DISCREPANCY MODELS IN THE IDENTIFICATION OF LEARNING DISABILITY

Kenneth A. Kavale, University of Iowa

On April 6, 1963, Samuel A. Kirk told a parent advocacy group that "Recently, I have used the term 'learning disability' to describe a group of children who have disorders in development, in language, speech, reading, and associated communication skills needed for social interaction" (Kirk, 1975, p. 9). By 1968, "specific learning disability" (LD) became a federally designated category of special education (U.S. Office of Education, 1968). The formal definition offered at the time has not changed substantively and was reaffirmed in the 1997 reauthorization of the Individuals With Disabilities Education Act [IDEA] (Public Law 105-17) as follows:

> The term "specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in imperfect ability to listen, think, speak, read, write, spell or do mathematical calculation. The term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such term does not include a learning problem that is primarily the result of visual, hearing or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage (IDEA Amendments of 1997, PL105-17, 11 Stat. 37 [20 USC 1401(26)]).

DISCREPANCY AND INTRA-INDIVIDUAL DIFFERENCES

The federal definition does not stipulate procedural guidelines for LD identification. In fact, the definition is primarily exclusive, describing what LD is not rather than identifying what LD is. Consequently, operational definitions necessary for practice have usually considered factors that may not have been articulated in the formal definition. One such factor that originated in the Kirk (1962) LD definition was the notion of intra-individual differences, the possibility of subaverage functioning in only a few areas with average or above functioning in other areas. Gallagher (1966) termed these "developmental imbalances" that were represented by discrepancies in psychoeducational functioning. One of the first such discrepancies investigated was related to the cognitive abilities of students with LD. Using subtest scores from cognitive assessments like the Wechsler Intelligence Scale for Children (WISC), patterns of strengths and weaknesses were examined to determine whether the resulting scatter ("profile") differentiated students with LD from other average or low achieving populations.

COGNITIVE DISCREPANCIES

The clinical use of scatter-analysis methods has precipitated debate about its relationship to the nature of LD (e.g., Miller, 1980; Wallbrown, Blaha, & Vance, 1980; Wallbrown, Vance, & Blaha, 1979). For example, hypotheses about uniqueness assume that the profile for samples of students with LD is characteristic of the entire LD population or that the LD subtest profile varies significantly from the average population. The empirical evidence, however, has not supported any assumptions about LD profile uniqueness (e.g., Dudley-Marling, Kaufman, & Tarver, 1981; Gutkin, 1979; Kaufman, 1981).

In a comprehensive quantitative synthesis, Kavale and Forness (1984) found no WISC profile for students with LD. For example, a discrepancy between Verbal IQ and Performance IQ (VIQ-PIQ) has been assumed to be a primary LD characteristic. The difference (PIQ > VIQ) was, on average, only 3 IQ points, which was well below the requisite 11 IQ points necessary for significance. In addition, although students with LD generally performed more poorly on Verbal subtests, no Verbal or Performance subtest score fell below the *average* level. Any measure of WISC inter-subtest variability ("scatter") was not significant and indicated no subtest strength or weakness that distinguished LD performance.

On the basis of hypotheses about cognitive performance, a number of different subtest score groupings have been proposed to reveal discrepant abilities. One method involves recategorizing subtest scores exemplified in the proposal by Bannatyne (1968) that included a Spatial, Conceptual and Sequential category, each based on three WISC subtests. An LD sample was presumed to show a Spatial > Conceptual > Sequential pattern, but, although exhibiting the required pattern, the magnitude of the score differences was well below required significance values. A second primary method was to seek a profile that either specifies particular subtest scores as high or low or identifies subtests where students with LD might score low. For example, Ackerman, Peters, and Dykman (1971) studied the ACID profile (low scores on the Arithmetic, Coding, Information, and Digit Span subtests) but, again, LD performance did not reach the required level of significant suppression. Similarly, WISC factor scores (e.g., Naglieri, 1981) and WISC patterns (e.g., Myklebust, Bannochie, & Killen, 1971) have also been investigated, but in no instance was discrepant LD performance at a level that could be termed significant.

The longstanding criticism (e.g., Bijou, 1942) of examining discrepancies in cognitive performance to identify LD appears justified. In summarizing the available research, Kavale and Forness (1984) concluded that "Regardless of the manner in which WISC subtests were grouped and regrouped, no recategorization, profile, pattern, or factor cluster emerged as a 'clinically' significant indicator of LD. In fact, when compared to average levels, the LD group was found to exhibit no significant deviations, and on average, revealed less variability than normal populations" (p. 150).

ORIGINS OF ABILITY-ACHIEVEMENT DISCREPANCY

The failure to find significant cognitive (IQ) discrepancies in LD populations and the desire to reinforce notions about the academic achievement deficits associated with LD directed attention to the possibility of conceptualizing IQ-achievement discrepancies as a feature of LD. The IQ-achievement discrepancy notion was introduced by Bateman (1965) in a definition of LD that included the description of "an educationally significant discrepancy between estimated intellectual potential and actual level of performance related to basic disorders in the learning processes" (p. 220).

The idea of IQ-achievement discrepancy was introduced by Franzen (1920) in the "Accomplishment Quotient" (AQ). The AQ is the ratio of Educational Quotient (EQ) to Intelligence Quotient (IQ). The importance of IQ "lies in its diagnosis of power of adaptation, and it has a high correlation with maximum possible school progress" (p. 434) while the EQ "is the quotient resulting from the division of the age level reached on the test in question by the chronological age of the pupil" (p. 435). "[T]he ratio of EQ to IQ [the AQ] gives the percentage of what that child could do, that he has actually done" (p. 436).

In cases where the AQ is less than 90, there is potential "underachievement." A number of analyses appeared to show that, in general, AQs were typically less than unity (1.00) (e.g., McPhail, 1922; Pintner & Marshall, 1921; Ruch, 1923). The resulting discrepancy demonstrated by the "laggards" was often attributed to "laziness (i.e., lack of effort) and if pupils are pushed to the extreme limit of their ability, the correlation between their educational achievement and their intelligence is not only high but actually reaches unity" (Whipple, 1922, p. 600). In general, there was a belief that "bright" students were achieving less than expected, relative to ability, than were "dull" students whose lagging performance was presumed to indicate limited effort. Interestingly, with the 1920s view of intelligence as a fixed entity, IQ was regarded as an index of the upper limit for academic attainment which meant AQs really could not exceed unity. As suggested by Franzen (1920), "One's differences when EQ is subtracted from IQ are always positive when they are large enough to be significant and small enough to seem spurious when they are negative.…It is safe, therefore, for practical use to assume that the optimum accomplishment is 1.00" (p. 436).

In reality, findings surrounding the AQ were unreliable because of a number of psychometric and statistical problems. In a comprehensive analysis, Toops and Symonds (1922) discussed a number of flaws with the AQ that were a presage of many later analyses of ability-achievement discrepancy. Many other critiques appeared; for example, J. C. Chapman (1923) pointed out the unreliability of using difference scores based on intelligence and achievement test scores. W. R. Wilson (1928) suggested that "Conclusions based on the use of the accomplishment quotient are misleading unless they take into account the reliability of the measures employed, the validity of the measures employed, and the part played by factors determining the intelligence quotient in school achievement under conditions of maximum maturation" (p. 10).

The major statistical criticism of AQ surrounded the operation of the "regression effect" (Crane, 1959; Cureton, 1937). The calculation of AQ assumed an almost perfect correlation between IQ and EQ, whereas the value is closer to 0.60. With less than perfect correlation between measures, scores well above average on one measure will be less superior on the second measure, and at the other end of the continuum, those scores well below average on the first measure will be less inferior on the second. Consequently, if AQ does not account for the effects of statistical regression, then there will be an overrepresentation of "bright" students and an underrepresentation of "dull" students. This result was demonstrated by Popenoe (1927) who found that "Instead of each pupil having an equal chance to get a favorable accomplishment quotient, it appears that out of almost five hundred pupils, in no case did an individual having a high intelligence quotient get a favorable accomplishment quotient, and that individuals having a low intelligence quotient obtained accomplishment quotients far above the average. So an AQ of 100 means an entirely different thing in a part of the range from what it does in another" (p. 45). The many difficulties with AQ led to the conclusion that "the administrative use of the accomplishment quotient is open to serious criticism" (p. 47) and foreshadowed many later issues about the use of ability-achievement discrepancy for LD classification.

Discrepancy and LD Identification: Rules and Regulations

The Bateman (1965) notion of discrepancy was not formally incorporated into the federal LD definition. In fact, there was no modification of the LD definition in the 1975 Education for All Handicapped Children Act (Public Law 94-142), indicating that an inherent vagueness and imprecision remained, as well as difficulties in using the definition in actual practice (Kavale & Forness, 2000). In an attempt to remedy the situation, the then Bureau of Education for the Handicapped issued regulations outlining procedures for LD sdentification. The U.S. Office of Education (USOE; 1976) regulations read as follows:

> A specific learning disability may be found if a child has a severe discrepancy between achievement and intellectual ability in one or more of several areas: oral expression, written expression, listening comprehension or reading comprehension, basic reading skills, mathematics calculation, mathematics reasoning, or spelling. A "severe discrepancy" is defined to exist when achievement in one or more of the areas falls at or below 50% of the child's expected achievement level, when age and previous educational experiences are taken into consideration (p. 52405).

Formula-Based Discrepancy

To assist the process, a formula to determine the presence of a severe discrepancy level (SDL) was proposed, but comments and testimonies about its usefulness were decidedly negative. For example, Lloyd, Sabatino, Miller, and Miller (1977) objected to the use of general intelligence measures and the negative effects of measurement error on accuracy, while Sulzbacher and Kenowitz (1977) objected to the standard 50% discrepancy across academic areas. In an empirical analysis of the SDL, Algozzine, Forgnone, Mercer, and Trifiletti (1979) cast doubt on the 50% discrepancy level "except for children whose measured intelligence falls exactly at 100" (p. 30). Danielson and Bauer (1978) reviewed the issues surrounding formula-based classification procedures and concluded by questioning whether "a technically adequate solution to the problem of LD sdentification exists" (p. 175).

By 1977, the SDL formula was dropped but not the concept of discrepancy as stipulated in regulations indicating the following:

> A team may determine that a child has a specific learning disability if: (1) The child does not achieve commensurate with his or her age and ability in one or more of the areas listed in paragraph (2) of this section, when provided with learning experiences appropriate for the child's

> age and ability levels; and (2) The team finds that a child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas: (i) oral expression, (ii) listening comprehension, (iii) written expression, (iv) basic reading skill, (v) reading comprehension, (vi) mathematics calculation, or (vii) mathematics reasoning. (USOE, 1977, p. 65083)

Thus, discrepancy was reinforced as the primary criterion for LD identification (see Chalfant & King, 1976) and, although not given precise specification in a particular formula, became over time almost the exclusive variable used for LD eligibility determination (Frankenberger & Fronzaglio, 1991; Mercer, Jordan, Allsopp, & Mercer, 1996).

Quantifying Discrepancy: Methods

With the idea that a severe discrepancy must be demonstrated, individual states were free to choose their own methodology, but wide variation in procedures introduced a substantial element of arbitrariness to LD identification (Divoky, 1974; Shepard, 1983). Nevertheless, an *in numeris veritas* [in numbers there is truth] mentally developed, and different means of quantifying the presence of a severe discrepancy were attempted even though "there is little reason to believe and much empirical reason to disbelieve the contention that some arbitrarily weighted function of two variables will properly define a construct" (Cronbach & Furby, 1970, p. 79). A significant question arose: Can two variables (ability and achievement) be combined to determine the presence or absence of a construct (LD)? The theoretical problems were exacerbated by practical difficulties surrounding the notion of prediction. As pointed out by Thorndike (1963), prediction is almost always imperfect because of (1) errors of measurement, (2) heterogeneity of the criterion (i.e., achievement), (3) limited scope of the predictors, and (4) impact of varied experiences upon the individual.

Grade-Level Deviation

The simplest but least sophisticated discrepancy method examines grade level deviations where an expected grade level (EGL) score is compared to an actual grade level (AGL) score and the discrepancy is calculated from the EGL-AGL difference. For example, expected grade level might be based on chronological age (CA), and then discrepancy calculated in terms of "years behind" (CA – 5). The 5 represents the 6 years of informal activity before school entry, with one year subtracted because the real AGL is 1.0, not 0. When the difference is "significant" (usually 1 to 2 years below grade level), a discrepancy exists. The most fundamental problem is the lack of consideration for the level and degree of instruction received. In place of CA, mental age (MA) was substituted because of the presumed closer relationship between intellectual ability and school achievement (Harris, 1961). The search for increased accuracy led to formulas with additional factors and differential weighing of variables (Harris, 1971; Monroe, 1932). Although no formula proved entirely satisfactory, the grade level deviation method was at one time a relatively common procedure for LD identification in research studies (e.g., J. S. Chapman & Boersma, 1979; Gottesman, 1979; Selz & Reitan, 1979).

Expectancy Formulas

The next type of discrepancy calculation involves more comprehensive expectancy formulas including some combination of variables (usually IQ and perhaps CA, MA, years in school [YS], or grade age [GA]). The USOE (1976) SDL formula provides an example:

$$SDL = CA\left(\frac{IQ}{300} + 0.17\right) - 2.5$$

Earlier examples were provided by Bond and Tinker (1973), Harris (1975), and Johnson and Myklebust (1967) .The Bond and Tinker formula is

$$\left(YSx\,\frac{IQ}{100} + 1.0\right) - AGL$$

The underlying logic for the Bond and Tinker formula seems confounded; an IQ score was included to account for unequal learning rates, but the included constant (1.0) makes this point moot because it negates the differential effects of IQ during the first 6 years of life (Dore-Boyce, Misner, & McGuire, 1975). To remedy this confounding, one set of proposed formulas (Horn, 1941) assigned different weights to MA and CA so formulas may be applied at four different age ranges and the problem of unequal learning rates presumably negated. Without some modification of this sort, the Bond and Tinker (as well as the Harris) formulas are poor predictors that over- and underidentify students with low and high IQs, respectively (Alspaugh & Burge, 1972; Rodenborn, 1974; Simmons & Shapiro, 1968).

Problems and Issues

The formula proposed by Johnson and Myklebust (1967) introduced the problem of interpreting ratio scores in determining discrepancy level. The Johnson and Myklebust formula calculates an expectancy level (MA + CA + GA / 3), but instead of a direct comparison (EGL – AGL), discrepancy is calculated from a ratio score (AGL / EGL 100) with a value less than 90 considered significant.

Because of the absence of an absolute zero and equal intervals, ratio scores do not possess inherent meaning. Only extreme scores are meaningful on what is really an ordinal scale, and a value such as 90 cannot be interpreted to mean 90% of average, for example. The situation is further complicated by the variable standard deviations (SDs) across age levels which means that the significance of a given discrepancy ratio will vary from one grade to another. The difficulties such SD variability causes were demonstrated by Macy, Baker, and Kosinski (1979) where the Johnson and Myklebust (1967) discrepancy quotients were quite variable across different combinations of age, grade, and content areas.

