavior indicated such a need rather than depend on a fixed, periodic schedule. The annual review would have the psychologists participate more actively in the IEP annual review. The psychologist can use consultation and behavioral intervention skills to effect meaningful changes in the student's education.

Group Differences

Many differences were found between the two special education groups. Explanations for these differences are presented, but the possible interpretations of them are myriad and highly speculative at this time.

Of the 135 subjects originally labelled learning disabled in this school system, only 44% (60 Subjects) met the discrepancy model criterion for having a learning disability in either math computation or reading comprehension. The remaining 56% had other learning problems, or perhaps were learning disabled in other areas. The latter explanation, however, is not probable. Other investigations in this school system have found that the number of students diagnosed as learning disabled for reasons other than reading comprehension and math computation is minimal. Therefore, it is estimated that very few of the subjects not meeting the discrepancy criteria are in fact truly learning disabled. These results are similar to the 42.6% learning disabled found by Shepard, Smith, and Vojir (1983). As this study and others found, those who are labelled learning disabled and receive services for their disability oftentimes do not qualify for special education remediation based on the federal definition.

When comparing the two groups' respective compositions, it is not surprising that the subjects' ages were not

significantly different. Approximate numbers of students classified as learning disabled were identified each year, and the learning disabled split between self-contained and resource services was consistent across grades. The mean age of 14.02 years for all learning disabled subjects is also consistent with that of a normal junior high population.

For all three IQ Scales, the ANOVA indicted a significant mean IQ score difference between the special education services. Consistently, the resource group showed higher mean IQ scores, similar to the results of Olson and Midgett (1984). Further, the difference between the two groups on the severity of their discrepancy was not significant (11.60 points resource versus 11.78 points self-contained). Since the self-contained group did not demonstrate a greater mean T-score point discrepancy than the resource group, the discrepancy size between ability and performance does not appear to be the major factor in deciding which special education service the student receives. Rather, the findings of this study suggests that placement in special education programs appears to be determined more by the student's IQ than the severity of his/her learning disability.

Associated with this is the significant difference between the groups on number of years spent in special education and number of years between testing. Having spent significantly fewer years in special education (4.07 versus 5.63), it can be assumed that those students receiving only

resource services entered special education later in their schooling, requiring more time for their disability to significantly hamper their academic achievement. Also, they were more likely to be exited from special education upon reevaluation because they were considered less handicapped. One or two years behind grade level at the secondary level is not as significant as it is at the primary level.

The significant difference in years between testing (2.92 resource versus 3.53 self-contained) is related to the total years each group spent in special education. The multidisciplinary team is aware that those students who only receive resource services, and are assumed to have milder learning problems, may need to have their programs changed after three years. This is especially important as the child moves to higher and more difficult grades. Alternatively, the chance of those students previously placed in a full-time special education program making dramatic gains or losses that would appreciably change their placement is thought to be minimal. Therefore, when it is time for three-year reevaluations (usually left to the end of the academic year), the evaluations of those in a self-contained program tend to be delayed unless there is a specific reason to evaluate them sooner. Although this is not in keeping with the federal regulations and unfair to those students whose assessments are being delayed, it is probably due to the large number of reevaluations needed to be completed each year.

Finding that the two groups differed on MAT percentile score changes is not surprising. When they were originally tested, the MAT percentile scores were already higher for the resource group, indicating their relatively stronger initial academic achievement. Further, the resource group has been shown to be more intelligent than the self-contained group, and are assumed to have milder problems. Consequently, they are able to remain in the regular class and receive only supportive assistance. Thus, they continue to learn information and skills commensurate with their age. The fact that their achievement has increased significantly suggests that resource services alone have proven successful for them as a group. Alternatively, a self-contained placement does not appear to increase one's academic skills, and maintains these skills at a lower level compared to the normal population.

This study found distinct and important differences between the two special education groups. It appears that determining where a truly learning disabled child is placed may be more dependent on the child's IQ than on other variables thought to be more significant (e.g., severity of the learning disability), but again, this interpretation needs further investigation.

Summary

In summary, this study supported the original hypothesis that learning disabled children would demonstrate stable WISC-R scores over a three-year retesting span. This finding, therefore, provides support for not automatically retesting children every three years with the WISC-R. Alternative assessment techniques should be utilized to learn more useful information about the child's present functioning. This study also found several differences between those children placed in a self-contained program versus those who only receive resource services. The interpretation of these findings is speculative at this time, however, and more research in needed in this area. For example, other school systems' placement policies should be investigated to ascertain whether similar differences between children placed in the two services are found.

References

Anastasi, A. (1982). Psychological Testing. New York: Macmillan.

Bloom, B. (1964). Stability and change in human characteristics. New York: Wiley.

Bryan, T., & Bryan, J. (1978). Understanding learning disabilities. Sherman Oaks, Ca: Alfred.

Covin, T. (1977). Stability of the WISC-R for 9 year-olds with learning difficulties. Psychological Reports, 40, 1297- 1298.

Education for all Handicapped Children Act, PL 94-142 (United States Office of Education, November 29, 1975).

Elliott, S., Piersel, W., Witt, J., Argulewioz, E., Gutkin, T., & Galvin, G. (1985). Three-year stability of WISC-R IQs for handicapped children from three racial/ethnic groups. Journal of Psychoeducational Assessment, 3, 233-244.

Galvin, G., & Elliott, S. (1985). Psychological reevaluations of handicapped children: A survey of practioners and policymakers. Professional Psychology: Research and Practice, 16, 64-75.

Hanna, G., Dyck, N., & Holen, M. (1979). Objective analysis of achievement aptitude discrepancies in LD classification. Learning Disabilities Quarterly, 2, 32-38.

Kaufman, A. (1979). Intelligence testing with the WISC-R. New York: Wiley.

Kirk, S., & Elkins, J. (1975). Characteristics of children enrolled in Child Service Demonstration Centers. Journal of Learning Disabilities, 8, 630-637.

Olson, J., & Midgett, J. (1984). Alternative placements: Does a difference exist in the LD populations? Journal of Learning Disabilties, 2, 101-103.

Shepard, L., Smith, M., & Vojir, C. (1983). Characteristics of pupils identified as learning disabled. American Educational Research Journal, 20. 309-331.

Smith, M. (1978). Stability of the WISC-R subtest profiles for learning disabled children. Psychology in the Schools, 15, 4-7.

Smith, M., & Rogers, C. (1978). Reliability of the standardized assessment intruments when used with learning disabled children. Learning Disabilities Quarterly, 1, 23-31.

Tuma, J., & Appelbaum, A. (1980). Reliability and practice effects of WISC-R IQ estimates in a normal population. Educational and Psychological Measurement, 40, 671-678.

United States Office of Education (1977). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. Federal Register, 42, 65082-65085.

Vance, H., Blixt, S., & Ellis, R. (1981). Stability of the WISC-R for a sample of exceptional children. Journal of Clinical Psychology, 37, 397-399.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children-Revised. New York: Psychological Corporation.

Winer, B. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Ysseldyke, J., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practice in classifying students as handicapped. Exceptional Children, 50, 160-166.

Ysseldyke, J., Algozzine, B., Shinn, M., & McGuire. (1982). Similarities and differences between low achievers and students classified learning disabled. The Journal of Special Education, 16, 73-85.

Zingale, S., Smith, M., & Dokecki, P. (1980). Temporal stability of the Metropolitan Achievement Test when used with learning disabled children. Learning Disabilities Quarterly, 3, 84-86.