AN EVALUATION OF THE REGRESSION DISCREPANCY METHOD FOR IDENTIFYING CHILDREN WITH LEARNING DISABILITIES

Lorrie Shepard, Ph.D.
Laboratory of Educational Research
University of Colorado

A learning disability (LD) exists if a child's academic achievement lags significantly behind intellectual ability and there is no other known cause for the discrepancy. The Regression Discrepancy Method using multiple regression for identifying LD children directly parallels the theoretical definition. It involves giving both an ability and an achievement test, which are normed together. An anticipated achievement score is computed for each child based on ability, grade level, and sex. Then, for each ability score, the 10% whose actual achievement is most discrepant from their anticipated achievement are identified as likely LD. In comparing this method with other identification techniques, the author discusses its advantages and limitations, pointing out that while it is conceptually and methodologically superior to other approaches, it is nonetheless seriously deficient as a sole criterion for LD identification. In an area so fraught with definitional and instrumentation problems, provision should be made to collect data independently and to trust only those diagnoses that are consistent and independently verifiable.

The Regresssion Discrepancy Method using multiple regression is one way of quantifying or operationalizing the definition of learning disabilities (LD). To decide if it is a valid identification tool, one must consider both its underlying logic and the statistical consequences of using it this way. Just as tests can be studied to see if they are valid for a specific purpose, the discrepancy assessment model (which involves a combination of test scores) can be evaluated to determine if it measures what it was intended to measure.

DEFINITION OF LEARNING DISABILITIES

Validity rests first upon the match between the measurement device and the conceptual definition of a trait. The National Advisory Committee on Handicapped Children (U.S. Office of Education, 1968) has adopted the following definition of LD:

> Children with special learning disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic. They include conditions which have been referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, developmental aphasia, etc. They do not include learning problems which are due primarily to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage. (p. 34)

Like all attempts to define LD, this definition is vague. It is similar to other formal definitions (Bateman, 1965; Kirk, 1962) in that it makes clear what LD

0022-4669/80/1401-0008$01.00/0

are *not* but fails to make clear what they are. This shortcoming in the LD definition is acknowledged by virtually all writers in the field (Gearhart, 1973; Lilly, 1977).

What stands out as the most important characteristic of children identified as LD is that they have learning problems that cannot be attributed to lack of intelligence. All LD definitions, either by connotation or denotation, rest on this discrepancy between achievement and ability. LD children are thereby distinguished from slow learners, who have low achievement but are presumably learning as fast as they are able. In some definitions, experts go on to attribute the discrepancy between performance and potential to some cerebral dysfunction (e.g., Cruickshank, 1972) or to a disruption in the system of perceptual processing. Others disagree, arguing that since the neurological causes cannot be verified and make no difference to instructional intervention, definitions should not rest on presumed etiology (Freeman, 1967; Kirk, 1971; McCarthy, 1971). However, all agree that LD exists when a child cannot learn as well as expected and there is no other known cause, such as lack of opportunity or emotional disturbance.

THE REGRESSION DISCREPANCY METHOD

The discrepancy assessment model discussed in this article is based on multiple regression and directly operationalizes the definition of LD as a discrepancy between potential and performance. Children are given a group-administered ability test, such as the Short Form Test of Academic Aptitude (SFTAA) (Sullivan, Clark, & Tiegs, 1970), and a group achievement test, such as the Comprehensive Tests of Basic Skills (CTBS) (CTB/McGraw Hill, 1974). Ability scores plus sex and grade level are used to estimate anticipated achievement for a child on each of the CTBS subtests; this is done by means of multiple regression equations developed by the test publisher. (Sample sizes vary between 1,500 and 2,000 depending on test level.) Without being distracted by the computational formulas, it is helpful to view this score as follows: The anticipated achievement score is the norm for children of the same ability, grade level, and sex. That is, the predicted score is the average achievement score for a national sample of similar children.

Confidence intervals ($\pm$ 1.28 σ) are established around the anticipated achievement scores for each examinee, using the standard error of estimate. The 10% of children whose achievement scores are furthest below expectation are considered to be significantly deficient. A child is a candidate for LD identification if this negative discrepancy occurs on any one of the major subtests. More than 10% of all children will be eligible by this criterion, since the 10% lowest are not the same individuals on all tests; i.e., a child could be low on math or language but not necessarily on both.