The expectancy formula approach to discrepancy calculation has been roundly criticized. McLeod (1979) discussed the negative influence of measurement errors and regression: "Regression means that if scores on two tests are positively correlated, as are intelligence, reading, arithmetic, and spelling scores, then individuals who obtain a particular score on one test will on the average obtain a score nearer to the population average, i.e., regress toward the mean on the other test" (p. 324). Hoffman (1980) suggested that the theoretical problems surrounding regression were not well understood, which led to "considerable uncertainty and possibly confusion among many professionals as to what the data mean at an applied level" (p. 11). If not considered, regression effects lead to increased possibility of misclassification, as pointed out by Thorndike (1963):

If a simple difference between aptitude and achievement standard scores, or a ratio of achievement to aptitude measure, is completed, the high aptitude group will appear primarily to be "underachievers" and the low aptitude group to be "overachievers." For this reason it is necessary to define "underachievement" as discrepancy of actual achievement from the *predicted* value, predicted upon the basis of the regression equation between aptitude and achievement. A failure to recognize this regression effect has rendered questionable, if not meaningless, much of the research in "underachievement" (p. 13).

The questionable reliability associated with some tests used in determining discrepancy almost ensures the presence of regression effects (Coles, 1978; Thurlow & Ysseldyke, 1979). The test validity question is captured in what Kelley (1927) long ago labeled the "jingle and jangle" fallacy—the assumption that tests with the same names measure similar functions, or that tests with different names measure different functions. Hanna, Dyck, and Holen (1979) focused their criticism on the psychometric difficulties associated with age- and grade-equivalent scores. The many associated problems made the expectancy approach a less than optimal means of determining and interpreting a "significant" discrepancy (Davis &

Shepard, 1983). L. R. Wilson, Cone, Busch, and Allee (1983) discussed the incorrect assumption that achievement follows a linear growth pattern which results in an inherent bias when discrepancy is defined as a fraction of some expected achievement value because of different slopes in the patterns.

When used in practice, the expectancy formula approach to discrepancy "yielded strikingly disparate results in terms of the number of children identified as learning disabled by each" (Forness, Sinclair, & Guthrie, 1983, p. 111). In actuality, the resulting prevalence rates ranged from 1% to 37% (Sinclair, Guthrie, & Forness, 1984). Confounding this variability was the additional finding that in a sample of students deemed eligible for LD programs, 64% were not identified by any expectancy formula (Sinclair & Alexson, 1986). Finally, O'Donnell (1980) found that a discrepancy derived from an expectancy formula was not a distinctive characteristic of LD and was equally likely to be found among other students with disabilities.

Discrepancy Score Components

Although discrepancy methods were the object of contentious debate, discrepancy continued to be reinforced as a primary criterion for LD identification (e.g., Chalfant, 1985) mainly because of a desire to reduce the reliance on clinical judgment in LD diagnosis (see Meehl, 1954). Thus, the continued use of discrepancy in the LD diagnostic process required improved methodology.

The first problem requiring attention was related to the types of test scores included in discrepancy formulas. Age-equivalent scores (e.g., MA), for example, lack a consistent unit of measurement. More problematic are grade-equivalent (GE) scores that possess difficulties related to the fact that they ignore both the dispersion of scores about the mean and the nonequivalent regression lines between grade and test scores across both grade levels and content areas (Gullicksen, 1950). Consequently, exact values are difficult to achieve, and GEs, therefore, usually involve an excess of extrapolation, especially at the upper and lower ends of a scale. The difficulties are compounded because scores calculated between testing periods (often 1 year) must be interpolated, but such a calculation is based on the invalid assumption of a constant learning rate. What this means is that achievement tests do not exhibit identical GEs. For example, a seventh grade student who is 2 years below grade level in reading will receive quite different percentile rankings (a possible range of 12 percentile ranks) depending on the reading achievement measure used (Reynolds, 1981). When included in discrepancy formulas, GEs from different tests assessing different academic areas may distort scores that may exaggerate small performance differences (Berk, 1981). The problem of GE comparability is thus significant and, by grade 8, GE scores may possess essentially no meaning (Hoover, 1984).

The problems associated with GEs may be partially remedied by the use of standard scores that hold the advantage of being scaled to a constant mean (M) and SD which permits more accurate and precise interpretation. Nevertheless, Clarizio and Phillips (1986) pointed out the potential limitations with standard scores: (a) no basis for comparisons across grade levels, (b) possible distortions in profile comparisons, and (c) inconsistency of unit size caused by within-grade variability. Although their use provides advantages over GEs, standard scores also need to be interpreted cautiously.

Standard Score Methods

Standard score (SS) discrepancy methods typically involve a direct comparison between common metrics for intellectual ability and academic achievement (Elliot, 1981; Erickson, 1975; Hanna et al., 1979). For LD determination, the standard scores for ability (IQ) and achievement most often have an M = 100 and SD = 15 with the SDL criterion usually being a minimum of 15-point IQ-achievement difference.

Although advancing discrepancy calculation, the SS procedure is not without limitation. One problem surrounds the invalid assumption that, on average, IQ and achievement scores should be identical (e.g., a child with an IQ of 115 should have a reading or math achievement score of 115). This assumption would be true only if IQ and achievement were perfectly correlated (r = 1.00). The actual correlation is about

0.60, which means that the expected achievement for an IQ of 130 is actually 122, not 130. With below-average IQs, an opposite effect occurs (i.e., an IQ of 85 actually has an expected achievement level of about 88). Thus, the SS approach to discrepancy will always possess a systematic bias (Thorndike, 1963). For LD identification, this means the overidentification of high-ability underachievers and the underidentification of low-ability achievers who may in fact be LD.

The less-than-perfect correlation between ability and achievement measures also produces measurement errors that may influence the resulting difference scores. When different IQ and achievement tests are used in calculating discrepancy, the use of particular test combinations will identify more students as LD than will other test combinations (Bishop & Butterworth, 1980; Jenkins & Pany, 1978). The measurement errors also affect the inherent meaning of score comparisons because of the possibility that unique elements may not be measured. Hopkins and Stanley (1981) illustrated the substantial overlapping variance possible between ability and achievement tests. Across grade levels on average, 47% of the variance overlaps, which means that almost half the time the same skills are being measured, making it questionable whether or not "true" differences are being revealed.

Difference Scores

The SS approach produces a difference score that is presumably an index of discrepancy. The difference score, however, often lacks adequate reliability, resulting in uncertainty as to whether or not the difference may have really occurred by chance (Feldt, 1967; Payne & Jones, 1957). For example, the acceptable individual reliabilities of most IQ and achievement tests (about 0.90) produce a difference score with a reliability of only about 0.75. Measurement error is again the primary factor producing this unreliability (see Cronbach, Gleser, Nanda, & Rajaratnam, 1972) which ultimately may distort the discrepancy score as discussed by Thorndike (1963), who concluded that

> if *nothing* but the errors of measurement in the predictor and criterion were operating, we could *still* expect to get a spread of discrepancy scores represented by a standard deviation of half a grade-unit. We would *still* occasionally get discrepancies between predicted and actual reading level of as much as a grade and a half. This degree of "underachievement" would be possible as a result of nothing more than measurement error (p. 9).

Algozzine and Ysseldyke (1981a), using various IQ-achievement test correlations, demonstrated the significantly lower reliabilities of difference scores compared with both of the reliabilities of the tests on which they were based. Using the standard error of measurement (SEM) (a theoretical range around the presumed true score), Schulte and Borich (1984) also demonstrated the unreliability of difference scores. The calculated SEMs of difference scores were substantial and would significantly influence the type and rate of errors made in LD identification. In an empirical analysis, Salvia and Clark (1973) showed how "the standard error of measurement for deficit scores is sufficiently large to preclude rigid adherence to deficits as a criterion for learning disabilities" (p. 308). Reynolds (1981) showed how it is possible to determine the significance of the difference between two scores, but it is a time-consuming process and does not fully answer the question about where to set the cut-off (i.e., criterion) score for LD identification (Schulte & Borich, 1984).

Regression Methods

With SS methods being problematic, alternative means of calculating discrepancy were considered. Shepard (1980) suggested a regression discrepancy method to remedy many of the existing problems. The measurement error associated with IQ and achievement measures ensures that statistical regression will occur, especially when dealing with IQ levels outside of a 95–105 range. The regression method involves calculating equations for IQ and achievement where "The anticipated [expected] achievement score is the norm for children of the same ability, grade level, and sex" (Shepard 1980, p. 80). Measurement error makes a "true" score indeterminate, and its value may be expressed through the SEM, a range surrounding the obtained score. The formula includes the SD of the test and its reliability estimate and is computed from

$$SEM = SD\sqrt{(1\text{-}r_x)}$$

The SEM is then used to calculate a CI that reflects a range within which the "true" score might be found. The formula is

$$CI = x \pm z(SEM)$$

where x is the obtained score and z is the normal curve value corresponding to confidence level (e.g., 95% level = 1.96).

The standard error of estimate (SEE) is a statistic similar to the SEM that is used in the case of two independent scores when one is used to predict the second. Essentially, the SEE places a CI around the predicted score. The formula is

$$SEE = SD\sqrt{(1\text{-}r_{xy}^{\,2})}$$

where SD is the standard deviation of the achievement test and is the squared correlation between IQ and achievement.

Because the correlation between IQ and achievement is not perfect, regression effects will operate (i.e., individuals who obtain an extreme score on one test will, on average, obtain a score closer to the population mean on the second test). The predicted achievement score may be adjusted for regression effects if both IQ and achievement test scores are expressed as SS with M = 100 and SD = 15, and the IQ-achievement correlation is multiplied by the obtained IQ minus the mean of the IQ test, which is then added to the mean of the achievement test (100) as follows:

$$\hat{y} = r_{xy}(IQ\text{-}100) + 100$$

The actual value is computed from the following equation:

$$(\hat{y} - y) > 15z\sqrt{(1\text{-}r_{xy})}$$

which includes (a) measuring IQ, (b) predicting achievement level, (c) measuring actual achievement (y), (d) establishing confidence intervals (CIs) around the predicted achievement score using the SEE, and (e) comparing the predicted and actual achievement scores $(\hat{y} - y)$ using the SEE to determine significant differences. For both IQ and achievement, standard scores (M = 100; SD = 15) are typically used in the formula.

An example of the regression method shows how it is used to determine the presence of a discrepancy. To illustrate, assume a student with a measured IQ of 115. Next, assume an r_{xy} of 0.54 (usually derived from the available research literature). With these values, a predicted achievement score is calculated and found to be 108.1. At the 95% level, the z-value is 1.96, which is used in the equation to obtain a value of 19.93. Using this value, a CI is constructed by adding and subtracting 19.93 from the predicted achievement score of 108.1 to create a CI of 88.17–128.03. If the student's actual achievement score (y) was 85, then it falls below the lower end of the CI (88.17) and a significant discrepancy is said to exist.

The actual calculations may be aided by computer programs that compute significant IQ-achievement discrepancies using a regression approach (e.g., McDermott & Watkins, 1985; Reynolds & Snow, 1985; Watkins & Kush, 1988). The computer programs reduce mathematical error and may be used to create tables for various combinations of IQ and achievement tests as exemplified in the Iowa Regression Discrepancy Tables (Iowa Department of Public Instruction, 1981).

EVALUATION OF REGRESSION METHODS

L. R. Wilson and Cone (1984) argued that the regression discrepancy method provides a "best fit" line for empirically establishing expected achievement values at various IQ levels, and "because regression is a real-world phenomenon, the equation automatically adjusts expected academic scores so that they are less extreme" (p. 99). Evans (1990) discussed six advantages of the regression discrepancy method including (a) determining whether IQ-achievement score differences are due to random error or real, nonchance differences, (b) determining expected achievement score based on individual IQ scores and the correlation between intelligence and achievement, (c) defining discrepancy as the difference between expected and actual achievement score, (d) measuring discrepancy in terms of the SD of the discrepancy difference score, (e) taking into account the SEM of the discrepancy by considering measurement error of IQ and achievement tests, and (f) determining if the discrepancy falls in a predetermined critical ("severe") range when measurement error is considered.

The regression discrepancy method still possesses some practical difficulties, however. Ideally, the regression equation calculated would be based on IQ and achievement scores obtained from large-scale random sampling from the population of interest. Because this is not usually feasible, population statistics for the correlations between individual IQ scores and specific achievement scores, the M and SD of the population IQ, and the M and SD for each specific achievement score must be estimated. With estimated values, the resulting equations may possess errors that limit generalizability [see methods proposed by Woodbury (1963) and McLeod (1979)], but with best estimates and noncontroversial assumptions about linear relationships and normal distributions, "[t]he regression equation approach provides the *best method* for determining academic discrepancy because unlike other approaches, it considers regression, measurement errors, and evidence" (L. R. Wilson & Cone, 1984, p. 107, emphasis added).

Although the regression discrepancy method provides the best answer to the question, "Is there a severe discrepancy between this child's score on the achievement measure and the average achievement score of all other children with the same IQ as this child?" (Reynolds, 1985, p. 40), another practical difficulty remains. A regression equation requires the choice of a value to denote "severity level" but the vagaries surrounding LD make this choice uncertain. The most usual value chosen is two SDs (gleaned from the historical two SDs below the mean IQ level used for the diagnosis of mental retardation [MR]), but while presumably meeting a criterion of "relative infrequency" in the population, the value remains uncertain because of the lack of a true prevalence rate for LD. The uncertainty may produce classification errors of two types: false positive (i.e., identifying a student as LD when he or she is not, in fact, LD) and false negative (i.e., failing to detect real LD). Shepard (1980) suggested that "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 assume that all real instances of LD will be detected" (p. 88).

EVALUATION OF DISCREPANCY METHODS

Cone and Wilson (1981) analyzed the four basic methods of quantifying a discrepancy and concluded that SS and regression equation methods are preferred. This conclusion has been affirmed in other comparative analyses of discrepancy methods (e.g., Bennett & Clarizio, 1988; Braden & Weiss, 1988; Clarizio & Phillips, 1989).

The primary difficulty with regression equation methods is the numerous and complex empirical calculations required that may be further exacerbated by assessment instruments which may not meet acceptable psychometric standards as well as other technical problems (e.g., calculating the correlation between measures or choosing a proper incidence figure). Berk (1984), in an analysis of discrepancy methods, urged caution because of questions surrounding reliability and validity of outcomes. In a similar analysis, Reynolds (1984–1985) validated the use of regression equation models but noted possible confusion in choosing one type of regression equation over another:

> Case *a* will be far too restrictive and is conceptually illogical in several regards: It will create a more homogeneous group of children; however, LD is characterized by the individuality of the learner, not traits or characteristics held in common with other children. Objections to application of case *b* are less conceptual than mathematical. Routinely applying both models and accepting qualification by either introduces a significantly greater probability of finding a severe discrepancy when none actually exists than does the application of either model.…Using both models with all children will then not aid in reducing the conceptual confusion in the field as might application of a uniform model (p. 465).

Even the most defensible method of discrepancy calculation (i.e., SS and regression equation) remains less than perfect with respect to optimal psychometric and statistical considerations. The problems are exacerbated by the many different measurement models that might be employed (see Willson & Reynolds, 1984–1985) and the curious situation involving the fact that as these models become more defensible statistically, they become more complicated to use in practice (Boodoo, 1984–1985). Consequently, actual diagnostic practice in the LD field lags behind state-of-the-art statistical models, which almost makes discrepancy "an atheoretical, psychologically uninformed solution to the problem of LD classification" (Willson, 1987, p. 28).

