DOCUMENT RESUME

ED 211 608 TM 820 066

AUTHOR	Epps, Susan; And Others
TITLE	Inter-Judge Agreement in Classifying Students as Learning Disabled.
INSTITUTION	Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY	Office of Special Education (ED), Washington, D.C.
REPORT NO	IRLD-RR-51
PUB DATE	Feb 81
CONTRACT	300-80-0622
NOTE	36p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Educational Diagnosis; *Handicap Identification; Intermediate Grades; *Judges; *Labeling (of Persons); *Learning Disabilities; Learning Problems; Special Education; *Student Evaluation
IDENTIFIERS	*Inter Rater Reliability

ABSTRACT

Eighteen judges with backgrounds in assessment, decision making, and learning disabilities were asked to use an array of information to differentiate learning disabled (LD) and non-learning disabled students. Each judge was provided with forms containing information on 42 test or subtest scores of 50 school-identified LD students and 49 non-LD students. Judges were extremely inaccurate in their classifications and in little agreement with each other. Also, it appeared that different judges emphasized different factors when making their decisions. The results suggest that, give current definitions of the condition called "learning disabilities," there is little basis for the hope that school personnel are going to be able accurately to identify such students. (Author/CE)

University of Minnesota

Research Report No. 51

INTER-JUDGE AGREEMENT IN CLASSIFYING STUDENTS AS LEARNING DISABLED

Susan Epps, Matthew McGue, and James E. Ysseldyke

IRLD

Institute for Research on Learning Disabilities

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mi.

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

Research Report No. 51

INTER-JUDGE AGREEMENT IN CLASSIFYING STUDENTS AS LEARNING DISABLED

Susan Epps, Matthew McGue, and James E. Ysseldyke

Institute for Research on Learning Disabilities

University of Minnesota

February, 1981

Abstract

Eighteen judges with backgrounds in assessment, decision making, and learning disabilities were asked to use an array of information to differentiate learning disabled and non-learning disabled students. Each judge was provided with forms containing information on 42 test or subtest scores of 50 school-identified LD students and 49 non-LD students. Judges were extremely inaccurate in their classifications and in little agreement with each other. Also, it appeared that different judges emphasized different factors when making their decisions. The results suggest that, given current definitions of the condition called "learning disabilities," there is little basis for the hope that school personnel are going to be able accurately to identify such students.

Inter-Judge Agreement in Classifying Students as Learning Disabled

Traditionally, research in the area of learning disabilities has compared learning disabled students to students identified as "educable mentally retarded," "emotionally disturbed," or "normal" (Ackerman, Peters, & Dykman, 1971; Bussell, Huls, & Long, 1975; Gajar, 1979; Wagonseller, 1973; Webster & Schenck, 1978). While numerous attempts have been made to identify the salient characteristics that distinguish LD students from other students, a major difficulty with such studies is that they compare learning disabled students to populations with obviously different characteristics. In comparison to the mentally retarded, we expect and observe intellectual differences; compared to students classified as emotionally disturbed, we expect and find differences in the number of behavior problems; in comparison to "normal" students we expect and identify differences in the level of academic achievement. We really learn very little from such investigations.

Diagnostic personnel usually have little difficulty distinguishing between discernably different kinds of students ("normal," MR, LD), relative to the difficulties they experience in trying to differentiate LD students from those who are simply slow learners. Numerous students fail to acquire academic skills and fail to meet the objectives that the school sets for them. A critical diagnostic problem is seen in efforts to decide, from among the population of students who are not "making it" in school, who should be declared LD and provided with special educational interventions. Research is nearly nonexistent on this critical diagnostic problem.

Although school personnel typically believe that few low-achieving students exist who have not been labeled as LD, Ysseldyke, Algozzine, Shinn, and McGue (in press) found a relatively large number of these students at the fourth-grade level. The students scored in the lowest quartile on a group achievement test, but had not been identified as LD by their school districts.