Although more than 10% of all children fit the discrepancy criterion, they are not all immediately considered for LD classification. Those so considered require professional staffing by an interdisciplinary team, including teachers, special educators, language specialists, and psychologists, to confirm the diagnosis. It is also suggested that any significant discrepancy be confirmed by a similar pattern on another pair of achievement and ability tests. However, because supplemental criteria are by no means consistently applied, the strengths and weaknesses of the Regression Discrepancy Method will be discussed as if it were the single diagnostic technique. Its usefulness in combination with other methods will also be considered.

STRENGTHS OF THE REGRESSION DISCREPANCY METHOD

As stated initially, the adequacy of a measurement model should be judged both logically and in terms of its statistical properties. The discrepancy assessment rule is intuitively appealing because it so directly parallels the definition of LD. Any conceptual problems that arise have to do with assumptions about the measurement of aptitude and ambiguities in the definition itself (see next section on model weaknesses). There is no apparent problem, however, with an operational criterion that does not conform to the logic of the formal definition.

The particular discrepancy method discussed in this article also has the appeal of being superior statistically and conceptually to other attempts at quantifying discrepancy. The strengths of the Regression Discrepancy Method will be demonstrated by briefly reviewing these other approaches.

Years-below method

The years-below grade-level criterion for identifying LD children uses the national norm on a standardized achievement test as the expected achievement level for all children. Those who score significantly below grade level (usually 1 year below in the early grades, more in upper grades) are considered LD.

This method of identification has been used in a number of school districts and is cited by Erickson (1975) as a common procedure. It has serious conceptual problems, however. Primarily, it will identify slow learners (including those with limited intellectual capacity) rather than those whose performance is significantly discrepant from their potential. Bright and average children who have trouble learning some subjects would not be identified.

A secondary problem is the use of an arbitrary cutoff for significance, such as 1 year below grade level. Because the standard deviation in grade-equivalent units increases with grade level, more and more children will be identified as grade level increases without any real change in learning ability.

Mental-age method

Some time ago, Harris (1970) proposed the simple technique of determining expected grade equivalent (EGE) in reading based on ability by subtracting 5 years from the child's mental age (MA):

$$EGE = MA - 5.$$

Thus, the average 6-year-old first grader $(6 - 5 = 1)$ has an achievement grade-equivalent expectation of 1. Children performing below this expectancy are in need of special help.

Although this approach takes ability into account, there are still serious weaknesses. No account is taken of normative differences in how children with very high and very low MAs can actually be expected to perform. (That is, the formula does not allow for our expectation that a 7-year-old whose MA is 5 will achieve differently from a 5-year-old with the same MA because of differences in exposure to instruction.) It also fails to consider the less than perfect correlation between ability and achievement. Because of regression effects, it will identify proportionally more bright pupils than dull pupils as LD. Further, since Harris gave no guidance as to how large a discrepancy to consider important, there is a danger of overinterpreting small discrepancies that are perfectly within the range of normal fluctuations in learning. Variations of the MA method, combined with use of chronological age (CA), such as the Horn formulas (Harris, 1970), have the same general problems.

Bond and Tinker discrepancy method (1967)

These authors attempted to take both grade level and ability into account by computing expectancy as:

$$\text{EGE} = \text{years in school} \times \frac{\text{IQ}}{100} + 1.0.$$

Their formulation, however, while conceptually appealing, uses the IQ score as if it were on a ratio scale of measurement — which it is not. No account is taken of the regression of IQ on achievement or the fact that the variance in normal grade-equivalent performance increases with grade level. Errors in computing expectancy will be greater the further one's IQ is from 100. Moreover, if a constant rule is used to identify significant discrepancy (e.g., 1 year below), many more children will be identified at higher grade levels. In an empirical study Erickson (1975) demonstrated that the Bond and Tinker method identifies essentially the same children as does the years-below criterion.