Practical Difficulties

The technical problems create real-world difficulties. Ross (1992a, b), in a survey of school psychologists, found that fewer than 10% were able to correctly evaluate whether four sets of ability-achievement scores reflected chance measurement differences or reliable, nonchance differences. Barnett and Macmann (1992) attributed much of the inaccuracy in discrepancy interpretation to basic misunderstandings surrounding test interpretation; statistical significance, confidence intervals, and measurement error. For example, Macmann, Barnett, Lombard, Belton-Kocher, and Sharpe (1989) found classification agreement rates ranging from 0.57 to 0.86 with different discrepancy calculation methods. When different achievement measures were used in the same calculations, however, the classification agreement rates fell to a range of 0.19 to 0.47. When both ability and achievement measures varied, agreement rates were consistently below 0.25 (Clarizio & Bennett, 1987). Thus, on average, only about 1 in 4 students deemed to possess a "severe" discrepancy would be identified as such with different sets of ability and achievement test scores. Macmann and Barnett (1985) affirmed this finding in a computer simulation study that concluded that "the identification of a severe discrepancy between predicted and actual achievement was disproportionately related to chance and instrument selection" (p. 371). The consequences become even more problematic in cases where more than one achievement test was administered, and the lowest score among them was used in discrepancy calculation. Sobel and Kelemen (1984) showed how this situation will likely result in a difference between the proportion of students actually classified LD and the proportion originally expected. For example, in the case of three achievement measures administered and the lowest score selected, the original criterion value of 6.68% LD cases identified would increase to 12.2%.

Instability of Discrepancy Scores

The inherent variability associated with discrepancy calculation is made more unsure by findings showing instability in discrepancy scores over time. O'Shea and Valcante (1986) found that SDL comparisons between groups of students with LD and low achieving students without LD differed significantly from grade 2 to grade 5. The groups appeared to develop diverging SDLs over time with increasingly larger differences for students with LD in language and mathematics compared to reading but, nevertheless, the SDL for reading doubled from grade 2 to grade 5. White and Wigle (1986), in a large-scale evaluation of school-identified students with LD, found four different patterns of discrepancy over time. The largest group (40%) revealed no ability-achievement discrepancy at initial placement or at reevaluation. The next largest groups demonstrated either a pattern of being discrepant at placement but not at reevaluation or, conversely, a pattern of not being discrepant at initial placement but discrepant at reevaluation. The smallest group showed a discrepancy at both placement and reevaluation. Considering that discrepancy is a primary identification criterion for LD, its instability over time is a source of concern, but the problem appears endemic. For example, Shepard and Smith (1983) reported that only 43% of a statewide sample of school-identified students with LD met strict identification criteria, with discrepancy being the primary criterion.

An early survey of 3,000 students with LD in Child Service Demonstration Centers showed that the average discrepancy was only about 1 year, leading to the conclusion that "[t]his discrepancy can be interpreted as a moderate retardation, rather than a severe disability" (Kirk & Elkins, 1975, p. 34). In a later similar analysis, Norman and Zigmond (1980) applied the federal (1976) SDL formula and found that, on average, 47% of students met the SDL criterion. For children aged 6 to 10 years (the likely age range of identification), less than 40% met the SDL criterion while the percentage for students aged 15 to 17 was 68%. Although providing greater confidence in the LD classification of the older children, the smaller percentage of younger children meeting the SDL criterion raises questions about the validity of their LD classification.

Discrepancy and the Identification of Learning Disability

Shepard and Smith (1983) suggested that "the validity of LD identification cannot be reduced to simplistic statistical rules" (p. 125), but the inconsistent application of existing criteria creates significant difficulties in the LD diagnostic process. Shepard, Smith, & Vojir (1983), using a "discrepancy criterion," found that 26% of identified students with LD in Colorado revealed no discrepancy while 30% revealed a significant discrepancy with the use of any reading or math test. When validated with a second achievement measure, 5% of all students with LD had a significant discrepancy on two math tests while 27% revealed a significant discrepancy on two reading tests. Thus, not only was the discrepancy criterion not validated, but a "below grade level" criterion was not affirmed either; "Many LD pupils were not achieving below grade level as measured by standardized tests" (p. 317).

In contrast, Cone, Wilson, Bradley, and Reese (1985) found that 75% of a school-identified LD population in Iowa met the required discrepancy criterion. As this LD population continued in school, achievement levels became increasingly discrepant. In a later analysis, L. R. Wilson, Cone, Bradley, and Reese (1986) found that the identified students with LD were clearly different from other students with mild disabilities in Iowa (e.g., MR and behavior disorders [BDs]): "The main factor providing differentiations was discrepancy between achievement and ability" (p. 556). They concluded that students with LD were primarily underachievers, not simply low achievers.

In a later analysis, Valus (1986a) found 64% of identified students with LD to be significantly underachieving. In a large-scale analysis of Iowa's LD population, Kavale and Reese (1992) found that 55% met the discrepancy criterion. In different locales, the percentage of students with LD meeting the discrepancy criterion ranged from 32% to 75%. Thus, in any LD population, there will be a significant proportion who do not meet a significant discrepancy criterion, and, because of possible differences in interpretation, considerable variability in the proportions that do meet the discrepancy criterion across settings.

The finding of significant inconsistencies about the percentage of students meeting the discrepancy criterion is common among studies analyzing identified LD populations. For example, McLeskey (1989) found that 64% of an Indiana LD population met the discrepancy criterion, but this figure was achieved only after more rigorous and stringent state guidelines for LD identification were implemented. The 64% figure was almost double the 33% found in an earlier study (McLeskey & Waldron, 1991). In general, about one third of identified LD samples have been found not to meet the stipulated discrepancy criterion (e.g., Bennett & Clarizio, 1988; Dangel & Ensminger, 1988; Furlong, 1988).

Statistical Classification vs. Clinical Judgment

Shepard and Smith (1983) referred to the approximately one third of identified students with LD as "clinical cases," meaning that their eligibility was a discretionary judgment made by a multidisciplinary team (MDT) which was at variance with the statistical (i.e., discrepancy) information. This situation may occur because (a) the LD may have caused ability level (i.e., IQ) to decline, and if achievement remained at a comparatively low level, then a discrepancy would not exist; (b) intact skills permitted the student to "compensate" for the effects of LD which means that achievement test scores may reveal an increase while ability level remained constant; or (c) a "mild" discrepancy was present but not unexplained because of factors such as limited school experience, poor instructional history, behavior problems, or second-language considerations. The essential question: Are such students "truly" LD, or is the inconsistency between team decisions and statistical status "truly" misclassification?

The many vagaries associated with "system identification" (Morrison, MacMillan, & Kavale, 1985) are the primary reason for the difficulty in decisions about the presence or absence of LD (Frame, Clarizio, Porter, & Vinsonhaler, 1982). In analyses of MDT decisions, it appears that LD identification criteria, especially the primary criterion of severe discrepancy, were neither rigorously or consistently applied (Epps, McGue, & Ysseldyke, 1982; Furlong & Yanagida, 1985; Furlong & Feldman, 1992). The difficulties begin with the lack of uniformity across educational agencies in setting "severe" discrepancy criterion levels (Perlmutter & Parus, 1983; Thurlow & Ysseldyke, 1979) which are then often exacerbated by differences in interpreting existing guidelines (Thurlow, Ysseldyke, & Casey, 1984; Valus, 1986b). The misapplication of criteria in LD identification procedures is further complicated by external pressures that might include the desire of MDTs to provide special education services, the request of general education teachers to remove difficult-to-teach students, and parental demands for LD placement (e.g., Algozzine & Ysseldyke, 1981b; Sabatino & Miller, 1980; Ysseldyke, Christenson, Pianta, & Algozzine, 1983).

When LD is viewed as primarily a socially constructed disability (Gelzheiser, 1987), the many external pressures often become primary considerations because a criterion like SDL is viewed as too child-centered in a medical model sense and does not permit examination of complex contextual interactions presumed relevant for valid diagnosis (Algozzine & Ysseldyke, 1987). Gerber and Semmel (1984) even argued that an instructional perspective rather than a statistical one should be the basis for determining LD eligibility. They suggested that the teacher become the "test" for determining whether a student has a "real" learning problem. Under such circumstances, it is not surprising to find that MDTs often do not "bother with the data" (Ysseldyke, Algozzine, Richey, & Graden, 1982).

The "clinical cases" of LD represent, at best, a "functional" LD because even though deemed eligible, the students in question really did not meet stipulated identification criteria with discrepancy often being the most tangible. The failure to meet stipulated criteria, however, raised serious questions about the reliability and validity of "clinical diagnoses" of LD (Shepard, 1983). It was, therefore, not surprising to find that judges were not able to differentiate students with LD based solely on an examination of test scores (Epps, Ysseldyke, & McGue, 1984).

VAGARIES OF IDENTIFICATION AND PREVALENCE

When LD determination is not based on the application of strict criteria, the diagnostic process may be likened to the U.S. Supreme Court's definition of pornography: "I know it when I see it." The lack of rigor in the diagnostic process has led to an accelerated rate of LD identification and LD becoming, by a wide margin, the largest category in special education. Presently, LD accounts for more than 50% of all students with disabilities and more than 5% of all students in school (U.S. Department of Education, 1999). In commenting on the magnitude of the increase in LD prevalence, MacMillan, Gresham, Siperstein, and Bocian (1996) suggested that "Were these epidemic-like figures interpreted by the Center for Disease Control, one might reasonably expect to find a quarantine imposed on the public schools of America" (p. 169). There is little justification for such numbers, and the problem is compounded by the lack of consistency in the way the LD population is distributed across settings (Kavale & Forness, 1995). Clearly, fewer students are identified as LD when a strict discrepancy criterion is implemented rigorously (e.g., Finlan, 1992), but external factors (e.g., financial resources) may significantly influence (and increase) the number of students identified as LD (Noel & Fuller, 1985). Forness (1985) showed how state special education policy changes in California significantly affected the number of students identified in the high-incidence mild disability categories. LD saw a 156% gain compared to the 104% gain nationally, and a comparison with concomitant losses for MR and BD led Forness to the conclusion "that California's relatively dramatic increase in children identified as learning disabled may be at the expense of two other related categories" (p. 41). Such state disparities were not uncommon and led to the conclusion that "Our results suggest that variation in LD diagnostic levels across states is significantly related to distinctions in diagnostic practice as well as or instead of actual disease prevalence" (Lester & Kelman, 1997, p. 605). In contrast, far greater consistency in classification rates has been found for hearing impairment and physical/multiple disability compared to LD (Singer, Palfrey, Butter, & Walker, 1989).

Confounding Among High-Incidence Mild Disabilities

The confounding among high-incidence mild disabilities appears to be primarily between LD and MR. MacMillan et al. (1996) found among 150 referred students 43 with IQ levels at 75 or below. Of the 43, only 6 were classified MR even though they met the requisite eligibility cut-off score, while 18 were classified LD primarily because the LD label was viewed as a more acceptable designation. Similarly, Gottlieb, Alter, Gottlieb, and Wishner (1994) found that an urban LD sample possessed a mean IQ level that was 1 SD lower than a suburban comparison sample. They concluded that "These children today are classified as learning disabled when in fact most are not" (p. 463). This view was affirmed by MacMillan, Gresham, _and Bocian (1998) who found that out of 61 students classified LD by schools, only 29 met the required discrepancy criterion. In analyzing the results, they remarked that "We did not anticipate the extent to which the process would yield children certified as LD who failed to meet the discrepancy required by the education code" (p. 322). Thus, even though discrepancy remains the primary (and sometimes sole) criterion for LD identification, it was often ignored in actual practice. Gottlieb et al. (1994) suggested "the discrepancy that should be studied most intensively is between the definition of learning disability mandated by regulation and the definition employed on a day-to-day basis in urban schools" (p. 455).

Because "public school practices for diagnosing children with LD bear little resemblance to what is prescribed in federal and state regulations (i.e., administrative definitions) defining LD" (MacMillan et al., 1998, p. 323), the LD population has become increasingly heterogeneous and the longstanding "problem of heterogeneity" firmly entrenched (Gallagher, 1986). For example, Gordon, Lewandowski, and Keiser (1999) analyzed the problems associated with the LD label for "relatively well functioning" students. By failing to rigorously adhere to a SDL criterion, students with LD may not demonstrate underachievement, a primary LD feature (Algozzine, Ysseldyke, & Shinn, 1982) which then makes the utility of the LD category open to question (Epps et al., 1984).

Confounding Between Learning Disability and Low Achievement

The vagaries of LD classification, especially the inability to differentiate LD and low achievement (LA), have been demonstrated in studies conducted by the University of Minnesota Institute for Research on Learning Disabilities (Minnesota studies). Ysseldyke, Algozzine, and Epps (1983) analyzed psychometric data obtained from students without LD using 17 operational definitions of LD. For 248 cases, 85% met the requirements for one operational definition of LD, while 68% qualified with two or more operational definitions. Only 37% of the non-LD sample did not meet the criteria specified in any of the 17 operational definitions of LD. A second analysis examined data for students with LD and students with LA to determine how many would qualify with each of the 17 operational definitions of LD used earlier. For the LD group, 1% to 78% were classified LD with each definition while the LA group was also classified LD from 0% to 71% of the time using each operational definition. Further analysis showed that 4% of the LD group was not classified by any of the 17 operational definitions while 88% of the LA group qualified as LD by using at least one operational definition. In a similar investigation, Epps, Ysseldyke, and Algozzine

(1983) examined the number of students identified as LD with each of 14 operational definitions that emphasized the discrepancy criterion. The definitions classified from 7% to 81% of students as LD, whereas 5% to 70% of a non-LD group were also classified LD using at least one of these 14 operational definitions. To determine the congruence among the 14 operational definitions, Epps, Ysseldyke, and Algozzine (1985) performed a factor analysis and found two factors. The first factor (I) emphasized LA whereas the second factor (II) was represented by discrepancy. In terms of their respective weights, Factor I accounted for 70% of the variance compared to 16% for Factor II. The difference in explained variance led to the conclusion that LD might be properly conceptualized as a category reflecting LA, rather than discrepancy.

Epps et al. (1985) also found that knowing how many LD definitions qualified a student provided little assistance in correctly predicting group membership (LD vs. LA). Algozzine and Ysseldyke (1983) also found considerable inaccuracy in decisions about group membership (LD vs. LA) and concluded that "To make classification dependent on these discrepancies seems somehow arbitrary and capricious" (p. 245). Consequently, discrepancy appeared to possess limited value, and suggestions about its worth as a criterion for LD identification possessed little merit because "there may be an equally large number of children exhibiting similar degrees of school achievement not commensurate with their measured ability who are not categorized and therefore are *not* receiving special education services even though they are eligible for them under the current conceptual scheme represented by the category of learning disabilities" (p. 246). Thus, the failure to make LD a classification predicated on discrepancy suggests that it has not been possible to unequivocally define a category different from LA, and it might be more appropriate to recognize LA as the major problem.

The Minnesota studies appeared to support the view that reliance upon a discrepancy criterion for LD identification may not be defensible because it does not provide a clear distinction between LD and LA. L. R. Wilson (1985), however, challenged the idea that the LD category should be eliminated in favor of a more general classification like LA because a more general category will do little to eliminate the ambiguities and inconsistencies associated with LD. In fact, the Minnesota studies may themselves possess ambiguities and inconsistencies that limit the findings. For example, the Minnesota studies used only a discrepancy criterion for LD identification, and failed to include other components of the federal definition such as the exclusion which "states that the academic deficit cannot be the result of other possible causes such as emotional and personality factors, cultural deprivation, impaired sensory acuity, or educational deprivation" (p. 45). Since this aspect of the federal definition was not applied, the identification process was necessarily incomplete and restricted.

The other major problem area was related to sampling, specifically the possibility of bias in the Minnesota samples. The final sample used in the Minnesota studies was selected from a much larger population, which raised the question, "Is there evidence to suggest that the selection was random or is there reason to believe that bias may have distorted the findings?" (L. R. Wilson, 1985, p. 45). With respect to the LA group, L. R. Wilson suggested that "there is good reason to suspect that selection factors may have produced a disproportionately large number of discrepant achievers in the group of low achievers who were not formally labeled as learning disabled" (p. 46). Finally, the restricted nature of the selected samples raised questions about the generalizibility of the Minnesota findings.