Given that the two groups of students (low achievers and LD) do exist, it becomes important to investigate the basis for distinguishing them. Ysseldyke et al. (in press) examined the extent to which measures of cognitive ability, academic achievement, perceptual-motor skills, self-concept, and behavior problems discriminated between low achievers and school-identified LD children. Their results indicated considerable similarity between the two groups; an average of 96% of the scores were within a common range and the performance of the LD and low-achieving students on many of the subtests was identical. Based on psychometric measures only, Ysseldyke et al. (in press) raised serious concerns regarding the differential classification of low-achieving students as either learning disabled or non-learning disabled.

Some school personnel have argued that individual psychometric measures do not provide the necessary information for distinguishing between LD and low-achieving students. Rather, they maintain that decisions actually are made on the basis of the pattern of scores that students obtain on different psychometric instruments. It thus becomes important to ascertain the extent to which school decision makers can distinguish between the two groups when given patterns of scores on multiple measures.

The use of a number of scores to identify learning disabled students is, in fact, mandated by the U.S. Office of Education. The August 23, 1977 Federal Register specifically states that "no single procedure is used as the sole criterion for determining an appropriate educational program for a child" (p. 42497). To determine the existence of a specific learning disability, "a severe discrepancy between achievement and intellectual ability" (Federal Register, 1976, p. 65083) must be demonstrated. Both of these federal guidelines necessitate the use of more than one test in an evaluation of a student, a procedure which, in turn, requires an examination of a student's pattern of scores.

Although an ability-achievement discrepancy is a typical requirement for classification as learning disabled, by no means is there consensus that it is a useful concept. Salvia and Ysseldyke (1981) caution that before a student's difference scores can be interpreted, it must be established that the differences are reliable; differences in standardization samples for the two tests also should be considered. In practice, these cautions frequently are ignored. In addition, a number of studies (Algozzine, Forgnone, Mercer, & Trifiletti, 1979; Algozzine, Ysseldyke, & Shinn, 1980; Ysseldyke & Algozzine, 1979) argue that learning disabilities is largely a category of underachievement, thus suggesting that the degree of students' achievement deficits also should be considered when determining eligibility for LD placement.

It is likely that the group of students identified as LD by an ability-achievement discrepancy will be different from the group identified by an achievement deficit; in turn, both of these groups may be different from the group identified as LD by school decision-making teams. The present study was designed to investigate the extent to which

professionals could distinguish between the learning disabled and non-learning disabled students identified by each of three definitions (actual school classification, ability-achievement discrepancy, achievement deficit) on the basis of an array of information similar to that typically presented in placement team meetings.

Method

Subjects

The sample consisted of 18 judges, 14 of whom were doctoral students in school psychology or special education. Of the other four judges, two had master's degrees in education, one had a doctorate in clinical psychology, and one had a doctorate in special education. Eleven of the judges had regular or special education teaching experience and seven were certified school psychologists. All 18 judges had previous experience in psychoeducational assessment and/or placement team decision making, plus experience in research on learning disabilities.

Materials

Forms containing the chronological age of 99 fourth-graders and information on their actual test scores in five domains were prepared. The domains and the tests used were: (a) aptitude (WISC-R, Woodcock-Johnson Tests of Cognitive Ability), (b) academic achievement (Woodcock-Johnson Tests of Achievement, PIAT, Stanford Achievement Test - math calculation and math concepts), (c) perceptual-motor (Bender, VMI), (d) self-concept (Piers-Harris), and (e) behavior problems (Peterson-Quay). Across these five domains, test scores were provided on 42 variables.

Fifty of the 99 students were school-identified learning disabled students and 49 were low achievers from metropolitan Minneapolis and

St. Paul schools; this classification was not indicated on the judges' test score forms. The exact criteria for identification of students as LD by the schools were unknown. The low-achieving group had not been identified as LD by their school districts, but scored at or below the 25th percentile on the Iowa Tests of Basic Skills which had been administered by the schools during the fall of the school year.