It should be noted that other formulas that attempt intuitive weightings of IQ and grade level or age will have essentially the same problems as the Bond and Tinker method. This is true of the definition of severe discrepancy proposed (but not adopted) by the Bureau of Education for the Handicapped (1976):

$$\text{severe discrepancy} = \text{CA}\ (\frac{\text{IQ}}{300} + .17) = 2.5.$$

This formula based on CA ignores both grade-level effects and the change in variance across grades (by using a constant 2.5 cutoff). It mistakenly treats IQ as if it were on a ratio scale and does not adequately take into account the regression of achievement on IQ.

Z-score discrepancy method

Erickson (1975) recommended the Z-score discrepancy method because it alleviates many of the problems of the foregoing procedures. It involves computing Z scores on both an IQ and achievement test — standard scores that express an individual's distance from the mean in standard deviation units: Z = (raw score − group mean) ÷ standard deviation. The achievement score is expected to be in about the same relative position in the distribution as the ability score. (Note that essentially the same comparison can be made using percentile ranks on the two tests.) Erickson did not specify a rule for determining how big a discrepancy should be considered significant, but in her study took the 10% with the largest negative discrepancies (IQ minus achievement).

This method takes ability and grade level into account, and, although it is not mentioned, the *Z*-score method adjusts for differences in variability across grades by using the standard deviation for the particular grade. The only serious drawback is that it does not take into account the regression between achievement and IQ. The regression phenomenon is well known in statistics and measurement theory, but it seems to have gone unrecognized in the literature pertaining to the quantification of LD.

The *Z*-score method is based on the assumption that, on the average, a child's *Z* score on achievement will be the same as his or her *Z* score on IQ. This would be true if achievement and IQ were perfectly correlated ($r = 1.0$). However, because the actual correlation is usually more on the order of .6, there will be regression to the mean. It can be shown both mathematically and theoretically that bright (high-IQ) children have above-average achievement, but their relative position (i.e., *Z* score) tends not to be as high as it is in the IQ distribution. Conversely, children with low IQs will, on the average, have relatively higher achievement status than IQ status, although they will still tend to be below the mean.

The figure illustrates the difference between the actual IQ-achievement correlation and the assumed perfect correlation. Because of actual regression the false, perfect-correlation expectation is too low for those of low ability and too high for those of high ability; therefore, the brighter a child is, the more likely he or she is to show a large discrepancy by this method. For example, students A and D are both achieving only slightly below the true expectation or norm, but by the *Z*-score method A should be considered significantly discrepant. At the same time, students F and B have the same serious deviations from the true regression line, but student F would be missed by the *Z*-score criterion.

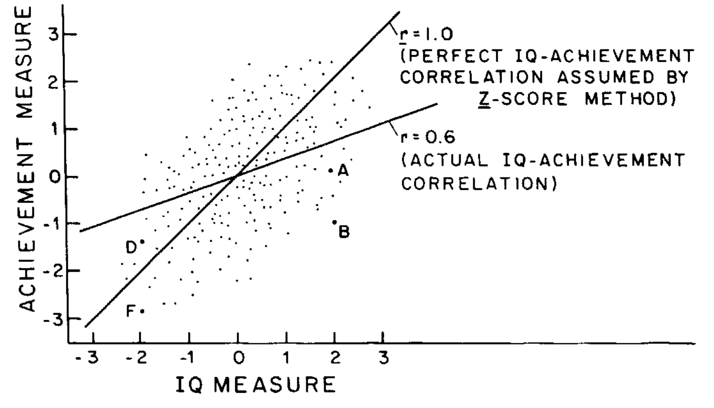

Figure 1. Example of exaggerated discrepancies for bright students when *r* = .6 is assumed to be 1.0.

Although Erickson did not consider it, there is a simple solution to the regression problem without resorting to the greater complexity of multiple regression; however, the solution is only workable with a large number of cases. Instead of taking the 10% overall who have the largest difference score, one could instead identify the 10% for each IQ score (or interval) who have the lowest achievement. Because of regression, brighter students will have larger discrepancies, but if the lowest 10% are identified score by score, the method will sample equally from the full performance continuum. This procedure would be the next-best alternative if multiple regression techniques could not be used.