In an analysis of a large-scale Iowa sample, L. R. Wilson (1985) demonstrated "that the federal definition of learning disabilities can be successfully used, that it can be consistently applied by a large group of special education professionals, that the various components of currently accepted learning disability definitions can provide the basis for discriminating a reasonably unique group of children, and that the exceptions found in this study, and other similar ones, do not automatically invalidate the previous conclusions" (pp. 49–50). The application of both a discrepancy *and* exclusion criterion resulted in a sound foundation for LD classification. As a result, the LD concept was quite defensible, and it would be "premature to eliminate it in favor of other concepts that probably have the very same weaknesses" (p. 51). In response, Algozzine (1985) suggested that there was really no reprieve for the LD concept and again LD was suggested to be a less than viable special education category because "creating the *new* concept of learning disabilities has not reduced the ambiguities, inconsistencies, and inadequacies that existed when low achievement was not a separate diagnostic category" (p. 75).

Learning Disability vs. Low Achievement Debate

The continuing debate about the LD-LA distinction began to erode the integrity of LD. Longstanding critiques of the LD definition (e.g., Reger, 1979; Senf, 1977; E. Siegel, 1968) evolved into suggestions that LD really did not exist as an independent entity as well as its depiction as myth (McKnight, 1982), questionable construct (Klatt, 1991), or imaginary disease (Finlan, 1994). The assumption that LD and LA could not be reliably distinguished became conventional wisdom. The primary evidence came from a study by Ysseldyke, Algozzine, Shinn, and McGue (1982) showing a substantial degree of overlap between the test scores of LD and LA groups and a conclusion raising "serious concerns regarding the differential classification of poorly achieving students as either LD or non-LD" (p. 82). Further confirmation was found in a study by B. A. Shaywitz, Fletcher, Holahan, and Shaywitz (1992) who concluded that "Our findings suggest more similarities than differences between the reading disabled groups" (p. 646). Group membership in this case was defined with a discrepancy criterion (LD) or low achievement (LA) criterion (scoring below 25th percentile in reading). When the LD and LA groups were compared across a number of child-, teacher-, and parent-based measures, few differences were found, with the major exception being in the ability (i.e., IQ) area. Nearly all the variance between groups was accounted for by IQ, but this may only be a reflection of the way groups were defined.

The findings from these studies have had significant impact and have been reported with remarkable consistency. For example, the Ysseldyke, Algozzine, Shinn, & McGue (1982) study has been used to conclude that limited LD-LA differences existed as exemplified in the following statements gleaned from the literature:

a. Certain researchers have suggested that LD is largely a category for low-achieving children.
b. [Ysseldyke et al.] found few psychometric differences between groups of students identified as learning disabled (LD) and low achievers who did not carry the label.
c. Recent studies of children diagnosed as learning disabled have shown that many such children...are virtually indistinguishable from low-achieving non-handicapped peers.

The difficulties in differentiating LD and LA groups were based on the Ysseldyke, Algozzine, Shinn, & McGue (1982) findings of a large number of identical scores between LD and LA subjects as well as a high percentage of overlap between scores. For example, on the Woodcock-Johnson Psychoeducational Battery, LD and LA groups showed identical scores 33 out of 49 times and an average overlap percentage of 95%. On five other psychoeducational measures, in better than half the cases there were identical scores and a 96% percentage of overlap. These metrics appeared, however, to be at variance with the reported statistical analyses. A comparison of Woodcock-Johnson scores revealed "that on average the LD group performed significantly poorer on 10 of the subtests" (p. 98), while statistical comparison of the five other psychoeducational measures showed "that the mean level performance of the LD children was lower on many of the measures, particularly the PIAT [Peabody Individual Achievement Test], and at times was significantly less than the mean level of their low-achieving peers" (p. 79).

Reanalysis of the Minnesota Studies

Kavale, Fuchs, and Scruggs (1994) reexamined the Minnesota studies using quantitative synthesis methods (meta-analysis) and demonstrated how the percentage of overlap metric used by Ysseldyke, Algozzine, Shinn, & McGue (1982) may have masked real performance differences. The overlap metric used in the Minnesota studies was calculated by using the range of scores found for one group and then comparing how many cases from the second group fell within that same range, but with such a methodology, "[t]he potential bias toward overlap is high because the comparison is based on the variability demonstrated by only one group with the other being forced into that distribution without regard to the characteristics of its own variability" (Kavale et al., 1994, p. 74). The effect size (ES) statistic used in meta-analysis, because it is a standard score (z-score), eliminates potential bias by representing the extent to which groups can be differentiated, or, conversely, the degree of group overlap. For example, an ES of 1.00 indicates that the two compared groups differed by 1 SD and that 84% of one group can be clearly differentiated from the other group with a 16% group overlap.

Using the data from the Ysseldyke, Algozzine, Shinn, & McGue (1982) study, Kavale et al. (1994) calculated ES's for 44 comparisons and found an average ES of 0.338. This means that, on average, it would be possible to reliably differentiate 63% of the LD group. Conversely, 37% could not be differentiated, and this represented the degree of overlap that was substantially less than the average 95% reported by Ysseldyke, Algozzine, Shinn, & McGue (1982). For the Woodcock-Johnson Cognitive Ability subtests, an average ES of 0.304 was found, while the Achievement subtests provided an average ES of 0.763. With little reason to expect cognitive (IQ) differences between LD and LA groups, the modest group differentiation was not surprising. On the other hand, almost 8 out of 10 members of the LD group scored at a level that made it possible to discern clear achievement differences when compared with the LA group members. Similar findings emerged with other cognitive and achievement tests. For example, Wechsler Intelligence Scale for Children–Revised (WISC-R) comparisons revealed an average ES of 0.141 (56% level of group differentiation) while PIAT comparisons showed an average ES of 1.14, indicating that in almost 9 out of 10 cases (87%), the LD group performance was substantially below the LA group. Consequently, "it appears that the lower achievement scores of the LD group are of a magnitude that distinguishes them from their LA counterparts" (Kavale et al., 1994, pp. 74–75).

Algozzine, Ysseldyke, and McGue (1995) contested the meta-analytic findings but agreed that students with LD may be the lowest of the low achievers. They suggested that the difficulty was in interpreting the meaning of that status: "Where we part company is in the inference that because students with LD may be the lowest of a school's low achievers, they necessarily represent a group of people with qualitatively different needs who require qualitatively different instruction" (pp. 143–144). What Algozzine et al. (1995) failed to consider, however, were the findings showing minimal group differentiation in the cognitive domain. With essentially no difference in ability but large differences in achievement, the LD group demonstrated "significant discrepancy" that was not shown by the LA group. Consequently, Kavale (1995) suggested that the LD and LA groups "represent two distinct populations. Because the LD group are lower on achievement dimensions but not on ability, they are, in addition to being the lowest of the low achievers, a different population defined by an ability-achievement distinction represented in a different achievement distribution but not in a different ability distribution" (p. 146).

Examining Learning Disability and Low Achievement Samples

In a similar comparison of LD and LA groups that also included comparisons with an MR group defined as IQ < 75, Gresham, MacMillan, and Bocian (1996) found an average LD-LA level of differentiation of 61% (compare with the 63% reported by Kavale et al. 1994). The differentiation levels for LD-MR and LA-MR averaged 68.5% and 67.5%, respectively. On achievement measures, LD-LA group comparisons revealed an average ES of 0.39 indicating a 65% level of differentiation and confirmation of the finding that "LD children performed more poorly in academic achievement than LA children" (p. 579). The LD group performed most poorly in reading, where almost 3 out of 4 students with LD could be reliably differentiated from LA students. The large achievement differences in reading between LD and LA groups were affirmed by Fuchs, Fuchs, Mathes, and Lipsey (2000) who found that 72% of the LA group performed better in reading than the LD group . Even larger ES differences were found with more rigorous measures, "suggest[ing] that researchers and school personnel in fact do identify as LD those children who have appreciably more severe reading problems compared to other low-performing students who go unidentified" (p. 95).

Gresham, MacMillan & Bocian (1996) also investigated cognitive ability (IQ) differences among the three groups. As expected, 94% of the LD group could be reliably differentiated from the MR group. The percentage fell to 73% in differentiating LD and LA groups, suggesting greater cognitive ability overlap between these two groups. Gresham et al., however, included an LA group defined differently from both the Ysseldyke, Algozzine, Shinn, & McGue (1982) and B. A. Shaywitz et al. (1992) studies: "Our LA group was closer to what might be considered a 'slow learner' group on the basis of their average-level intellectual functioning relative to the LA groups in [the other] studies" (p. 579). The result was that even though achievement was depressed, it was not discrepant when compared to IQ level. In contrast, the LD group revealed significant discrepancies and was thus properly classified because "Children with LD perform more poorly in reading than LA children, even when the former group has higher cognitive ability" (p. 580). This finding has been confirmed by Short, Feagans, McKinney, and Appelbaum (1986) in an analysis of LD subtypes. In examining reading achievement across five groups, they found that "the joint application of IQ- and age-discrepancy criteria appeared to be useful for distinguishing between seriously disabled students and those who might be more appropriately classified as slow learners or underachievers" (p. 223). In summary, Gresham, MacMillan, & Bocian (1996) concluded that LD, LA, and MR groups "could be reliably differentiated using measures of cognitive ability and tested academic achievement" (p. 580). When LD is defined with an ability-achievement difference criterion, the resulting discrepancy appears to be an appropriate metric that permits reliable differentiation between LD and LA groups.

Learning Disability and Intelligence

Although empirical evidence appeared to indicate that LD and LA could be reliably differentiated with a discrepancy criterion, questions about its use continued. One form of questioning focused on IQ and whether it was necessary in defining LD. Beginning with the finding that IQ was not useful in locating students with reading disability (L. S. Siegel, 1988), questions arose about whether or not IQ was a necessary component in definitions of LD (L. S. Siegel, 1989, 1990). A major problem surrounded IQ tests and what they presumably measure. Stanovich (1991b) concluded that "an IQ test score is not properly interpreted as a measure of a person's potential" (p. 10). Yet, "the LD field has displayed a remarkable propensity to latch onto concepts that are tenuous and controversial….The LD field seems addicted to living dangerously" (Stanovich, 1989, p. 487). At a practical level, for example, there was controversy about what type of IQ score should be used in discrepancy calculation. Although it was commonly recommended that performance or nonverbal IQ be used (e.g., Stanovich, 1986a; Thomson, 1982), an equally compelling case could be made for the use of verbal IQ (e.g., Hessler, 1987). Without resolution about what IQ tests actually measure, "IQ is a superordinate construct for classifying a child as reading disabled. Without clear conception of the construct of intelligence, the notion of a reading disability, as currently defined, dissolves into incoherence" (Stanovich, 1991a, p. 272).

The ability-achievement discrepancy criterion treats intelligence and achievement as separate and independent variables, but L. S. Siegel (1989) suggested that this may not be valid because "A lower IQ score may be a consequence of the learning disability, and IQ scores may underestimate the real intelligence of the individual with LD" (p. 471). Further confounding was introduced by findings that the IQ of students with LD may actually decline over time (Share & Silva, 1987; Van den Bos, 1989). If this is a valid finding and also assuming that students remain close to their original reading levels over time (see Juel, 1988), then discrepancies should increase over time, but McLeskey (1992) found a negative association between discrepancy level and CA where "students in the elementary grades were most likely to manifest a severe discrepancy between expected and actual achievement, while high school students were least likely to have such a discrepancy" (p. 18).

A partial explanation may be found in what Stanovich (1986b) termed the "Matthew effect," referring to the Biblical statement (Matthew 13:12) that suggests that each advantage leads to further advantage, or conversely, initial disadvantage multiplies into even more disadvantage. For reading, this means that the poor get poorer: "Children with inadequate vocabularies—who read slowly and without enjoyment—read less, and as a result have slower development of vocabulary knowledge, which inhibits further growth in reading ability" (p. 381). B. A. Shaywitz et al. (1995), however, found no evidence of a Matthew effect in reading but a modest Matthew effect for IQ in a large-scale LD sample. For both IQ and reading, however, "the influence of the regression-to-the-mean effect tends to mask the relatively small Matthew effect" (p. 902) which suggests that the presumed cumulative disadvantage (Matthew effect) really refers to the rate of gain or loss in reading ability compared to initial level (see Walberg & Tsai, 1983). There are thus complex reciprocal relationships between reading ability and cognitive skills that are seen to confound the discrepancy notion because "the logic of the learning disabilities field has incorrectly assigned all the causal power to IQ. That is, it is reading that is considered discrepant from IQ rather than IQ that is discrepant from reading" (Stanovich, 1991b, p. 275).

The Role of Intelligence in Definitions

The problem of confounding is most likely to arise in situations where concepts are defined with dual criteria. For example, although the psychometric characteristic IQ has long defined MR (e.g., Hollingworth, 1926), there was a later decision to include a second criterion in the form of adaptive behavior: the effectiveness and degree to which individuals meet standards of self-sufficiency and social responsibility (Heber, 1959). There was, however, concern over the inclusion of adaptive behavior in the MR definition primarily because of measurement issues (Clausen, 1972; MacMillan & Jones, 1972). Specifically, there were no adequate instruments to evaluate adaptive behavior that made it a psychometric characteristic comparable to IQ. [Of course, this situation was remedied with instruments like the American Association on Mental Retardation (AAMD) Adaptive Behavior Scale and the Vineland Social Maturity Scale]. With only one measure acceptable, there would be no means to evaluate *both* criteria, and this situation would create the possibility of students identified as MR who did not meet the dual criteria definition as well as students not identified who would meet the definition if appropriate assessments for *both* criteria were available.

When reliable and valid assessments are not available, clinical judgment was likely substituted but was often equally unreliable, especially in the "milder" regions of MR. With significant impairment in intellectual ability (IQ < 50), the corresponding adaptive behavior was probably equally impaired and not difficult to judge. As the upper limit of the IQ criterion was approached (IQ 70–75), however, the probability that adaptive behavior would correspond similarly decreased and clinical judgment became more problematic.

In defining LD, Kavale and Forness (1985) recommended a dual criteria definition similar to MR that included (a) significant subaverage academic impairment and (b) IQ in the average range. The advantage would be that both criteria can be reliably measured and little clinical judgment would be necessary. The two criteria can be readily compared and decision rules adopted to determine when the obtained difference ("discrepancy") was significant. If an additional exclusion criterion was added, then the identification process would avoid the myriad difficulties surrounding attempts to include other definitional parameters (e.g., psychological process deficits, central nervous system dysfunction) that cannot be reliably assessed. For this reason, IQ remains an important component in LD definition.

Defining Learning Disability Without Intelligence

Even though IQ should be considered a necessary criterion, L. S. Siegel (1989) suggested that the LD field "abandon the use of the IQ test in the definition of learning disabilities.…[T]he IQ-achievement deviation definition should be abandoned because of its illogical nature" (p. 477). Stanovich (1989) suggested, however, that such a position might be "too extreme" (p. 489) and "perhaps ends up saying too little about too much" (p. 490). Lyon (1989) concluded that "Siegel has raised some interesting and compelling issues but has confounded her position by taking a narrow conceptual and methodological stance in addressing the relationship between intelligence and the LD definition" (p. 506). Baldwin and Vaughn (1989) suggested that "Siegel's position might be illogical because the reasoning was convoluted and misleading" (p. 513).