In addition to the schools' classification as a criterion for considering a student LD, the federal definition of learning disabilities, taken from the December 29, 1977 Federal Register, was used to classify students as LD or non-LD. That definition lists seven areas in which a child may be determined to have a specific learning disability: oral expression, basic reading skill, reading comprehension, mathematics calculation, mathematics reasoning, listening comprehension, and written expression. However, the federal definition does not specify the amount of discrepancy between ability and achievement that is required. In the present study, the federal definition was operationalized in two ways using subjects' observed aptitude-achievement test scores in the areas of intelligence and achievement. The following measures were used to ascertain discrepancy in five of the seven areas: (a) written expression (WISC-R Full - W-J Written Language Achievement), (b) basic reading skill (WISC-R Full - W-J Reading Achievement), (c) reading comprehension (WISC-R - PIAT Reading Comprehension), (d) mathematics calculation (WISC-R Full - Stanford Mathematics Calculation), and (e) mathematics reasoning (WISC-R Full - Stanford Mathematics Concepts). In one condition (1.0 SD), the student was considered LD by definition if

there was between a 1.0 and 1.5 standard deviation difference on at least one of the five aptitude-achievement discrepancy scores (difference scores of 15 to 22). In the other condition (1.5 SD), the student was considered LD if there was at least a 1.5 standard deviation difference on at least one of the five aptitude-achievement discrepancy scores (difference scores of 23 or more). Table 1 indicates the number of school-identified LD students and low achievers who were classified as LD and non-LD based on the school's definition and the federal definitions.

Insert Table 1 about here

Along with the schools' classification and the two operationalizations of the federal definition as criteria for considering a student LD, the degree of students' achievement deficits was used to classify students as LD or non-LD. The achievement deficits were operationalized in two ways using subjects' observed test scores in written expression, basic reading skill, reading comprehension, mathematics calculation, and mathematics reasoning. In one condition (1.0 SD), the student was considered LD by definition if the score earned on at least one of the criterion measures was between 1.0 and 1.5 standard deviations below the mean (standard scores of 78 to 85). In the other condition (1.5 SD), the student was considered LD if the score earned was at least 1.5 standard deviations below the mean (standard scores of 77 or below). Table 2 indicates the number of school-identified LD students and low achievers who were classified as LD and non-LD based on the schools' definition and the low achievement definitions.

Insert Table 2 about here

Along with the "test scores" form, the packet of materials given to judges included a "tests administered" form that specified which tests were used in each of the five areas. A third sheet provided normative data for those measures that did not have a mean of 100 and standard deviation of 15. These included the Bender Visual-Motor Gestalt Test, the Developmental Test of Visual-Motor Integration, and the Peterson-Quay Behavior Problem Checklist. Also provided were the full group mean and standard deviation for the Piers-Harris Children's Self-Concept Scale, based upon the total sample of 99 students.

Procedure

Each judge was given a packet of materials and a response form, and then instructed to examine the test scores for the 99 students and to indicate which he/she believed were learning disabled and which were non-learning disabled.

Results

Judges' Accuracy

Judges' accuracy in identifying learning disabled students is reported in Tables 3 and 4. A comparison was made between judges' ratings and school classification, between judges' ratings and the classification from each of the two federal definitions, and between judges' ratings and the classifications from each of the two levels of achievement deficit. The accuracy percentages were computed separately for LD and non-LD students within each

of the five conditions and dependent-sample ts were calculated. Across all five conditions, judges were more accurate in identifying the non-LD than the LD students [school definition: t (17) = 3.89, $p < .001$; 1.0 SD federal definition: $t = 5.23$, $p < .001$; 1.5 SD federal definition: t (17) = 3.44, $p < .001$; 1.0 SD low achievement definition: t (17) = 4.08, $p < .001$; 1.5 SD low achievement definition: t (17) = 2.71, $p < .01$]. For the identification of LD students across all five criterion conditions, the percentages of judges' accuracy ranged from 2.9 to 72.0; for non-LD students, the percentages of accuracy ranged from 35.4 to 100.0.