Regression Discrepancy Method

The Regression Discrepancy Method, based on multiple regression, has several advantages over other statistical models:

1. Expected performance is predicted from aptitude scores, so that children from the full ability continuum will be identified. Therefore, LD children are clearly distinguished from slow learners.
2. Grade level is used in the equation, so that apparent discrepancies will not be created by differences in opportunity to learn.
3. By measuring discrepancies in standard errors and taking the 10% with the greatest deviation at each grade level, the percent identified will automatically be the same at each grade level and will not be influenced by scaling artifacts.
4. The multiple regression technique takes variables into account according to their actual relationship to achievement (based on data) rather than concocting a formula (like Bond and Tinker's) that is only accurate at the mean.
5. Finally, as implemented using the CTBS and SFTAA or other co-normed tests, errors due to differences in standardization populations are precluded, since these tests were normed together. If the Regression Discrepancy Method were to be used with an achievement and ability test not concurrently normed, some discrepancies would be created by sampling differences and other real discrepancies would be obscured.

WEAKNESSES OF THE REGRESSION DISCREPANCY METHOD

Two types of errors

A useful way to consider weaknesses in the Regression Discrepancy Model is to analyze how likely it is to create errors in classification. Whenever categorization or selection decisions are made, two types of errors can result (as well as two kinds of correct identification). These are generally called false-positive and false-negative errors; i.e., respectively identifying a child as having LD when he or she does not or, conversely, failing to detect real disabilities.

In Table 1, hypothetical data have been arranged to illustrate what a two-way table would look like if a test (or combination of test scores) were perfectly accurate in identifying LD children. In this idealized example, there are no

TABLE 1
IDEAL EXAMPLE OF TRUE-TEST
CLASSIFICATION OF LEARNER TYPE

		True classification: LD	True classification: Normal	
Test classification	LD	5	0	5
	Normal	0	95	95
		5	95	

instances where the test labeled as LD a child who was actually normal; nor are there cases where the test designated as normal a child who really had LD.

Table 2 is also based on hypothetical data, but is more like a real situation where a test will make some errors in categorization. Notice that the marginals were kept the same as in the previous example; that is, 5% really are LD and the test identified 5% as LD. However, the example has been contrived to illustrate what usually happens when one seeks to detect a relatively rare characteristic with a fallible measure. That is, there are more normal children falsely identified as LD than there are true instances of the disorder correctly diagnosed.

The 3:1 ratio of errors (two types) to correct LD identification in this example would stay the same even if one were to change the test cutoff score to designate the lowest 10% rather than the lowest 5% as requiring help. As can be seen in Table 3, a more liberal test criterion will reduce the number of real LD children missed — but at the expense of incorrectly labeling many more normal learners as LD. The only way to reduce both types of errors at the same

TABLE 2
EXAMPLE OF TRUE-TEST
CLASSIFICATION OF LEARNER TYPE WITH ERRORS

		True classification: LD	True classification: Normal	
Test classification	LD	2	3	5
	Normal	3	92	95
		5	95	

TABLE 3
EXAMPLE OF TRUE-TEST CLASSIFICATION
WITH ERRORS AFTER ADJUSTING TEST CUTOFF

		True classification: LD	True classification: Normal	
Test classification	LD	3	7	10
	Normal	2	88	90
		5	95	

time is to increase the validity of the test; changing the cutoff score only shifts the errors.

Causes of false-positive errors

Misidentification of normal children as LD will occur both because of statistical artifacts and because there are other real causes of discrepant profiles. Difference scores are notoriously unreliable (Bereiter, 1967; Salvia & Ysseldyke, 1978; Thorndike & Hagen, 1977; Webster & Bereiter, 1967). This is because there is measurement error in both instruments and because the reliable portions of each are largely redundant and therefore do not contribute to the reliability of the difference. Although the Regression Discrepancy Method is slightly more complicated than a simple difference score, it is still vulnerable to this problem. Using large confidence intervals (that is, only trusting extreme deviations) provides protection against this problem so long as one does not repeatedly administer tests until a discrepancy is found.