Meyen (1989) objected to the suggestion that IQ should be eliminated in the LD definition because "challenging the use of intelligence measures in defining learning disabilities, in essence, questions the efficacy of the category of learning disabilities itself as a means to identify students who warrant special education services" (p. 482). By eliminating IQ, a situation would be created where "we would largely serve low achievers and have no basis for determining whether or not a student is achieving at a reasonable level given his or her ability" (p. 482). The result would be an even more contentious LD-LA differentiation debate. The situation would not be remedied with a different IQ cut-off score which L. S. Siegel (1989) suggested as an alternative solution. In applying the discrepancy criterion in LD identification, there has long been the implicit assumption that IQ is at an average or above level in order to "discriminate between poor achievement that is expected (that is, on the basis of intellectual ability or sensory handicaps) and poor achievement that is not expected (that is, the probable presence of LD)" (Scruggs, 1987, p. 22). With an IQ cutoff of, for example, 75 (a level closer to the MR criterion) less than average academic achievement would be neither unexpected nor unexplained. There may be a need for special education, but such a student would not be properly classified as LD.

The primary difficulty with a lower IQ cutoff score in defining LD would be the potential confounding with MR. The AAMD (see Grossman, 1973) shifted the upper cutoff score for MR from –1 to –2 SD, that is, an IQ level of 70 instead of 85. Grossman (1983) later suggested the IQ cutoff could be as high as 75 since IQ should be viewed as only a rough guideline. Thus, cutoff scores really represent arbitrary statistical decisions rather than being based on scientific classification procedures (Zigler & Hodapp, 1986). Such arbitrary decisions create real dilemmas because they cause widely varying prevalence rates. For example, Reschly (1992) demonstrated that the use of an IQ cutoff of 75 and below results in twice as many individuals potentially eligible than would using IQ 70 and below. In addition, more cases fall in the interval 71–75 than in the entire range associated with mild MR (IQ 55–70). For LD with a 75 IQ cutoff, an additional 22.5% of the population would be eligible (given an "average" IQ level arbitrarily defined at 92.5) with perhaps 3% of this group potentially eligible for either MR or LD. With a discrepancy criterion, eligibility for LD can also be defined in SD units similar to MR (–1 to –2 SD depression) (see Mercer et al., 1996). As with MR, however, the choice of criterion level remains arbitrary and will also affect prevalence: the smaller the required discrepancy, the larger the prevalence. The current high prevalence rate for LD suggests a decision including smaller discrepancy levels, but the resulting LD classifications also suggest an increased probability of confounding with MR.

The consequences of the confounding between LD and MR are seen in large variations across states in prevalence rates with the typical outcome being more LD and less MR than expected (U.S. Department of Education, 1999). Gresham et al. (1996) showed that the percentage of students classified as MR was inversely related to the percentage of students classified as LD ($r = -0.24$). Thus, states serving a small percentage of students with MR classify a larger percentage of students as LD, and vice versa. It is entirely possible then that students with similar cognitive abilities and disabilities are served in one state as LD and in another as MR (MacMillan, Siperstein, & Gresham, 1996).

Although average or above IQ has been considered a prerequisite for LD, a longstanding view holds that average or above intelligence is not a necessary or desirable criterion (e.g., Ames, 1968; Belmont & Belmont, 1980; Cruickshank, 1977). Support for this view was found in large-scale evaluations of LD populations that have found mean IQ levels in both the low average (IQ 80–90) range (e.g., Koppitz, 1971; Smith, Coleman, Dokecki, & Davis, 1977; J. D. Wilson & Spangler, 1974) and the lower regions of the average (IQ 90–100) range (e.g., Kirk & Elkins, 1975; McLeskey & Waldron, 1990; Norman & Zigmond, 1980). In addition, IQ levels of students with LD tended to be quite variable, and anywhere from 10% to 40% of LD samples were found to have IQ scores falling below 85 (e.g., Gajar, 1980; Kavale & Reese, 1992; Shepard et al., 1983).

To explain why the actual IQ level of students with LD might be below average, Burns (1984) used the bivariate normal distribution to show how LD samples can have average IQ scores well below 100. With the known relationship between IQ and achievement, the average IQ of LD samples will decrease as the correlation between IQ and achievement increases. For example, if cases below a given cutoff for achievement (e.g., $z < -1.0$) and above a given IQ cutoff (e.g., IQ > 80) are considered while postulating a correlation of 0.50 between IQ and achievement, then the average IQ of a sample on the bivariate normal distribution will be about 93. Piotrowski and Siegel (1986), however, suggested that using the bivariate normal distribution to explain mean IQ levels less than 100 for LD samples may not be appropriate. The primary difficulty was found in the use of *fixed* achievement cutoff scores regardless of IQ score, as achievement is likely to vary as a function of both MA and CA. For example, a student with an IQ of 80 and achievement z-score of –0.05 would meet the LD discrepancy criterion under the bivariate normal distribution, but, in reality, demonstrate almost no underachievement. Conversely, a student with an IQ of 130 and achievement z-score of –0.95 would in fact be underachieving significantly but would not meet the discrepancy criterion for LD. These problems are compounded further as the correlation between IQ and achievement increases. Finally, the bivariate model requires IQ scores to be normally distributed, but this is unlikely given the finding that the IQ of students with LD reveals less stability over time (Kaye & Baron, 1987; Lally, Lloyd, & Kulberg, 1987).

With a proportion of the LD population showing IQ levels falling more than 1 SD below the mean, this group would, at one time, be considered as having borderline MR (see Heber, 1961). As such, this group would qualify under the rubric "slow learners" and likely manifest *generalized* academic deficiencies. The essential question: Is this group also LD? In some instances the answer might be affirmative, but the majority of this group would probably exhibit academic deficits across *all* achievement domains that would run counter to the assumption that students with LD exhibit achievement deficits in one or more (but not all) academic areas. When all academic achievement areas are equally depressed, the notion of *specificity*, in the sense of the presence of intra-individual differences, would not be achieved, even though the idea that LD results from a circumscribed set of problems that interfere selectively with academic performance has received support (Stanovich, 1986a). Thus, instead of specific LD (as defined in the federal definition), there is a more generalized LD, a concept closer to that defined by MR particularly at the borderline levels.

The "unexpected" failure idea often associated with LD has been the source of other concerns about IQ and LD. When identified as LD, a student presumably possesses average or above IQ and meets the discrepancy criterion which then suggests that the cause of the academic problems cannot be attributed to low intelligence. On the other hand, the academic deficiencies of slow learners should not be surprising because the demonstrated achievement problems are consistent with the lower than average intellectual ability. These differences suggest that the etiology of the two conditions is really not the same, and consequently, LD and LA groups appear to possess quantitative and qualitative differences.

Learning Disability and Low Achievement: Quantitative or Qualitative Differences?

The origins of assumptions about possible qualitative differences between LD and LA can be found in the Isle of Wight epidemiological studies (Rutter & Yule, 1975; see also Rutter & Yule, 1973; Yule, Rutter, Berger, & Thompson, 1974). Essentially, the LA sample of poor readers was differentiated into two groups: general reading backwardness (GRB) and specific reading retardation (SRR). The GRB group was defined as reading below expected CA (i.e., no discrepancy between IQ and achievement) while SRR was defined as reading below grade level predicted from IQ (i.e., the presence of an IQ-achievement discrepancy).

In analyzing the population, Rutter and Yule (1975) found that while IQ scores were approximately normally distributed, reading achievement scores did not show the same normal distribution because, at the lower end of the distribution, there was a "hump" indicating the presence of a greater proportion than the 2.3% expected in a normal distribution. This "hump" contained the SRR group whose problems were viewed as "specific" to the reading process. As Yule et al. (1974) suggested, "*Extreme* underachievement in reading occurs at appreciably above the rate expected on the basis of a normal distribution and so constitutes a hump at the lower end of the Gaussian curve.…There are no grounds for believing that the hump is anything but a true finding, and the finding implies that there is a group of children with severe and specific reading retardation which is *not* just the lower end of a normal continuum" (p. 10, emphasis in original).

Rutter and Yule (1975) concluded that, in addition to IQ differences, "Reading retardation is shown to differ significantly from reading backwardness in terms of sex ratio, neurological disorder, pattern of neurodevelopmental deficits and educational prognosis. It is concluded that the concept of specific reading retardation is valid" (p. 195). Rutter (1978) later affirmed the GRB-SRR distinction and the possibility of etiological differences particularly as manifested in the minimal brain dysfunction syndrome (Clements, 1966).

QUALITATIVE DISTINCTIONS IN MENTAL RETARDATION

The idea of distributional and etiological differences in a population was first proposed in the MR field. At IQ 50, it becomes possible to distinguish between mild and severe MR. Severe MR (about 25% of the MR population) typically represents "clinical" MR in the sense of probably possessing, besides limited cognitive ability, central nervous system pathology and associated disabilities. The larger mild MR group typically shows no neurological signs or associated clinical signs, and represents what is termed "familial" MR (Zigler, 1967). In the severe cases, the pathological factors significantly interfere with intellectual development (see Tarjan, Wright, Eyman, & Keeran, 1973) to such an extent that they distort the IQ score distribution as shown by Dingman and Tarjan (1960). In comparing the IQ distributions of low IQ populations (mild and severe) with those of the general population, there was an indication of an excess of cases ("hump") at the lower end of the distribution. Above IQ 50, there were few discrepancies between expected and actual percentages in the distribution but an excess of cases in the 0–19 IQ and 20–49 IQ ranges. This excess population formed a hump: an additional normal distribution of IQs with a mean IQ of 32 and an SD of 16. Clearly, when compared with IQ levels, the two groups appeared to differ with respect to etiology and clinical manifestations (Jastak, 1967; Weir, 1967).

The qualitative differences between the two MR "populations" became a source of debate and evolved into what was termed the "developmental-difference controversy" (Zigler & Balla, 1982). Generally, "this controversy centers around the question of whether the behavior of those retarded persons with no evidence of central nervous system dysfunction is best understood by those principles in developmental psychology that have been found to be generally applicable in explaining the behavior and development of non-retarded persons, or whether it is necessary to involve specific differences over and above a generally lower rate and asymptote of cognitive development" (p. 3).

QUALITATIVE DISTINCTIONS IN LEARNING DISABILITY

Because of the developmental-difference controversy, the related GRB-SRR distinction also became contentious. For example, many studies have failed to find a GRB-SRR bimodal distribution (e.g., Rodgers, 1983; Share, McGee, McKenzie, Williams, & Silva, 1987; Stevenson, 1988). Van der Wissel and Zegers (1985) suggested that no hump was found because it may, in reality, be an artifact resulting from floor and ceiling effects associated with the reading measures used. Using designs where students differed in reading level but were comparable in age (CA design) or comparable in reading level but varied in age (reading-level match design), a number of studies failed to demonstrate that SRR groups (achievement scores below levels predicted by IQ, i.e., discrepant) could be differentiated from a GRB group (depressed achievement not discrepant from IQ) (Fletcher et al., 1989; Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1992; Foorman, Francis, Fletcher, & Lynn, 1996; Rispens, van Yperen, & van Duijn, 1991; Share & Silva, 1986; B. A. Shaywitz et al., 1992; L. S. Siegel, 1992; Vellutino, Scanlon, & Lyon, 2000). Consequently, IQ was not a major factor associated with SRR, which was interpreted to mean that SRR was not a discrete entity, but rather

> …occurs along a continuum that blends imperceptibly with normal reading ability. These results indicate that no distinct cut-off exists to distinguish children with dyslexia clearly from children with normal reading ability; rather, the dyslexic children simply represent a lower portion of a continuum of reading capabilities (S. E. Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992, p. 148).

Rutter (1990) suggested that "the crucial test of the SRR hypothesis, however, does not depend on the presence or absence of a hump in the distribution but whether the correlates and outcomes of SRR serve to differentiate the syndrome from GBR" (p. 637). A number of studies have failed to differentiate GRB and SRR groups, however. For example, GRB groups (i.e., no I.Q.-achievement discrepancy) performed no differently on independent measures of reading achievement or on assessments of the cognitive abilities presumed to underlie the ability to learn to read (Fletcher et al., 1994; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Morris et al., 1998; Share, McGee, & Silva, 1989; Stanovich & Siegel, 1994). With respect to gender differences, the presumption of a disproportionately greater number of boys than girls in SRR groups has not received support (Pennington, Gilger, Olson, & DeFries, 1992; Share et al., 1987; S. E. Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). Finally, SRR groups were presumed to have a poorer educational prognosis than GRB groups (Yule, 1973), but no evidence supports the validity of this assumption (Francis et al., 1996; Share et al., 1989; B. A. Shaywitz et al., 1992; Vellutino et al., 1996).

In a summary of the available evidence, Fletcher et al. (1998) concluded that

> Under no circumstances is wholesale use of IQ test for learning disabilities justified. We have shown numerous problems with the discrepancy model, regardless of whether IQ tests or some other measures are used to operationalize the aptitude index. It is not the use of the IQ test that creates the problems with discrepancy. Classifications of children as discrepant versus low achievement lack discriminative validity (p. 200).

It was then suggested that the discrepancy criterion not be part of the LD identification process primarily because "it is not the score on the IQ test that identifies the child as having learning disabilities, but rather the score on the test of academic achievement that identifies the child with LD" (p. 201). Similarly, Aaron (1997) concluded that "a review of research in the area of reading disabilities indicates that classifying poor readers on the basis of a discrepancy formula into LD and non-LD categories lacks validity on both theoretical and empirical grounds" (p. 488). As an alternative, Aaron suggested a more pragmatic approach based on the Reading Component Model that identifies the source of the reading problem for all students and then focuses remedial efforts on that particular source.

The Status of Discrepancy in the Identification of Learning Disability

The discrepancy criterion for LD identification has thus been seriously challenged, with some anticipating its "impending demise" (Aaron, 1997). One difficulty, however, is in interpreting what that means for LD. The many analyses investigating discrepancy focused attention on the GRB-SRR distinction where, in both cases, the primary problem was an inability to read. Consequently, there was little question about the presence of reading disability (RD), but the presence or absence of LD was not really considered except by implication. Although students with LD are quite likely to manifest reading difficulties, they may not, and this fact makes any generalization from a GRB-SRR comparison suspect. The primary difficulty is conceptual and relates to the fact that if RD and LD are considered equivalent, then the law of parsimony is violated (Swanson, 1991). There appears, however, to be a decided tendency to view LD and RD as the same thing as evidenced by statements such as, "It is time to abandon the discrepancy-based classification of poor readers into LD and non-LD categories and expand the boundaries of LD to include all children who experience difficulties in learning to read" (Aaron, 1997, p. 488). Instead of providing conceptual clarity, such a suggestion would result in even greater confounding between concepts.

The same possible confounding is found with RD itself. The focus on GRB and SRR as discrete groups tends to obscure the fact that almost all students with SRR could be classified as GRB, while half of students with GRB can be classified as SRR (Hinshaw, 1992). Even when considering SRR itself, there are questions about its proper relationship with dyslexia, an RD equally difficult to define with precision (Benton & Pearl, 1978). The many similarities between the conditions raise the question as to whether SRR and dyslexia are the same thing (Yule & Rutter, 1976). Regardless of the answer, discussion about LD seems inappropriate as it is a different (and distinct) phenomenon that may or may not include students with these types of reading problems.