Insert Tables 3 and 4 about here

Judges' Leniency

Also examined was the judges' leniency, a factor which reflects the percentage of children classified as LD. By definition, a lenient judge classifies more students as LD than as non-LD. A non-lenient judge classifies fewer students as LD than as non-LD. As can be seen in Table 5, only 4 of the 18 judges were lenient; that is, they identified 50% or more of the students as LD.

Insert Table 5 about here

Considering that lenient judges classify a greater number of students as LD, they will have a relatively higher accuracy in identifying the LD than in identifying the non-LD students. That is, the four lenient judges (4, 10, 14, and 18) are four of the most accurate judges in identifying

LD students (although none of the judges were in agr ement with the schools' classification for as many as three-fourths of the students) and are relatively inaccurate in identifying the non-LD students. Similarly, the non-lenient judges, those who classify more students as non-LD will be relatively more accurate in identifying non-LD students, but at the expense of their accuracy in identifying LD students. The negative relationship between accuracy in identifying LD and non-LD students is a direct result of the confounding between leniency and accuracy. This confounding is best illustrated by considering the totally lenient judge, that is, one who classifies all students as LD. Such a judge would correctly identify all LD students as LD, but would incorrectly identify all non-LD students as LD. Consequently, caution is necessary in interpreting esults of Tables 3 and 4; one must take into account the inherent confounding between leniency and accuracy of classification.

Inter-judge Agreement

A third factor investigated was inter-judge agreement, the extent to which judges agreed with each other in their ratings of the students. In deriving the index of interrater agreement, the point biserial correlation between each judge's ratings and the average rating of the remaining 17 judges was first computed. The resulting point biserial correlation was then divided by the maximum possible point biserial which could result given the proportion of children classified as learning disabled and non-LD for that given judge. The agreement index was defined as the ratio of actual point biserial to the maximum point biserial, and can be interpreted as the agreement of that judge's

ratings with all other judges' ratings relative to the maximum level of agreement possible given that judge's leniency.

As can be seen in Table 5, there is a wide range in inter-judge agreement, from .067 (judge 11) to .803 (judge 14). Yet in examining their accuracy in identifying LD and non-LD students, judge 11 and judge 14 were very similar in their percentages of agreement with the school definition and the 1.5 SD federal definition. In addition, the two most accurate judges in overall agreement with the schools (judges 14 and 17) were actually considerably different in the extent to which each of them agreed with the other 17 judges. Judge 14 was in high agreement with the other judges (.803), yet judge 17 was in low agreement (.319).

Frequency of Judges' LD Ratings

For each of the 99 students, a frequency count was made of the number of judges who rated the student as LD. For 94 of the students, there was at least one judge who rated the student as LD. There were no students whom everyone identified as LD, but there were five students whom everyone rated as non-LD. Three of these five students, however, were actually identified as LD by the schools.

Factors Influencing Decisions

The standard procedure adopted by several investigators to model the judges' decision-making process is the method of discriminant function (Shavelson, 1979). This method is preferred since it results in a parsimonious and statistically optimal description of the judge's decisions; we sought an alternative to it because of two shortcomings in the method.

First, the discriminant function is simply a statistical model of the decision-making process and does not necessarily identify those variables that were important in making the decision (Goldberg, 1970). Second, the low ratio of subjects (99) to variables (48) here would make the results of any discriminant function analysis highly unstable.

As an alternative to the discriminant function, we analyzed the mean differences between the judge-identified LD and non-LD students for each of the 18 judges. This was accomplished as follows. First, for each judge, the mean difference between the judge-identified LD and non-LD students was calculated on each of the 42 original test scores plus an additional six discrepancy scores (WISC-R Verbal - WISC-R Performance, W-J Reading Aptitude - W-J Reading Achievement, W-J Mathematics Aptitude - W-J Mathematics Achievement, W-J Written Language Aptitude - W-J Written Language Achievement, WISC-R Full - W-J Broad Cognitive, and WISC-R Full - PIAT Total). The 48 absolute mean differences were then rank ordered from largest to smallest. For example, for a given judge, the variable ranked first would be that variable for which the students judged learning disabled were most different from the students judged non-learning disabled, and thus a variable which apparently played a major role in the differentiation of LD from non-LD for that judge. Conversely, the variable ranked last would be that variable for which the students judged LD were most similar to the students judged non-LD, and thus of no apparent importance in differentiating the two groups.