There is a further, slightly more esoteric, statistical problem caused by differences in the reliabilities of the tests used in identifying discrepant profiles. Cronbach, Gleser, Nanda, and Rajoratnam (1972) have emphasized that not only can imperfect reliabilities create discrepancies when there are none and vice versa; in addition, imperfect and differential reliabilities can also change the direction of the discrepancy — e.g., a child who appears to be worse off in achievement compared with ability could actually turn out to be a significant overachiever when the dependability of the measures is taken into account. The solution to this problem is to use estimated true scores or universe scores for profile interpretation (see Cronback et al., 1972, p. 310). This problem is given only passing attention here because it is so technical; because the Regression Discrepancy Method at least takes some sources of error in the ability measure into account by using expected achievement for comparison; and because the differences in reliabilities are not extreme in this case, being on the order of .92 for the SFTAA and .98 for the CTBS. The inaccuracies caused by differences in reliabilities will be more serious in situations where ability and achievement measures other than the SFTAA and CTBS (or equally reliable measures) are used. The greater the difference in reliabilities, the greater the error in the pattern of profiles.

Thus far in the discussion it has been assumed that the SFTAA and CTBS — or other co-normed tests such as the Otis-Lennon Mental Ability Test (Otis & Lennon, 1967) and Stanford Achievement Test (Madden, Gardner, Rudman, Karlsen, & Merwin, 1973) or the Cognitive Abilities Test (Thorndike, Hagen, & Lorge, 1971–74) and the Iowa Tests of Basic Skills (Lindquist, Hieronymus, et al, 1972) — would be used in implementing the Regression Discrepancy Method. If this is not the case, statistical problems escalate dramatically, and there is not even any way of knowing the likely direction of the errors. If the two tests used were not normed on the same national sample, discrepancies can easily be created or obscured by noncomparable norm samples. This dilemma will be confronted frequently because clinicians prefer the better individual intelligence tests, which have greater demonstrated validity than group-administered tests. Although the best IQ tests (e.g., the WISC-R and Stanford-Binet) will still be useful to confirm level of functioning and to provide other clinical data, it would be very difficult to accurately interpret discrepancies between percentiles or standard scores from these tests and any achievement test, since joint norms are not available. Normative differences are likely to be so large and misleading that they would completely confound corrections for lesser statistical problems discussed earlier, for example, regression based on the actual, but unknown, correlation between the two measures.

In addition to measurement errors that generate false discrepancies, an ability-achievement discrepancy can be caused by factors other than an underlying LD. For example, prolonged absence from school or a nontraditional curriculum that greatly underemphasizes basic skills could create such a profile. (A child in these circumstances would need more instruction, but not instruction tailored to a particular learning handicap.) It is, of course, well recognized in the LD literature that lack of opportunity to learn should not be mistaken for an LD diagnosis. This possibility, however, underscores the need for identification not to rest solely on arbitrary interpretation of test criteria.

It is also plausible that ability-achievement discrepancies reflect poor motivation not extreme enough to warrant being termed emotional disturbance. Experts in special education should be somewhat disconcerted by the realization that the Regression Discrepancy Method, which is the best available identification approach to LD, is exactly the same as the method recommended to identify underachievers in the counseling literature for the last 20 years (see Thorndike, 1963). In this context, the explanation for the discrepancy is assumed to be motivational, and treatments are designed to increase effort either directly or indirectly by influencing enthusiasm or self-concept.

Of course, the best defense against these criticisms of the Regression Discrepancy Method is to argue that the 10% identified include some of each of these kinds of cases (i.e., true LD, measurement errors, motivational problems, and lack of opportunity to learn) and that further clinical diagnosis will sort out the true LD cases from the rest. The weakness in this claim, however — besides the false negatives discussed in the next section — is that there is virtually no validity evidence to demonstrate that experts can tell which is which within this pool of cases. From an extensive review of validation studies of the 10 tests most frequently used in diagnosing LD, Coles (1978) concluded that "the predominant finding in the literature suggests that each test fails to correlate with a diagnosis of learning disabilities" (p. 326); that is, the tests cannot distinguish LD from normal learners. One of the possible explanations offered by Coles, apart from the failure to establish construct validity for each of the tests, was the frequently faulty identification of LD used in the studies to test the tests. For example, several authors have found that LD children have significantly below-average IQ scores (Hallahan & Kauffman, 1977; Kirk & Elkins, 1975), even though by definition IQ should be uncorrelated with LD. In a study by Routh and Roberts (1972), significant correlations between academic achievement and neurological "soft" signs disappeared when IQ and age were partialled out. Other major reviews cast doubt on the validity of certain tests used in diagnosing LD (q.v. Arter & Jenkins, 1979; Newcomer & Hammill, 1976).