Thus, both LD and RD are complex entities, and eliminating the discrepancy criterion does not appear to be a sensible solution for resolving these complexities. Any suggested alternative, as, for example, in the Reading Component Model proposed by Aaron (1997), does not appear to be a viable solution in any significant sense unless it is also accompanied by a belief that LD is not a legitimate construct. When LD is not considered legitimate, there is a general theme that calls for a cessation of the illegitimate and unnecessary LD labeling, and a focusing instead on the difficulties of some students in learning to read by providing them with effective and responsive interventions (e.g., Christensen, 1992; McGill-Franzen & Allington, 1993; Swerling & Sternberg, 1996). As suggested by Aaron (1997), "When the discrepancy formula disappears from the educational scene, so will the concept of LD. After 40 years of wandering in the wilderness of learning disabilities, we are beginning to get a glimpse of the promised land" (p. 489). Whether or not the disappearance of the discrepancy formula leads to a promised land is certainly moot and would do little to resolve the complex and vexing problems associated with defining LD.

A major roadblock to problem resolution is the lack of a precise description of LD (Kavale & Forness, 2000). Although the description of LD is far from complete, the field has witnessed unprecedented growth and has accomplished this expansion not by using formal, albeit limited, definitions but rather by using a number of singular operational definitions stipulating rules about how a term is to apply in a particular case if specified actions yield characteristic results. Thus, a concept like LD may have a set of operations that define it, and knowing these operations presumably provides complete understanding of the concept (Benjamin, 1955).

For LD, the primary (and often sole) operation has been the application of a discrepancy criterion. Beginning with the USOE (1976) regulations and reaffirmed in proposed operational definitions (e.g., Chalfant & King, 1976; Shaw, Cullen, McGuire, & Brinckerhoff, 1995), discrepancy has emerged as the major means of LD identification. The LD identification process, however, may be more difficult and complicated than it appears to be with the use of a discrepancy criterion. For example, a problem surrounds the theoretical validity of operations. In a scientific sense, an operational definition must bear a logical and rational relationship with the verified theoretical constructs stipulated in the formal definition (Bergmann, 1961). For LD, a problem is created because the formal definition includes no mention of discrepancy (or underachievement) (Kavale, 1993). The resulting lack of congruence between definitions means that essentially two distinct views of LD are being presented: a formal representation and an operational representation.

The lack of correspondence creates a consequential problem: an increased probability that the operational definition may not be justified and may lead to potentially meaningless and insignificant operations that do not meet formal criteria (Deese, 1972). The operations specified may not actually "define" anything but merely state procedures required to test for the presence of the phenomenon to which the operations refer (Kavale, Forness, & Lorsbach, 1991). For example, assume an operational definition of LD that is based on the Learning Disability Coefficient (LDC) whose procedures require a calculation including an individual's white blood cell count multiplied by body weight in ounces, divided by head circumference in centimeters. Although possible to calculate, the LDC would possess little meaning or significance because the available validated knowledge about LD clearly indicates that the LDC does not "fit" any of it.

A less obvious example surrounds the different meanings that may be conveyed when different operational indicators are chosen. For example, discrepancy is defined as the difference between ability and achievement, but any number of ability (i.e., IQ) measures and probably even a greater number of achievement measures might be chosen for comparison. The problem is that when different combinations of measures are used to define discrepancy, it is not at all clear that the assessments are operationally, and thus, definitionally equivalent (Deese, 1972). It may, therefore, be difficult to "make sense" of the calculated discrepancy.

The use of operational definitions is thus neither a simple nor straightforward process but one that requires significant theoretical verification. Unfortunately, the LD field has not achieved the necessary verification primarily because discrepancy was so quickly embraced: "The debate that rages over what LD might be and the lack of consensus over the importance of any given variable is in sharp contrast to the relative unanimity regarding discrepancy. The consensus regarding discrepancy as the primary identification variable for LD has entrenched discrepancy to the point where it now represents the foundation concept for LD diagnosis" (Kavale & Forness, 1994, p. 23). In fact, discrepancy has become a deified concept as evidenced in its ascension to the status of "imperial criterion" (Mather & Healey, 1990) and a reified concept as seen in its elevation to an almost tangible property of students with LD (Kavale, 1987). Such deification and reification do not appear justified given the fact discrepancy itself is a hypothetical construct defined by hypothetical constructs (see Messick, 1981) resulting in the possibility that, in a theoretical sense, discrepancy may be a "fictitious concept" (Hempel, 1952).

The wide embrace of discrepancy has obscured some fundamental considerations. One such consideration surrounds the relationship between discrepancy and LD. With discrepancy often the only criterion used for LD identification, there has been an accompanying assumption that discrepancy represents the operational definition of LD. In reality, "Discrepancy is best associated with the concept of underachievement. This is true now and has historically been the case" (Kavale, 1987, p. 18). In a theoretical context, Shepard (1983, 1989) argued that discrepancy is the operational definition of underachievement. Thus, when a student meets the discrepancy criterion, what is being affirmed is underachievement, not LD. The scientific law of parsimony would suggest that underachievement and LD are not the same thing. To avoid confounding, the proper conclusion when the discrepancy criterion is met is that underachievement has been identified. If it is believed that underachievement is associated with LD (certainly a valid assumption), then discrepancy becomes a necessary but not sufficient criterion for LD identification (Kavale, 1987; Reynolds, 1984–1985).

Within the context of LD identification, discrepancy and the documentation of underachievement should represent only the first step in diagnosis (Kavale & Forness, 1994). Discrepancy is important in the identification process because it establishes a sound theoretical foundation for later LD determination. Although the discrepancy concept possesses psychometric and statistical problems, they have been satisfactorily addressed, and a technically defensible procedure to indicate the presence or absence of underachievement has been achieved. The findings from large-scale investigations appear to have affirmed the relationship between discrepancy and underachievement, and the possibility of reliably differentiating LD (i.e., students who meet the discrepancy criterion) from LA (i.e., students who do not meet the discrepancy criterion). Although critical as the initial step in LD determination, discrepancy should not be elevated to the status of being LD but rather viewed simply as the most useful means for defining underachievement, a necessary part of LD.

With discrepancy placed in proper perspective, attention needs to be directed at what else should be considered in the identification process in order to capture the complex and multivariate nature of LD (Kavale & Nye, 1991). Kavale and Forness (1995) suggested a way the process might proceed. The initial step is the formulation of foundation principles aimed at developing a theoretical framework for elucidating the basic nature of LD. Kavale and Forness (2000) elucidated the process further by proposing an operational definition in the form of a hierarchical scheme where each level depicts a decision point in the determination of LD. The scheme includes five levels where the first includes an ability (IQ)-achievement discrepancy to document the presence or absence of underachievement. The next levels focus on other stipulated criteria (e.g., psychological process deficits, exclusion), and a final LD designation is predicated on a student proceeding through each level. The process ceases if a student cannot meet the requisite criterion at any level. With its initial position, discrepancy provides the foundation and would be further strengthened if the difference score were based on the most reliable *total* IQ score and *total* achievement test score. In this way, a too narrowly focused discrepancy, as in, for example, a comparison between a Performance IQ and a Social Studies achievement subtest, would be eliminated. With such a scheme, a more comprehensive view of LD is achieved along with greater confidence in declaring that a student is "truly" LD.

Conclusion

Discrepancy is an important and legitimate concept applied to LD. Beginning with its status as a measure of educational progress, discrepancy evolved into an index of underachievement. Because LD has always been viewed as a construct associated with underachievement, discrepancy became a necessary component of LD. Although subject to debate about statistical and psychometric properties, discrepancy calculation can be made adequate and defensible for use in LD identification. Because of pragmatic reasons, discrepancy has become the primary LD identification criterion, and this emphasis has led to a number of difficulties, most noticeably the failure to appropriately differentiate LD and LA. When viewed properly, discrepancy is a useful component for LD identification and any presumed problems can be resolved satisfactorily. The most important point is that discrepancy not be used alone for LD identification. Discrepancy is the operational definition of underachievement and, when present, reliably and appropriately documents the presence of underachievement, not LD. With the valid assumption that LD and underachievement are not equivalent, the task becomes one of deciding what other factors need to be considered before there is confidence that LD has been determined. When placed in proper context, any arguments about the use of discrepancy for LD determination would cease. It would, therefore, be an error to eliminate discrepancy when considering the best means of defining the LD construct. The task is one of using discrepancy so that it is not LD itself but rather only part of a more comprehensive identification process.

References

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research*, *67*, 461–502.

Ackerman, P., Peters, J., & Dykman, P. (1971). Children with learning disabilities: WISC profiles. *Journal of Learning Disabilities*, *4*, 150–166.

Algozzine, B. (1985). Low achiever differentiation: Where's the beef? *Exceptional Children, 52*, 72–75.

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

Algozzine, B., & Ysseldyke, J. E. (1981a). An analysis of difference score reliabilities on three measures with a sample of low-achieving youngsters. *Psychology in the Schools, 18*, 133–138.

Algozzine, B., & Ysseldyke, J. E. (1981b). Special education services for normal children: Better safe than sorry. *Exceptional Children, 48*, 238–243.

Algozzine, B., & Ysseldyke, J. (1983). Learning disabilities as a subset of school failure: The oversophistication of a concept. *Exceptional Children, 50*, 242–246.

Algozzine, B., & Ysseldyke, J. E. (1987). Questioning discrepancies: Retaking the first step 20 years later. *Learning Disability Quarterly, 10*, 301–312.

Algozzine, B., Ysseldyke, J. E., & McGue, M. (1995). Differentiating low-achieving students: Thoughts on setting the record straight. *Learning Disabilities Research and Practice, 10*, 140–144.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. R. (1982). Identifying children with learning disabilities: When is a discrepancy severe? *Journal of School Psychology, 20*, 299–305.

Alspaugh, J., & Burge, P. (1972). Determination of reading expectancy. *The Journal of Experimental Education, 40*, 1–5.

Ames, L. B. (1968). A low intelligence quotient often not recognized as the chief cause of many learning difficulties. *Journal of Learning Disabilities, 1*, 735–738.

Baldwin, R. S., & Vaughn, S. (1989). Why Siegel's arguments are irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 513, 520.

Bannatyne, A. (1968). Diagnosing learning disabilities and writing remedial prescriptions. *Journal of Learning Disabilities, 1*, 242–249.

Barnett, D. W., & Macmann, G. M. (1992). Aptitude-achievement discrepancy scores: Accuracy in analysis misdirected. *School Psychology Review, 21*, 494–508.

Bateman, B. D. (1965). An educational view of a diagnostic approach to learning disabilities. In J. Hellmuth (Ed.), *Learning disorders* (Vol. 1, pp. 219–239). Seattle, WA: Special Child Publications.

Belmont, I., & Belmont, L. (1980). Is the slow learner in the classroom learning disabled? *Journal of Learning Disabilities, 13*, 496–499.

Benjamin, A. C. (1955). *Operationism*. Springfield, IL: C. C. Thomas.

Bennett, D., & Clarizio, H. (1988). A comparison of methods for calculating a severe discrepancy. *Journal of School Psychology, 20*, 359–369.

Benton, A. L., & Pearl, D. (Eds.). (1978). *Dyslexia: An appraisal of current knowledge*. New York: Oxford University Press.

Bergmann, G. (1961). Sense and nonsense in operationism. In P. G. Frank (Ed.), *The validation of scientific theories* (pp. 45–56). New York: Collier.

Berk, R. A. (1981). What's wrong with using grade-equivalent scores to identify learning disabled children? *Academic Therapy, 17*,(2), 133–140.

Berk, R. A. (1984). An evaluation of procedures for computing an ability-achievement discrepancy score. *Journal of Learning Disabilities, 17*, 262–266.

Bijou, S. W. (1942). The psychometric pattern approach as an aid to clinical analysis—a review. *American Journal of Mental Deficiency, 46*, 354–362.

Bishop, D. V., & Butterworth, G. E. (1980). Verbal-performance discrepancies: Relationship to birth risk and specific reading retardation. *Cortex, 16*, 375–389.

Bond, G. L., & Tinker, M. (1973). *Reading difficulties: Their diagnosis and correction (3rd ed.).* New York: Appleton-Century-Crofts.

Boodoo, G. (1984–1985). A multivariate perspective for aptitude-achievement discrepancy in learning disability assessment. *Journal of Special Education, 18*, 489–494.

Braden, J. P., & Weiss, L. (1988). Effects of simple difference versus regression discrepancy methods: An empirical study. *Journal of School Psychology, 26*, 133–142.

Burns, E. (1984). The bivariate normal distribution and the I.Q. of learning disability samples. *Journal of Learning Disabilities, 17*, 294–295.

Chalfant, J. C. (1985). Identifying learning disabled students: A summary of the National Task Force Report. *Learning Disabilities Focus, 1,* 9–20.

Chalfant, J. C., & King, F. S. (1976). An approach to operationalizing the definition of learning disabilities. *Journal of Learning Disabilities, 9*, 228–243.

Chapman, J. C. (1923). The unreliability of the difference between intelligence and educational ratings. *Journal of Educational Psychology, 14*, 103–108.

Chapman, J. S., & Boersma, F. J. (1979). Learning disabilities, locus of control, and mother attitudes. *Journal of Educational Psychology, 71*, 250–258.

Christensen, C. A. (1992). Discrepancy definitions of reading disability: Has the quest led us astray? A response to Stanovich. *Reading Research Quarterly, 27*, 276–278.

Clarizio, H. F., & Bennett, D. E. (1987). Diagnostic utility of the K-ABC and WISC-R/PIAT in determining severe discrepancy. *Psychology in the Schools, 24*, 309–315.

Clarizio, H. F., & Phillips, S. E. (1986). The use of standard scores in diagnosing learning disabilities: A critique. *Psychology in the Schools, 23*, 380–387.

Clarizio, H. F., & Phillips, S. E. (1989). Defining severe discrepancy in the diagnosis of learning disabilities: A comparison of methods. *Journal of School Psychology, 27*, 383–391.

Clausen, J. (1972). Quo vadis, AAMD? *Journal of Special Education, 6*, 51–60.

Clements, S. D. (1966). *Minimal brain dysfunction in children: Terminology and identification*. (NINDS Monograph No. 3, U.S. Public Health Service Publication No. 1415). Washington, DC: Department of Health, Education, and Welfare.

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

Cone, T. E., & Wilson, L. R. (1981). Quantifying a severe discrepancy: A critical analysis. *Learning Disability Quarterly, 4*, 359–371.

Cone, T. E., Wilson, L. R., Bradley, C. M., & Reese, J. H. (1985). Characteristics of LD students in

Iowa: An empirical investigation. *Learning Disability Quarterly, 8*, 211–220.

Crane, A. R. (1959). An historical critical account of the accomplishment quotient idea. *British Journal of Educational Psychology, 29*, 252–259.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"—or should we? *Psychological Bulletin, 74*, 68–80.

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

Cruickshank, W. M. (1977). Myths and realities in learning disabilities. *Journal of Learning Disabilities, 10*, 51–58.

Cureton, E. E. (1937). The accomplishment quotient technique. *Journal of Experimental Education, 5*, 315–326.

Dangel, H. L., & Ensminger, E. E. (1988). The use of discrepancy formulas with LD students. *Learning Disabilities Focus, 4*, 24–31.

Danielson, L. C., & Bauer, J. N. (1978). A formula-based classification of learning disabled children: An examination of the issues. *Journal of Learning Disabilities, 11*, 163–176.

Davis, W. A., & Shepard, L. A. (1983). Specialists' use of tests and clinical judgment in the diagnosis of learning disabilities. *Learning Disability Quarterly, 6*, 128–138.

Deese, J. (1972). *Psychology as science and art*. New York: Harcourt Brace Jovanovich.

Dingman, H. F., & Tarjan, G. (1960). Mental retardation and the normal distribution curve. *American Journal of Mental Deficiency, 64*, 991–994.

Divoky, D. (1974). Education's latest victim: The "LD" kid. *Learning, 3*, 4–8.