Considering that all four lenient judges were more accurate in identifying LD students and that the 14 non-lenient judges were more accurate in identifying non-LD students, a judge's leniency is an apparent factor in the accuracy of the ratings. In order to determine the extent to which the differences between

the lenient and the non-lenient judges could be a function of their using different variables to arrive at a decision, we calculated the average rank for each of the 48 measures for the lenient and non-lenient judges. Table 6 lists the 10 most important variables, as judged by average rank, for the lenient and non-lenient judges.

Insert Table 6 about here

As can be seen in Table 6, the lenient judges were in very high agreement regarding the importance of an aptitude-achievement discrepancy (as measured by the WISC-R Full Scale IQ minus the PIAT Total Score); the average ranking was 2.00. The rankings for the non-lenient judges also indicated the importance of the aptitude-achievement discrepancy, although they were not in as high agreement; the variable had an average ranking of 13.79. Of the 10 ranked variables listed for each group of judges, 5 of them are common to both the lenient and non-lenient judges, namely, the aptitude-achievement discrepancy and 4 achievement measures. Overall, the rank ordering of variables suggests that both lenient and non-lenient judges as a group gave more consideration to measures of discrepancy and achievement than to measures of scholastic aptitude, perceptual-motor skills, or behavior problems.

At the group level, judges apparently were emphasizing aptitude-achievement discrepancy and achievement when differentiating LD from non-LD students. Yet the salient factors that individual judges used in making their decisions remain undelineated. To examine the variables that individual judges used to classify the students, mean difference scores were examined. For each judge, the means of each of the 42 test or subtest scores were calculated for students whom the judge classified as LD and for students whom the judge

classified as non-LD. Thus, the means indicated the average scores for particular variables for the judge-identified LD and non-LD groups. At the idiographic level of analysis, eight case studies are presented and include the two most lenient and two least lenient judges and the two most accurate and two least accurate judges (compared with the schools' ratings). The variables that they used to classify students, in order of magnitude, are presented in Table 7

Insert Table 7 about here

Most Lenient Judges

Case #1. The most lenient judge (judge 10) classified 60.6% of the students as LD. The mean difference scores suggested that this judge primarily used aptitude-achievement and verbal-performance discrepancies to differentiate the LD from the non-LD students.

Case #2. The second most lenient judge (judge 14) classified 52.5% of the students as LD. This judge also seemed to examine aptitude-achievement discrepancy, but placed more emphasis on achievement measures.

Least Lenient Judges

Case #3. The least lenient judge (judge 9) only classified 5.1% of the students as LD. Mean difference scores suggest that school-related aptitude scores were the primary consideration. Behavior ratings and an achievement measure apparently were examined as well.

Case #4. The second least lenient judge (judge 2) classified 11.1% of the students as LD. In a similar manner as the most lenient judges, this judge also emphasized aptitude-achievement discrepancy together with cognitive measures.

Most Accurate Judges

Case #5. One of the most accurate judges (judge 17) was in overall agreement with the schools' classification 70.7% of the time. This judge was considerably more accurate in identifying the non-LD (87.8%) than in identifying the LD (54.0%) students. As for several other judges, aptitude-achievement discrepancy apparently was examined. Cognitive measures and a school-related aptitude score also were emphasized with less consideration given to achievement measures.

Case #6. The other most accurate judge (judge 14), who was also one of the most lenient judges, was in overall agreement with the schools' classifications 70.7% of the time and was generally equally accurate in identifying both the LD and the non-LD students. This judge apparently emphasized aptitude-achievement discrepancy, as did the other highly accurate judge, but gave more consideration to achievement measures.

Least Accurate Judges

Case #7. One of the least accurate judges (judge 1) was in overall agreement with the schools' classifications 52.6% of the time. Aptitude-achievement discrepancy and discrepancy measures apparently were emphasized.