Causes of false-negative errors

The causes of false-negative errors — i.e., failing to identify a child who really has LD — are again both statistical and conceptual. Since the measurement and prediction problems dicussed above can operate in either direction, they are the source of negative as well as positive errors. The only difference is that negative errors will be relatively less frequent because there are many fewer LD than normal children to be identified.

The most serious flaw in the Regression Discrepancy Method which will cause LD children to be missed is conceptual. It is the assumption that the paper-and-pencil, group-administered ability test is a reasonably good measure of potential. To the extent that a child has a LD hindering performance on both the ability and achievement tests, he or she will be classified as a slow learner and will not have a sufficient discrepancy to warrant the LD designation. (For example, Nelson Rockefeller is often cited as a person with a severe LD because he was bright, but could not read.) Such a person is likely to do poorly on both tests, since both require reading. It would take an individually administered oral IQ test to discover the true discrepancy in performance. (And, unfortunately, no individual ability and achievement tests have been conormed.)

CONCLUSION: PROTECTION AGAINST TWO TYPES OF ERRORS

Although the Regression Discrepancy Method can be considered the best available quantification procedure for diagnosing LD, it must still be judged seriously deficient. Actual numbers cannot be computed without empirical studies, but it is likely that the Regression Discrepancy Method falsely labels more normal children as LD than it correctly identifies children who really have a disorder. At the same time, errors of overidentification do not assure that all real instances of LD will be detected. Rather, as the contingency tables illustrate, some LD children are missed, and at the same time, normal children are misclassified.

The best way to prevent false identification of normal learners is to require confirmation of the specific pattern of discrepancies (i.e., in the same subject areas) by another pair of tests given at a different time. It would be acceptable to repeat the first pair of tests given in a different year; however, it is not permissible to use one of the current scores and retest only on the other, since as likely as not one would be retaining an aberrant score which "caused" the apparent discrepancy.

To avoid missing LD children because they do poorly on both the ability and achievement measures, other eligibility criteria would have to be entertained. Teacher nominations should be considered, since they have the opportunity to observe contradictions between oral and written performance. However, since teacher judgments have been shown to be less reliable than standardized tests and are just as prone to false-positive errors, these observations would have to be independently confirmed by individually administered tests of both IQ and achievement.

It is a good general rule when dealing with fallible measures never to trust a diagnosis unless it is independently confirmed by other measures. This provides for a sort of triangulation (cf. Webb, 1966). Since errors are of different kinds in each measurement approach, they are not likely to be consistent; therefore, when results concur, there is some assurance that the conclusion is valid. The requirement for independent confirmation calls for some double testing, but it distinctly precludes the practice of repeatedly testing until a discrepancy is found and then recommending LD classification. In such a case the weight of evidence is against LD, and the single discrepancy should be considered a fluke. Knowing the flaws in diagnostic tests and the discrepancy model, experts must avoid abusing the factor of chance by overinterpreting single indicators of LD in the same way that a statistician must resist interpreting a significant result after a string of nonsignificant tests (i.e., if you do it enough times, significance will occur just by chance).

Unfortunately, the diagnostic team approach, which is so highly regarded in the field of learning disabilities (Kirk, 1972; Learner, 1976; Myers & Hammill, 1969; Swanson & Willis, 1979), prevents the kind of independence that is essential to demonstrate the validity of diagnoses. The process of seeking concensus in team staffing tends to accentuate similarities in perceptions and ignores the frequent disparities which are inevitable due to faulty measures and the enormous amount that remains unknown in this field.

Diagnostic practices should be better tailored to reflect how subject to error they are. Because the information from various experts is so important, diagnostic teams will not be disbanded; but provision should be made to collect judgments independently prior to group discussion. Lack of convergence should be seriously considered as evidence that a specific LD does not exist.

References

Arter, J. A., & Jenkins, J. R. Differential diagnosis-prescriptive teaching: A critical appraisal. *Review of Educational Research,* 1979, *49,* 517–555.

Bateman, B. An educator's view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders* (Vol. I). Seattle: Special Child Publications, 1965.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1967.

Bond, G., & Tinker, M. *Reading difficulties: Their diagnosis and correction* (2nd ed.). New York: Appleton-Century-Crofts, 1967.