Dore-Boyce, K., Misner, M., & McGuire, D. (1975). Comparing reading expectancy formulas. *The Reading Teacher, 29*, 8–14.

Dudley-Marling, C., Kaufman, N., & Tarver, S. (1981). WISC and WISC-R profiles of learning disabled children: A review. *Learning Disability Quarterly, 4*, 307–319.

Elliot, M. (1981). Quantitative evaluation procedures for learning disabilities. *Journal of Learning Disabilities, 14*, 84–87.

Epps, S., McGue, M., & Ysseldyke, J. E. (1982). Inter-judge agreement in classifying students as learning disabled. *Psychology in the Schools, 19*, 209–220.

Epps, S., Ysseldyke, J. E., & Algozzine, B. (1983). Impact of different definitions of learning disabilities on the number of students identified. *Journal of Psychoeducational Assessment, 1*, 341–352.

Epps, S., Ysseldyke, J. E., & Algozzine, B. (1985). An analysis of the conceptual framework underlying definitions of learning disabilities. *Journal of School Psychology, 23*, 133–144.

Epps, S., Ysseldyke, J. E., & McGue, M. (1984). "I know one when I see one"—Differentiating LD and non-LD students. *Learning Disability Quarterly, 7*, 89–101.

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

Evans, L. D. (1990). A conceptual overview of the regression discrepancy model for evaluating severe discrepancy between I.Q. and achievement scores. *Journal of Learning Disabilities, 23*, 406–412.

Feldt, L. A. (1967). Reliability of differences between scores. *American Educational Research Journal, 4*, 139–145.

Finlan, T. G. (1992). Do state methods of quantifying a severe discrepancy result in fewer students with learning disabilities? *Learning Disability Quarterly, 15*, 129–134.

Finlan, T. G. (1994). *Learning disability: The imaginary disease*. Westport, CT: Bergin & Garvey.

Fletcher, J. M., Espy, K. A., Francis, P. J., Davidson, K. C., Rourke, B. P., & Shaywitz, S. E. (1989). Comparisons of cutoff and regression-based definitions of reading disabilities. *Journal of Learning Disabilities, 22,* 334–338.

Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, S. E., & Shaywitz, B. A. (1992). The validity of discrepancy-based definitions of reading disabilities. *Journal of Learning Disabilities, 25*, 555–561, 573.

Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research and Practice, 13*, 186–203.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6–23.

Foorman, B. R., Francis, D. J., Fletcher, J. M., & Lynn, A. (1996). Relation of phonological and orthographic processing to early reading: Comparing two approaches to regression-based, reading-level-match design. *Journal of Educational Psychology, 88*, 639–652.

Forness, S. R. (1985). Effects of public policy at the state level: California's impact on MR, LD, and ED categories. *Remedial and Special Education, 6*, 36–43.

Forness, S. R., Sinclair, E., & Guthrie, D. (1983). Learning disability discrepancy formulas: Their use in actual practice. *Learning Disability Quarterly, 6*, 107–114.

Frame, R. E., Clarizio, H. F., Porter, A. F., & Vinsonhaler, J. R. (1982). Inter-clinician agreement and bias in school psychologists' diagnostic and treatment recommendations for a learning disabled child. *Psychology in the Schools, 19*, 319–327.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*, 3–17.

Frankenberger, W., & Fronzaglio, K. (1991). A review of states' criteria and procedures for identifying children with learning disabilities. *Journal of Learning Disabilities, 24*, 495–500.

Franzen, R. (1920). The accomplishment quotient: A school mark in terms of individual capacity. *Teachers College Record, 21*, 432–440.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Lipsey, M. W. (2000). Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 81–104). Mahwah, NJ: Lawrence Erlbaum.

Furlong, M. J. (1988). An examination of an implementation of simple difference score distribution model in learning disability identification. *Psychology in the Schools, 25*, 132–145.

Furlong, M. J., & Feldman, M. G. (1992). Can ability-achievement regression to the mean account for MDT discretionary decisions? *Psychology in the Schools, 29*, 205–212.

Furlong, M. J., & Yanagida, E. H. (1985). Psychometric factors affecting multidisciplinary team identification of learning disabled children. *Learning Disability Quarterly, 8,* 37–46.

Gajar, A. H. (1980). Characteristics across exceptional categories: EMR, LD, and ED. *Journal of Special Education, 14*, 165–173.

Gallagher, J. J. (1966). Children with developmental imbalances: A psychoeducational definition. In W. M. Cruickshank (Ed.), *The teacher of brain-injured children* (pp. 20–34). Syracuse, NY: Syracuse University Press.

Gallagher, J. J. (1986). Learning disabilities and special education: A critique. *Journal of Learning Disabilities, 19*, 595–601.

Gelzheiser, L. M. (1987). Reducing the number of students identified as learning disabled: A question of practice, philosophy, or policy? *Exceptional Children, 54*, 145–150.

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*, 137–148.

Gordon, M., Lewandowski, L., & Keiser, S. (1999). The LD label for relatively well-functioning students: A critical analysis. *Journal of Learning Disabilities, 32*, 485–490.

Gottesman, R. C. (1979). Follow-up of learning disabled children. *Learning Disability Quarterly, 2*, 60–69.

Gottlieb, J., Alter, M., Gottlieb, B. M., & Wishner, J. (1994). Special education in urban America: It's not justifiable for many. *Journal of Special Education, 27*, 453–465.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1996). Learning disabilities, low achievement, and mild mental retardation: More alike than different? *Journal of Learning Disabilities, 29*, 570–581.

Grossman, H. J. (Ed.). (1973). *Manual on terminology and classification in mental retardation*. Washington, DC: American Association on Mental Deficiency.

Grossman, H. J. (Ed.). (1983). *Classification in mental retardation* (3rd rev.). Washington, DC: American Association on Mental Deficiency.

Gullicksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Gutkin, T. B. (1979). WISC-R scatter indices: Useful information for differential diagnosis? *Journal of School Psychology, 17*, 368–371.

Hanna, G. S., Dyck, N. A., & Holen, M. C. (1979). Objective analysis of achievement-aptitude discrepancies in LD classification. *Learning Disability Quarterly, 2*, 32–38.

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

Harris, A. J. (1971). *How to increase reading ability* (5th ed.). New York: McKay.

Harris, A. J. (1975). *How to increase reading ability* (6th ed.). New York: McKay.

Heber, R. F. (1959). A manual on terminology and classification in mental retardation. *American Journal of Mental Deficiency Monograph (Supp. 64).*

Heber, R. (1961). Modifications in the manual on terminology and classification in mental retardation. *American Journal of Mental Deficiency, 65*, 499–500.

Hempel, C. (1952). Fundamentals of concept formation in empirical science. *International Encyclopedia of Unified Science (Vol. II, No. 7)*. Chicago: University of Chicago Press.

Hessler, G. L. (1987). Educational issues surrounding severe discrepancy. *Learning Disabilities Research, 3*, 43–49.

Hinshaw, S. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127–155.

Hoffman, J. V. (1980). The disabled reader: Forgive us our regressions and lead us not into expectations. *Journal of Learning Disabilities, 13*, 7–11.

Hollingworth, L. S. (1926). *The psychology of subnormal children*. New York: Macmillan.

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GEs. *Educational Measurement: Issues and Practice, 3*, 8–14.

Hopkins, K. D., & Stanley, J. C. (1981). *Educational and psychological measurement and evaluation* (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Horn, A. (1941). *The uneven distribution of the effects of special factors*. Southern California Education Monographs, No. 12.

Iowa Department of Public Instruction. (1981). *The identification of pupils with learning disabilities*. Des Moines, IA: Author.

Jastak, J. F. (1967). Mental Retardation. *Science, 155*, 577–578.

Jenkins, J. R., & Pany, D. (1978). Standardized achievement tests: How useful for special education? *Exceptional Children, 44*, 448–453.

Johnson, D. J., & Myklebust, H. R. (1967). *Learning disabilities: Educational principles and practices*. New York: Grune & Stratton.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 78*, 243–245.

Kaufman, A. S. (1981). The WISC-R and learning disabilities assessment: State of the art. *Journal of Learning Disabilities, 14*, 520–526.

Kavale, K. A. (1987). Theoretical issues surrounding severe discrepancy. *Learning Disabilities Research, 3*, 12–20.

Kavale, K. A. (1993). A science and theory of learning disabilities. In G. R. Lyon, D. B. Gray, J. F. Kavanagh, & N. A. Krasnegor (Eds.), *Better understanding of learning disabilities: New views from research and their implications for education and public policies* (pp. 171–195). Baltimore: Brookes.

Kavale, K. A. (1995). Setting the record straight on learning disability and low achievement: The tortuous path of ideology. *Learning Disabilities Research and Practice, 10*, 145–152.

Kavale, K. A., & Forness, S. R. (1984). A meta-analysis of the validity of Wechsler Scale profiles and recategorizations: Patterns or parodies? *Learning Disability Quarterly, 7*, 136–156.

Kavale, K. A., & Forness, S. R. (1985). *The science of learning disabilities*. San Diego, CA: College-Hill Press.

Kavale, K. A., & Forness, S. R. (1994). Learning disabilities and intelligence: An uneasy alliance. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 8, pp. 1–63). Greenwich, CT: JAI Press.

Kavale, K. A., & Forness, S. R. (1995). *The nature of learning disabilities: Critical elements of diagnosis and classification*. Mahwah, NJ: Lawrence Erlbaum.

Kavale, K. A., & Forness, S. R. (2000). What definitions of learning disability say and don't say: A critical analysis. *Journal of Learning Disabilities, 33*, 239–256.

Kavale, K. A., Forness, S. R., & Lorsbach, T. C. (1991). Definition for definitions of learning disabilities. *Learning Disability Quarterly, 14*, 257–266.

Kavale, K. A., Fuchs, D., & Scruggs, T. E. (1994). Setting the record straight on learning disability and low achievement: Implications for policymaking. *Learning Disabilities Research and Practice, 9*, 70–77.

Kavale, K. A., & Reese, J. H. (1992). The character of learning disabilities: An Iowa profile. *Learning Disability Quarterly, 15*, 74–94.

Kaye, D. B., & Baron, M. B. (1987). Long-term stability of intelligence and achievement scores in specific-learning-disabilities samples. *Journal of Psychoeducational Assessment, 3*, 257–266.

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: World Book.

Kirk, S. A. (1962). *Educating exceptional children*. Boston: Houghton Mifflin.

Kirk, S. A. (1975). Behavioral diagnosis and remediation of learning disabilities. In S. A. Kirk & J. M. McCarthy (Eds.), *Learning disabilities: Selected ACLD papers*. Boston: Houghton Mifflin.

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

Klatt, H. J. (1991). Learning disabilities: A questionable construct. *Educational Theory, 41*, 47–60.

Koppitz, E. M. (1971). *Children with learning disabilities: A five year follow-up study*. New York: Grune & Stratton.

Lally, M. J., Lloyd, R. D., & Kulberg, J. M. (1987). Is intelligence stable in learning-disabled children? *Journal of Psychoeducational Assessment, 3*, 411–416.

Lester, G., & Kelman, M. (1997). State disparities in the diagnosis and placement of pupils with learning disabilities. *Journal of Learning Disabilities, 30*, 599–607.

Lloyd, J., Sabatino, D., Miller, T., & Miller, S. (1977). Proposed federal guidelines: Some open questions. *Journal of Learning Disabilities, 10*, 69–71.

Lyon, G. R. (1989). I.Q. is irrelevant to the definition of learning disabilities: A position in search of logic and data. *Journal of Learning Disabilities, 22*, 504–506, 512.

Macmann, G. M., & Barnett, D. W. (1985). Discrepancy score analysis: A computer simulation of classification stability. *Journal of Psychoeducational Assessment, 4,* 363–375.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *Journal of Special Education, 23*, 127–143.

MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998). Discrepancy between definitions of learning disabilities and school practices: An empirical investigation. *Journal of Learning Disabilities, 31*, 314–326.

MacMillan, D. L., Gresham, F. M., Siperstein, G. N., & Bocian, K. M. (1996). The labyrinth of IDEA: School decisions on referred students with subaverage general intelligence. *American Journal on Mental Retardation, 101*, 161–174.

MacMillan, D. L., & Jones, R. L. (1972). Lions in search of more Christians. *Journal of Special Education, 6*, 81–91.

MacMillan, D., Siperstein, G., & Gresham, F. (1996). A challenge to the viability of mild mental retardation as a diagnostic category. *Exceptional Children, 62*, 356–371.

Macy, M., Baker, J., & Kosinski, S. (1979). An empirical study of the Myklebust learning quotient. *Journal of Learning Disabilities, 12*, 93–96.

Mather, N., & Healey, W. C. (1990). Deposing aptitude-achievement discrepancy as the imperial criterion for learning disabilities. *Learning Disabilities: A Multidisciplinary Journal, 1*, 40–48.

McDermott, P. A., & Watkins, M. (1985). McDermott Multidimensional Assessment of Children. [Computer program]. New York: Psychological Corporation.

McGill-Franzen, S., & Allington, R. L. (1993). Flunk 'em or get them classified: The contamination of primary grade accountability data. *Educational Researcher, 22*, 19–22.

McKnight, P. T. (1982). The learning disability myth in American education. *Journal of Education, 164*, 351–359.

McLeod, J. (1979). Educational underachievement: Toward a defensible psychometric definition. *Journal of Learning Disabilities, 12*, 322–330.

McLeskey, J. (1989). The influence of level of discrepancy on the identification of students with learning disabilities. *Journal of Learning Disabilities, 22*, 435–438, 443.

McLeskey, J. (1992). Students with learning disabilities at primary, intermediate, and secondary grade levels: Identification and characteristics. *Learning Disability Quarterly, 15*, 13–19.

McLeskey, J., & Waldron, N. L. (1990). The identification and characteristics of students with learning disabilities in Indiana. *Learning Disabilities Research, 5*, 72–78.

McLeskey, J., & Waldron, N. L. (1991). Identifying students with learning disabilities: The effect of implementing statewide guidelines. *Journal of Learning Disabilities, 24*, 501–506.

McPhail, A. H. (1922). The correlation between the I.Q. and the A.Q. *School and Society, 16*, 586–588.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Mercer, C. D., Jordan, L., Allsopp, D. H., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. *Journal of Learning Disabilities, 19*, 217–232.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin, 89*, 575–588.

Meyen, E. (1989). Let's not confuse test scores with the substance of the discrepancy model. *Journal of Learning Disabilities, 22*, 482–483.

Miller, M. (1980). On the attempt to find WISC-R profiles for learning and reading disabilities: A response to Vance, Wallbrown, and Blaha. *Journal of Learning Disabilities, 13,* 338–340.

Monroe, M. (1932). *Children who cannot read*. Chicago: University of Chicago Press.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., Katz, L., Francis, D. J., & Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*, 347–373.

Morrison, G. M., MacMillan, D. L., & Kavale, K. A. (1985). System identification of learning disabled children: Implications for research sampling. *Learning Disability Quarterly, 8*, 2–10.

Myklebust, H. R., Bannochie, M. N., & Killen, J. R. (1971). Learning disabilities and cognitive processes. In H. R. Myklebust (Ed.), *Progress in learning disabilities* (Vol. II, pp. 213–251). New York: Grune & Stratton.

Naglieri, J. A. (1981). Factor structure of the WISC-R for children identified as learning disabled. *Psychological Reports, 49*, 891–895.

Noel, M. M., & Fuller, B. C. (1985). The social policy construction of special education: The impact of state characteristics on identification and integration of handicapped children. *Remedial and Special Education, 6*, 27–35.

Norman, C. A., & Zigmond, N. (1980). Characteristics of children labeled and served as learning disabled in school systems affiliated with Child Service Demonstration Centers. *Journal of Learning Disabilities, 13*, 542–547.

