American Educational Research Journal
Fall 1983, Vol. 20, No. 3, Pp. 309–331

Characteristics of Pupils Identified as Learning Disabled

LORRIE A. SHEPARD, MARY LEE SMITH, and CAROL P. VOJIR
University of Colorado, Boulder

This study was aimed at describing the characteristics of school-aged children whom educators had identified as learning disabled (LD). A probability sample of 800 was selected from the population of all children served as learning disabled in the state of Colorado. A coding form was used by trained coders to extract relevant features from the case files of the children. The sample was characterized by (1) distributions of single variables (e.g., below grade level achievement, discrepancy between IQ and achievement, medical indicators), and (2) hierarchical classification creating clusters or subgroups within the LD sample. Fewer than half the sample exhibited characteristics consistent with definitions of LD in federal regulations and professional literature. Included in this group were subgroups of hyperactive, brain-injured children, children with significant discrepancies between IQ and achievement and those with signs of perceptual processing disorders. Slightly more than half the sample did not match conventional definitions of LD but exhibited learning problems such as language interference, emotional disorders, or mild retardation. The inclusion of the latter groups among the learning disabled is a particular problem in the validation of the construct and will confound research on prevalence rates and treatment efficacy.

In 1975, the U.S. Congress created an entitlement, the right of the handicapped to a "free, appropriate education." By including the learning disabled among those considered handicapped, Congress required educa-

The research reported in this paper was funded in part by the Colorado Department of Education as part of a larger evaluation study. The authors wish to thank Gene V Glass for his helpful criticism of our writing and both Ann Riley and W. Alan Davis for their extensive efforts in data collection.

tors to identify and serve a type of handicapped child that researchers have so far failed to define. The learning disabled were referred to in governmental regulations as those individuals with severe discrepancies between their achievement and intellectual ability, the discrepancy having been caused by disorders in basic psychological processes (e.g., memory, perception). Those whose underachievement was caused by environmental or cultural deprivation or sensory, mental, or emotional deficiency were excluded from the category.

Although the specifications in the federal law reflect the "consensus definition" (McCarthy & McCarthy, 1969) in the professional literature, they serve poorly as the foundation for a psychological construct. Furthermore, they provide equivocal guidelines for educators to use in determining which children are afflicted and thus entitled to services for the handicapped. Despite the weaknesses of the consensus definition, it provides a standard by which one can compare the characteristics of pupils actually identified as learning disabled by educators.

METHODOLOGICAL WEAKNESSES OF RESEARCH ON LD

Basic research has so far failed to clarify the psychological construct of learning disabilities, their symptoms or origins. Researchers in the 1970s typically selected clinical samples of children already identified as learning disabled and contrasted their characteristics with those of normal or low-achieving children. The search for variables that distinguish these groups was aimed at identifying patterns of behavior (e.g., impulsivity, poor attention span, perceptual disorder) or organic and presumably causal conditions (e.g., cerebral lesion, prematurity, low birth weight) reliably associated with the syndrome.

Ex post facto studies of this type suffer from several methodological limitations:

1. *Absence of appropriate controls* (Werner, 1980). Regression artifacts in matched group designs, which have been noted by Hopkins (1969), confuse normal variation in traits with symptoms of the disorder.

2. *Lack of comparable definitions* (Werner, 1980). Some researchers defined learning disability as significant underachievement in reading and chose samples based on that definition; others thought of it as something more like hyperactivity. Indeed, many researchers indiscriminately mixed what seemed to be meaningful subtypes of learning disability, thereby confounding the results. Routh and Roberts (1972), after partialing out age and IQ, found no common variance in the behavioral symptoms thought to underlie learning disabilities and addressed the inevitable false positives and false negatives encountered in identifying the disorder.

3. *Confounding of the disorder with its identification.* When researchers use clinical samples, they rely on the validity of the original identification.

If the identification was valid, someone (teacher, parent, pediatrician, or psychometrician) had at one time compared the pupil with the definition and found a match. But if the pupil was labeled LD because he was troublesome to his teacher or was merely not living up to his parents' expectations (Coles, 1978; Schrag & Divoky, 1975; Smith, 1982), the identification was invalid. Including the pupil in the LD sample with others who were validly identified nearly guarantees that the empirical study of the construct of learning disabilities will encounter perplexing difficulties.

4. *Biased samples* (Werner, 1980). Even well-designed, matched group studies such as those of Ysseldyke and Warner and their colleagues (e.g., Warner, Alley, Deshler, & Schumaker, 1980; Ysseldyke, Algozzine, Shinn, & McGue, 1979) used accessible samples rather than probability samples from known populations. Results of such studies may be confounded by sample-specific characteristics.