Case #8. The other least accurate judge (judge 3) was in overall agreement with the schools 52.5% of the time. Aptitude-achievement discrepancy was examined. Unlike the other least accurate judge, however, cognitive measures, rather than achievement measures, seemed to have been given strong consideration.

Discussion

Classification of students as learning disabled is clearly problematic. Ysseldyke et al. (in press) provided evidence that there are no meaningful psychometric differences between students identified by schools as learning

disabled and those who are low achievers and not identified as LD. They reported a 96% overlap between the two groups in performance on psychometric measures. Psychologists and educators who have viewed these results have argued repeatedly that while there is considerable overlap between the two groups, they can, given pupils' pattern of scores on numerous devices, differentiate between low achievers and those who are learning disabled. This argument, often called the "I know one when I see one" argument, was tested in this study.

Judges were provided with the scores earned on 42 tests or subtests by 50 school-identified LD students and 49 non-LD students. When asked to discriminate between the LD and non-LD students, judges were extremely inaccurate (whether compared to the school's classification, the classifications based upon the two federal definitions, or the two low achievement definitions) and in little agreement with each other. They also emphasized different factors. Almost all of the students were labeled LD by at least one judge, and only five were consistently labeled by all judges as non-LD (despite the fact that three of these actually were identified as LD by the schools). Given current definitions of the condition called "learning disabilities," there is little basis for the hope that school personnel are going to be able accurately to identify such students. Of course, the judges in the present study were not currently within the educational system and were not currently making decisions about the classification of students. A final test would be to have actual decision makers attempt to use patterns in the data to differentiate LD and low-achieving students.

References

Ackerman, P. T., Peters, J. E., & Dykman, R. A. Children with learning disabilities: WISC profiles. Journal of Learning Disabilities, 1971, 4, 150-166.

Algozzine, B., Forgnone, C., Mercer, C., & Trifiletti, J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives. Learning Disability Quarterly, 1979, 2, 25-31.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980.

Bussell, C., Huls, B., & Long, L. Positive reinforcers for modification of auditory processing skills in LD and EMR children. Journal of Learning Disabilities, 1975, 8, 373-376.

Federal Register. Department of Health, Education, and Welfare. Washington, D.C., 41(230), Monday, November 29, 1976, Part III.

Federal Register. Department of Health, Education, and Welfare, Washington, D.C., 42(163), Tuesday, August 23, 1977, Part II.

Federal Register. Department of Health, Education, and Welfare, Washington, D.C., 42(250), Thursday, December 29, 1977.

Gajar, A. Educable mentally retarded, learning disabled, emotionally disturbed: Similarities and differences. Exceptional Children, 1979, 45, 470-472.

Goldberg, L. R. Man versus model of man: A rationale plus some evidence for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education (2nd ed.). Boston: Houghton-Mifflin, 1981.

Shavelson, R. Teachers' decision making. Colloquium presentation, Institute for Research on Learning Disabilities, Minneapolis, October 1979.

Wagonseller, B. R. Learning disabilities and emotional disturbances: Factors relating to differential diagnosis. Exceptional Children, 1973, 40, 205-206.

Webster, R. R., & Schenck, S. J. Diagnostic test pattern differences among LD, ED, EMH, and multihandicapped students. Journal of Educational Research, 1978, 72, 75-80.

Ysseldyke, J. E., & Algozzine, B. Perspectives on assessment of learning disabled students. Learning Disability Quarterly, 1979, 2, 3-13.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. Similarities and differences between low achievers and students labeled learning disabled. Exceptional Children, in press.

Table 1

Number of Students within Categories of School and Federal Definitions of Learning Disabilities

Federal Definition[a]	School Definition	
	LD	Non-LD
Non-LD (1-14)	13	14
1.0 SD LD (15-22)	18	20
1.5 SD LD (23 or more)	19	15

[a]Numbers in parentheses indicate the ability-achievement discrepancy scores required for each operationalization of the federal definition.