O'Donnell, L. E. (1980). Intra-individual discrepancy in diagnosing specific learning disabilities. *Learning Disability Quarterly, 3*, 10–18.

O'Shea, L. J., & Valcante, G. (1986). A comparison over time of relative discrepancy scores of low achievers. *Exceptional Children, 53*, 253–259.

Payne, R. W., & Jones, H. G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13*, 115–121.

Pennington, B. F., Gilger, J. W., Olson, R. K., & DeFries, J. C. (1992). The external validity of age- versus I.Q.-discrepancy definitions of reading disability: Lessons from a twin study. *Journal of Learning Disabilities, 25*, 562–573.

Perlmutter, B., & Parus, M. (1983). Identifying children with learning disabilities: A comparison of diagnostic procedures across school districts. *Learning Disability Quarterly, 6*, 321–328.

Pintner, R., & Marshall, H. (1921). A combined mental-educational survey. *Journal of Educational Psychology, 12*, 32–43.

Piotrowski, R. J., & Siegel, D. J. (1986). The I.Q. of learning disability samples: A reexamination. *Journal of Learning Disabilities, 19*, 492–493.

Popenoe, H. (1927). A report of certain significant deficiencies of the accomplishment quotient. *Journal of Educational Research, 16*, 40–47.

Reger, R. (1979). Learning disabilities: Futile attempts at a simplistic definition. *Journal of Learning Disabilities, 12*, 31–34.

Reschly, D. J. (1992). Mental retardation: Conceptual foundation, definitional criteria, and diagnostic operations. In S. R. Hooper, G. W. Hynd, & R. E. Mattison (Eds.), *Developmental disorders: Diagnostic criteria and clinical assessment* (pp. 23–67). Hillsdale, NJ: Lawrence Erlbaum.

Reynolds, C. R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. *Journal of School Psychology, 19*, 350–358.

Reynolds, C. R. (1984–1985). Critical measurement issues in learning disabilities. *Journal of Special Education, 18*, 451–476.

Reynolds, C. R. (1985). Measuring the aptitude-achievement discrepancy in learning disability diagnosis. *Remedial and Special Education, 6*, 37–55.

Reynolds, C. R., & Snow, M. (1985). Severe discrepancy analysis. [Computer program]. College Station, TX: TRAIN.

Rispens, J., van Yperen, T. A., & van Duijn, G. A. (1991). The irrelevance of I.Q. to the definition of learning disabilities: Some empirical evidence. *Journal of Learning Disabilities, 24*, 434–438.

Rodenborn, L. V. (1974). Determining and using expectancy formulas. *The Reading Teacher, 28*, 286–291.

Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology, 53*, 369–373.

Ross, R. P. (1992a). Accuracy in analysis of discrepancy scores: A nationwide study of school psychologists. *School Psychology Review, 21*, 480–493.

Ross, R. P. (1992b). Aptitude-achievement discrepancy scores: Accuracy in analysis ignored. *School Psychology Review, 21,* 509–514.

Ruch, G. M. (1923). The achievement quotient technique. *Journal of Educational Psychology, 14*, 334–343.

Rutter, M. (1978). Prevalence and types of dyslexia. In A. L. Benton, & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 5–28). New York: Oxford University Press.

Rutter, M. (1990). Isle of Wight revisited: Twenty-five years of child psychiatric epidemiology. *Journal of American Academy of Child and Adolescent Psychiatry, 29*, 633–653.

Rutter, M., & Yule, W. (1973). Specific reading retardation. In L. Mann, & D. Sabatino (Eds.), *The first review of special education* (pp. 49–62). Philadelphia: JSE Press.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181–197.

Sabatino, D. A., & Miller, T. L. (1980). The dilemma of diagnosis in learning disabilities: Problems and potential directions. *Psychology in the Schools, 17*, 76–86.

Salvia, J., & Clark, J. (1973). Use of deficits to identify the learning disabled. *Exceptional Children, 39*, 305–308.

Schulte, A., & Borich, G. D. (1984). Considerations in the use of difference scores to identify learning-disabled children. *Journal of School Psychology, 22*, 381–390.

Scruggs, T. E. (1987). Theoretical issues surrounding severe discrepancy: A discussion. *Learning Disabilities Research, 3*, 21–23.

Selz, M., & Reitan, R. (1979). Rules for neuropsychological diagnosis: Classification of brain function in older children. *Journal of Consulting and Clinical Psychology, 47*, 258–264.

Senf, G. M. (1977). A perspective on the definition of LD. *Journal of Learning Disabilities, 10*, 537–539.

Share, D. L., McGee, R., McKenzie, D., Williams, S., & Silva, P. A. (1987). Further evidence relating to the distinction between specific reading retardation and general reading backwardness. *British Journal of Developmental Psychology, 5*, 34–44.

Share, D. L., McGee, R., & Silva, P. D. (1989). I.Q. and reading progress: A test of the capacity notion of I.Q. *Journal of American Academy of Child and Adolescent Psychiatry, 28*, 97–100.

Share, D. L., & Silva, P. A. (1986). The stability and classification of specific reading retardation: A longitudinal study from age 7 to 11. *British Journal of Educational Psychology, 56*, 32–39.

Share, D. L., & Silva, P. A. (1987). Language deficits and specific reading retardation: Cause or effect? *British Journal of Disorders of Communication, 22*, 219–226.

Shaw, S. F., Cullen, J. P., McGuire, J. M., & Brinckerhoff, L. C. (1995). Operationalizing a definition of learning disabilities. *Journal of Learning Disabilities, 28*, 586–597.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326*, 145–150.

Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., & Shaywitz, S. E. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. *Journal of Learning Disabilities, 25*, 639–648.

Shaywitz, B. A., Holford, T. R., Holohan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for I.Q. but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30*, 894–906.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association, 264*, 998–1002.

Shepard, L. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. *Journal of Special Education, 14*, 79–91.

Shepard, L. A. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. *Educational Measurement: Issues and Practices, 2*, 4–8.

Shepard, L. A. (1989). Identification of mild handicaps. In. R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 545–572). New York: Macmillan.

Shepard, L. A., & Smith, M. L. (1983). An evaluation of the identification of learning disabled students in Colorado. *Learning Disability Quarterly, 6*, 115–127.

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

Short, E. J., Feagans, L., McKinney, J. D., & Appelbaum, M. I. (1986). Longitudinal stability of LD subtypes based on age-and I.Q.-achievement discrepancies. *Learning Disability Quarterly, 9*, 214–225.

Siegel, E. (1968). Learning disabilities: Substance or shadow. *Exceptional Children, 35*, 433–437.

Siegel, L. S. (1988). Evidence that I.Q. scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42*, 201–215.

Siegel, L. S. (1989). I.Q. is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469–478, 486.

Siegel, L. S. (1990). IQ and learning disabilities: R.I.P. In H. L. Swanson & B. Keogh (Eds.), *Learning disabilities: Theoretical and research issues* (pp. 111–128). Hillsdale, NJ: Erlbaum.

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*, 618–629.

Simmons, G. A., & Shapiro, B. J. (1968). Reading expectancy formulas: A warning note. *Journal of Reading, 2*, 625–629.

Sinclair, E., & Alexson, J. (1986). Learning disability discrepancy formulas: Similarities and differences among them. *Learning Disabilities Research, 1*, 112–118.

Sinclair, E., Guthrie, D., & Forness, S. R. (1984). Establishing a connection between severity of learning disabilities and classroom attention problems. *Journal of Educational Research, 78,* 18–21.

Singer, J., Palfrey, J., Butter, J., & Walker, D. (1989). Variation in special education classification across school districts: How does where you live affect what you are labeled? *American Educational Research Journal, 26*, 261–281.

Smith, M. D., Coleman, J. M., Dokecki, P. R., & Davis, E. E. (1977). Intellectual characteristics of school labeled learning disabled children. *Exceptional Children, 43*, 352–357.

Sobel, M., & Kelemen, E. (1984). On an application of statistical ranking and selection tables to the identification of the learning disabled. *Journal of Educational Statistics, 9*, 215–226.

Stanovich, K. E. (1986a). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. In J. K. Torgesen, & B .Y. L. Wong (Eds.), *Psychological and educational perspectives on learning disabilities* (pp. 87–131). Orlando, FL: Academic Press.

Stanovich, K. E. (1986b). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Stanovich, K. E. (1989). Has the learning disabilities field lost its intelligence? *Journal of Learning Disabilities, 22*, 487–492.

Stanovich, K. E. (1991a). Conceptual and empirical problems with discrepancy definitions of reading disability. *Learning Disability Quarterly, 14*, 269–280.

Stanovich, K. E. (1991b). Discrepancy definitions of reading disability: Has intelligence led us astray?

Reading Research Quarterly, 26, 7–29.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24–53.

Stevenson, J. (1988). Which aspects of reading disability show a "hump" in their distribution? *Applied Cognitive Psychology, 2*, 77–85.

Sulzbacher, S., & Kenowitz, L. A. (1977). At last, a definition of learning disabilities we can live with? *Journal of Learning Disabilities, 10*, 67–69.

Swanson, H. L. (1991). Operational definitions and learning disabilities: An overview. *Learning Disability Quarterly, 14*, 242–254.

Swerling, S. L., & Sternberg, R. J. (1996). *Off track: When poor readers become learning disabled.* Boulder, CO: Westview Press.

Tarjan, G., Wright, S. W., Eyman, R. K., & Keeran, C. V. (1973). Natural history of mental retardation: Some aspects of epidemiology. *American Journal of Mental Deficiency, 77*, 369–379.

Thomson, M. (1982). Assessing the intelligence of dyslexic children. *Bulletin of the British Psychological Society, 35*, 94–96.

Thorndike, R. L. (1963). *The concepts of over- and underachievement*. New York: Teachers College, Columbia University Press.

Thurlow, M. L., & Ysseldyke, J. E. (1979). Current assessment and decision-making practices in model LD programs. *Learning Disability Quarterly, 2*, 15–24.

Thurlow, M. L., Ysseldyke, J. E., & Casey, A. (1984). Teachers' perceptions of criteria for identifying learning disabled students. *Psychology in the Schools, 21*, 349–355.

Toops, H. A., & Symonds, P. M. (1922). What shall be expected of the AQ? *Journal of Educational Psychology, 13*, 513–528.

U.S. Department of Education. (1999). *To assure the free appropriate public education of all children with disabilities: Twenty-first annual report to the Congress on the implementation of the Individuals With Disabilities Education Act*. Washington, DC: Author.

U.S. Office of Education. (1968). *First annual report of the National Advisory Committee for Handicapped Children.* Washington, DC: U.S. Department of Health, Education, and Welfare.

U.S. Office of Education (1976, December 29). Proposed rule-making. *Federal Register, 41*, (230), 52404–52407. Washington, DC: U.S. Government Printing Office.

U.S. Office of Education (1977, December 29). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42*, (250), 65082–65085. Washington, DC: U.S. Government Printing Office.

Valus, A. (1986a). Achievement-potential discrepancy status of students in LD programs. *Learning Disability Quarterly, 9*, 200–205.

Valus, A. (1986b). Teacher perceptions of identification criteria emphasized in initial learning disabilities placement. *Learning Disabilities Research, 2*, 21–25.

Van den Bos, K. P. (1989). Relationship between cognitive development, decoding skill, and reading comprehension in learning disabled Dutch children. In P. Aaron & M. Josh (Eds.), *Reading and writing disorders in different orthographic systems* (pp. 75–86). Dordrecht, The Netherlands: Kluwer.

Van der Wissel, A., & Zegers, F. E. (1985). Reading retardation revisited. *British Journal of Developmental Psychology, 3*, 3–9.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the I.Q.-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33,* 223–238.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601–638.

Walberg, H. J., & Tsai, S. (1983). Matthew effects in education. *American Educational Research Journal, 20*, 359–373.

Wallbrown, F. H., Blaha, J., & Vance, H. B. (1980). A reply to Miller's concerns about WISC-R profile analysis. *Journal of Learning Disabilities, 13*, 340–345.

Wallbrown, F. H., Vance, H. B., & Blaha, J. (1979). Developing remedial hypotheses from ability profiles. *Journal of Learning Disabilities, 12*, 557–561.

Watkins, M., & Kush, J. (1988). The research assistant. [Computer program]. Phoenix, AZ: South West Ed Psych.

Weir, M. (1967). Mental Retardation. *Science, 155*, 576–577.

Whipple, G. M. (1922). Educational determinism: A discussion of Professor Bagley's address at Chicago. *School and Society, 15*, 599–602.

White, W. J., & Wigle, S. E. (1986). Patterns of discrepancy over time as revealed by a standard-score comparison formula. *Learning Disabilities Research, 2*, 14–20.

Willson, V. L. (1987). Statistical and psychometric issues surrounding severe discrepancy. *Learning Disabilities Research, 3*, 24–28.

Willson, V. L., & Reynolds, C. R. (1984–1985). Another look at evaluating aptitude-achievement discrepancies in the diagnosis of learning disabilities. *Journal of Special Education, 18*, 477–487.

Wilson, J. D., & Spangler, P. F. (1974). The Peabody Individual Achievement Test as a clinical tool. *Journal of Learning Disabilities, 7*, 384–387.

Wilson, L. R. (1985). Large-scale learning disability identification: The reprieve of a concept. *Exceptional Children, 52*, 44–51.

Wilson, L. R., & Cone, T. (1984). The regression equation method of determining academic discrepancy. *Journal of School Psychology, 22*, 95–110.

Wilson, L. R., Cone, T., Bradley, C., & Reese, J. (1986). The characteristics of learning disabled and other handicapped students referred for evaluation in the state of Iowa. *Journal of Learning Disabilities, 19*, 553–557.

Wilson, L. R., Cone, T. E., Busch, R., & Allee, T. (1983). A critique of the expectancy formula approach: Beating a dead horse? *Psychology in the Schools, 20*, 241–249.

Wilson, W. R. (1928). The misleading accomplishment quotient. *Journal of Educational Research, 17*, 1–10.

Woodbury, C. (1963). The identification of underachieving readers. *The Reading Teacher, 16*, 218–223.

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. E., Algozzine, B., Richey, L., & Graden, J. L. (1982). Declaring students eligible for learning disability service: Why bother with the data? *Learning Disability Quarterly, 5*, 37–44.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *Journal of Special Education, 16*, 73–85.

Ysseldyke, J. E., Christenson, S., Pianta, B., & Algozzine, B. (1983). An analysis of teachers' reasons and desired outcomes for students referred for psychoeducational assessment. *Journal of Psychoeducational Assessment, 1*, 73–83.

Yule, W. (1973). Differential prognosis of reading backwardness and specific retardation. *British Journal of Educational Psychology, 43*, 244–248.

Yule, W., & Rutter, M. (1976). Epidemiology and social implications of specific reading retardation. In R. M. Knights & D. J. Bakker (Eds.), *The neuropsychology of learning disorders: Theoretical approaches* (pp. 25–39). Baltimore: University Park Press.

Yule, W., Rutter, M., Berger, M., & Thompson, J. (1974). Over- and under-achievement in reading: Distribution in the general population. *British Journal of Educational Psychology, 44,* 1–12.

Zigler, E. (1967). Familial mental retardation: A continuing dilemma. *Science,155*, 292–298.

Zigler, E., & Balla, D. (Eds.). (1982). *Mental retardation: The developmental-difference controversy*. Hillsdale, NJ: Lawrence Erlbaum.

Zigler, E., & Hodapp, R. (1986). *Understanding mental retardation*. New York: Cambridge University Press.