Bureau of Education for the Handicapped. *Federal Register,* November 29, 1976, *41*(230), 52407.

Coles, G. S. The learning-disabilities test battery: Empirical and social issues. *Harvard Educational Review,* 1978, *48,* 313–340.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley, 1972.

Cruickshank, W. M. Some issues facing the field of learning disabilities. *Journal of Learning Disabilities,* 1972, *5,* 380–388.

CTB/McGraw Hill. *Comprehensive Tests of Basic Skills.* Monterey, Calif.: Author, 1974.

Erickson, M. T. The Z-score discrepancy method for identifying reading-disabled children. *Journal of Learning Disabilities,* 1975, *8,* 308–312.

Freeman, R. Special education and the electroencephalogrami: A marriage of convenience. *Journal of Special Education,* 1967, *2,* 61–73.

Gearhart, B. R. *Learning disabilities: Educational strategies.* St. Louis: Mosby, 1973.

Hallahan, D. P., & Kauffman, J. M. Labels, categories, behaviors: ED, LD, EMR reconsidered. *Journal of Special Education,* 1977, *11,* 139–149.

Harris, A. *How to increase reading ability* (4th ed.). New York: McKay, 1961.

Kirk, S. A. *Education exceptional children.* Boston: Houghton Mifflin, 1972.

Kirk, S. A. Learning disabilities: The view from here. In D. D. Hammill & N. R. Bartell (Eds.), *Educational perspectives in learning disabilities.* New York: Wiley, 1971.

Kirk, S. A., & Elkins, J. Characteristics of children enrolled in the Child Service demonstration centers. *Journal of Learning Disabilities,* 1975, *8,* 630–637.

Learner, B. W. *Children with learning disabilities.* Boston: Houghton Mifflin, 1976.

Lilly, M. S. A merger of categories: Are we finally ready? *Journal of Learning Disabilities,* 1977, *10,* 115–121.

Lindquist, E. F., Hieronymus, A. N., et al. *Iowa Tests of Basic Skills.* Boston: Houghton Mifflin, 1972.

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. *Stanford Achievement Test.* New York: Psychological Corp., 1973.

McCarthy, J. M. Learning disabilities: Where have we been? Where are we going? In D. D. Hammill & N. R. Bartell (Eds.), *Educational perspectives in learning disabilities.* New York: Wiley, 1971.

McGrady, H. J. From diagnosis to remediation. In D. D. Hammill & N. R. Bartel (Eds.), *Educational perspectives in learning disabilities.* New York: Wiley, 1971.

Myers, P., & Hammill, D. *Methods for learning disorders.* New York: Wiley, 1969.

Newcomer, R. L., & Hammill, D. D. *Psycholinguistics in the schools.* Columbus, Ohio: Merrill, 1976.

Otis, A. S., & Lennon, R. T. *Otis-Lennon Mental Ability Test.* New York: Psychological Corp., 1967.

Routh, D. K., & Roberts, R. D. Minimal brain dysfunction in children: Failure to find evidence for a behavioral syndrome. *Psychological Reports,* 1972, *31,* 307–314.

Salvia, J., & Ysseldyke, J. E. *Assessment in special and remedial education.* Boston: Houghton Mifflin, 1978.

Sullivan, E. T., Clark, W. W., & Tiegs, E. W. *Short Form Test of Academic Aptitude.* Monterey, Calif.: CTB/McGraw Hill, 1970.

Swanson, B. M., & Willis, D. J. *Understanding exceptional children and youth.* Chicago: Rand McNally, 1979.

Thorndike, R. L. *The concepts of over- and* *under-achievement.* New York: Columbia University, Teachers College, 1963.

Thorndike, R. L., & Hagen, E. P. *Measurement and evaluation in psychology and education* (4th ed.). New York: Wiley, 1977.

Thorndike, R. L., Hagen, E. P., & Lorge, I. *Cognitive Abilities Test.* Boston: Houghton Mifflin, 1971–74.

U.S. Office of Education. *National Advisory Committee on Handicapped Children: First Annual Report.* Washington, D.C.: Author, 1968.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally, 1966.

Webster, H., & Bereiter, C. The reliability of changes measured by mental test scores. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1967.