Table 2

Number of Students Within Categories of School and

Low Achievement Definitions of Learning Disabilities

Low Achievement Definition[a]	School Definition	
	LD	Non-LD
Non-LD (86 or above)	11	19
1.0 SD LD (78 to 85)	22	23
1.5 SD LD (77 or below)	17	7

[a]Numbers in parentheses indicate the achievement deficit scores required for each operationalization of the low achievement definition.

Table 3

Judge's Accuracy in Identifying LD and Non-LD Students According to School and Federal Definitions

	School Definition			1.0 SD Federal Definition			1.5 SD Federal Definition		
Judge No.	% Correct LD	% Correct Non-LD	% Overall Accuracy	% Correct LD	% Correct Non-LD	% Overall Accuracy	% Correct LD	% Correct Non-LD	% Overall Accuracy
1	34.0	71.4	52.6	36.1	81.5	48.5	44.1	75.4	64.7
*2	20.0	98.0	58.6	15.3	100.0	38.4	26.5	96.9	72.7
3	24.0	81.6	52.5	27.8	96.3	46.5	35.3	86.2	68.7
#4	62.0	59.2	60.6	63.9	81.5	68.7	70.6	58.5	62.6
5	34.7	91.7	62.9	18.3	69.2	32.0	14.7	74.6	53.7
6	22.0	91.8	56.6	18.1	92.6	38.4	20.6	87.7	64.7
7	44.0	81.6	62.6	33.3	74.1	44.4	35.3	70.8	58.6
8	52.0	63.3	57.6	51.4	74.1	57.6	61.8	64.6	63.6
*9	10.0	100.0	54.6	4.2	92.6	28.3	2.9	93.8	62.6
#10	68.0	46.9	57.5	70.8	66.7	69.7	82.4	50.8	61.6
11	64.0	65.3	64.6	37.5	18.5	32.4	20.6	35.4	30.3
12	52.0	59.2	55.6	56.9	81.5	63.6	61.8	61.5	61.6
13	22.0	87.8	54.5	22.2	96.3	42.5	26.5	87.7	66.7
#14	72.0	69.4	70.7	63.9	77.8	67.7	64.7	53.8	57.6
15	50.0	59.2	54.6	55.6	81.5	62.6	61.8	63.1	62.6
16	46.0	67.3	56.5	40.3	63.0	46.5	41.2	61.5	54.5
17	54.0	87.8	70.7	40.3	85.2	52.5	41.2	70.8	60.6
#18	65.3	61.7	63.5	55.1	55.6	55.2	47.1	45.2	45.9
	t (17) = 3.89, p < .001			t (17) = 5.23, p < .001			t (17) = 3.44, p < .001		

* = least lenient judges

= lenient judges

Table 4

Judge's Accuracy in Identifying LD and Non-LD Students According to School and Low Achievement Definitions

	School Definition			1.0 SD Low Achievement Definition			1.5 SD Low Achievement Definition		
Judge No.	% Correct LD	% Correct Non-LD	% Overall Accuracy	% Correct LD	% Correct Non-LD	% Overall Accuracy	% Correct LD	% Correct Non-LD	% Overall Accuracy
1	34.0	71.4	52.6	41.4	93.1	56.6	41.7	72.0	64.6
*2	20.0	98.0	58.6	15.7	100.0	40.4	12.5	89.3	70.7
3	24.0	81.6	52.5	27.1	93.1	46.5	25.0	80.0	66.7
#4	62.0	59.2	60.6	60.0	69.0	62.6	66.7	53.3	56.6
5	34.7	91.7	62.9	30.4	100.0	50.5	56.5	89.2	81.4
6	22.0	91.8	56.6	20.0	96.6	42.4	16.7	85.3	68.6
7	44.0	81.6	62.6	32.9	72.4	44.4	29.2	68.0	58.6
8	52.0	63.3	57.6	62.9	100.0	73.7	66.7	62.7	63.7
*9	10.0	100.0	54.6	7.1	100.0	34.4	16.7	98.7	78.7
#10	68.0	46.9	57.5	67.1	55.2	63.7	62.5	40.0	45.5
11	64.0	65.3	64.6	51.4	55.2	52.6	70.8	57.3	51.6
12	52.0	59.2	55.6	42.9	44.8	43.4	41.7	52.0	49.5
13	22.0	87.8	54.5	24.3	100.0	46.5	20.8	84.0	68.4
#14	72.0	69.4	70.7	62.9	75.9	66.6	70.8	54.7	58.4
15	50.0	59.2	54.6	52.9	72.4	58.6	54.2	57.3	56.5
16	46.0	67.3	56.5	50.0	86.2	60.7	50.0	64.0	60.6
17	54.0	87.8	70.7	42.9	89.7	56.6	45.8	70.7	64.6
#18	65.3	61.7	63.5	50.0	42.9	47.9	52.2	47.9	49.0
	t (17) = 3.89, p < .001			t (17) = 4.08, p < .001			t (17) = 2.71, p < .01		

* = least lenient judges

\# = lenient judges

Table 5

Individual Judges' Leniency and Agreement with Other Judges

Judge	Leniency	Inter-judge Agreement
1	31.3	.644
2	11.1	.713
3	21.2	.22?
#4	51.5	.657
5	21.6	.418
6	15.2	.405
7	31.3	.256
8	44.4	.700
9	5.1	.275
#10	60.6	.669
11	49.5	.067
12	46.5	.271
13	17.2	.340
#14	52.5	.803
15	45.5	.575
16	39.4	.410
17	33.3	.319
#18	50.5	.333

#Lenient judges

Table 6

Rank Ordering of the Average Mean Differences Between LD and Non-LD Students for Lenient and Non-Lenient Judges[a]

Lenient	Mean Rank	Non-Lenient	Mean Rank
WISC-R Full - PIAT Total	2.00	WISC-R Comprehension	13.14
PIAT Spelling	7.38	WISC-R Full - PIAT Total	13.79
W-J Basic Skills	7.75	W-J Basic Skills	14.07
PIAT Read Recog	8.75	WISC-R Similarities	15.18
Picture Arrangement	11.38	PIAT Reading Comp	16.07
W-J Math Achiev	12.38	W-J Reading Achiev	16.32
PIAT Total	13.50	W-J Knowledge Apt	17.68
W-J Reading Achiev	14.25	WISC-R Obj Assembly	18.07
PIAT Reading Comp	14.50	PIAT Spelling	18.18
WISC-R Verbal-Performance	14.75	WISC-R Vocabulary	18.39

[a]The ranking of variables ranged from 1 to 48. The table presents the first 10 variables for the two groups of judges, starting with the one of greatest average mean difference.

Table 7

Variables Used by Judges to Classify Students

Judge 1	Judge 2
Picture Arrangement (WISC-R subtest)	WISC-R Full - PIAT Total
WISC-R Full - PIAT Total	Object Assembly (WISC-R subtest)
W-J Skills Achievement	Similarities (WISC-R subtest)
PIAT Reading Comprehension	WISC-R Performance
PIAT Reading Recognition	Block Design (WISC-R subtest)
Judge 3	**Judge 9**
WISC-R Full - PIAT Total	W-J Knowledge Aptitude
WISC-R Full	W-J Reading Aptitude
W-J Reasoning	W-J Written Language Aptitude
Similarities (WISC-R subtest)	Peterson-Quay Total
Picture Arrangement (WISC-R subtest)	PIAT Reading Comprehension
Judge 10	**Judge 14**
Picture Arrangement (WISC-R subtest)	W-J Memory
WISC-R Full - PIAT Total	WISC-R Full - PIAT Total
WISC-R Verbal - WISC-R Performance	W-J Skills Achievement
WISC-R Performance	PIAT Spelling
W-J Skills Achievement	

Judge 17

WISC-R Full - PIAT Total
W-J Skills Achievement
W-J Memory
WISC-R Full - W-J Broad Cognitive
W-J Written Language Aptitude

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, A. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D. The reliability of formative measures of written expression (Research Report No. 50). January, 1981.

Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy: A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.