ADVANCES IN RECENT RESEARCH

The work of Ysseldyke and his colleagues (e.g., Ysseldyke et al., 1979) has given us a start in unraveling the tangle of "disorder vs. identification of the disorder." The researchers selected a sample of fourth graders whom the school had labeled learning disabled and contrasted them to a sample of children with similar levels of underachievement but who had not been labeled learning disabled. The two groups were given a battery of psychoeducational tests and their performance was compared.

> An analysis of the results indicated considerable similarities between the two groups; in fact, an average of 96% of the scores were within a common range, and the performance of LD and underachieving children on many subtests was identical. Comparing characteristics of these children with Federal definition showed that as many as 40% of the students may be misclassified. (Ysseldyke et al., 1979, p. i)

Our research explores this discrepancy further, asking the following questions: "What are the characteristics of children identified as learning disabled?" "How well do these characteristics match government and research definitions?" "Are there meaningful subgroups in the school-identified population whose comparison sheds light on the construct of learning disabilities?"

METHOD

The research reported herein was part of a larger attempt to evaluate the identification of learning disabled[1] children in Colorado (Shepard & Smith, 1981).

[1] In Colorado, learning disabilities are called perceptual-communicative disorders. The terms are equivalent both conceptually, and according to specific guidelines for identification.

Sampling and Inferential Statistics and Standard Errors

A stratified, two-stage cluster sampling design was used to sample the population of identified LDs. Of the state's 48 special education administrative units, 22 were sampled at random from strata representing size and type of unit (district or cooperative board serving several districts). All 22 units agreed to participate. The number of units to be selected from each stratum was chosen to keep the number of students in the sample roughly proportional to corresponding population sizes. Exact proportionality was achieved by post hoc weighting. (The sampling frame and weighting procedures are described in Technical Appendix A, Shepard & Smith, 1981.) From the administrative units sampled in the first stage, probability samples of LD case files were selected in the second stage from population lists submitted to the Colorado Department of Education to obtain reimbursement of expenses for services provided to handicapped children. The resulting sample of 1,000 LD cases (3.8 percent of the Colorado LD population) was randomly subdivided into 200 cases to be used in a qualitative study and 800 cases for quantitative analysis. The 800 cases designated for quantitative study (eventually 790 because of clerical errors or logistical problems) are the subjects of this research.

Standard errors in estimating population values were calculated for all the major statistical analyses and were used to establish confidence intervals. The estimation of standard errors was complex because of the multistage cluster sampling design. The derivation of these estimation procedures and the choice between a ratio-to-size estimator or an unbiased estimator (Cochran, 1963) are explained in Technical Appendix A (Shepard & Smith, 1981).

Characteristics Studied

The LD pupils were studied through the medium of their special education files. The files typically included minutes of special education placement staffings and written reports from professionals who participated in the assessment and identification. For the sampled cases the mean number of specialist reports at the time of initial staffing was 3.7, typically from the school psychologist, LD teacher, social worker, and school nurse.

Coding protocols were created to quantify the information found in pupil files. The coded variables included indicators relevant to various clinical and legal definitions of LD. For example, achievement-ability discrepancies were computed for each achievement test and examined for statistical significance.

Variables included both objectively and subjectively coded information. For example, an IQ was recorded for each subject. When more than one IQ was available, coding rules specified which test and test time (at the time of the initial LD placement) was to be given precedence. More

subjectively determined characteristics such as "other plausible causes of learning difficulties" also had coding rules. The variables that relied more on the judgment of the individual coder had relatively more conservative and more stringent guidelines; for example, plausible "environmental causes" for learning difficulties were not coded unless a child had a history of missing about 30 days of school per year, or moving three or four times per year. Furthermore, coders followed the rule not to code characteristics such as "emotionally disturbed" or "not English dominant" if the data in clinicians' reports were minimal or equivocal. Complete definitions and coding rules are available in the technical report (Shepard & Smith, 1981).

The first 100 cases were read again at the end of the data collection period to check for changes in the coders' use of the definitions. A reliability study was done on three case files selected apart from the probability sample. The concordance rate among the three coders (for 242 variables) was 96 percent.

ANALYSIS AND RESULTS

This research determined the characteristics of children identified as learning disabled in two steps: (1) a descriptive single-variable analysis, results of which are summarized in the first section; (2) the more significant contribution of this research, the creation of identifiable subgroups within the LD population. These subgroups are defined by combinations of variables. The hierarchical algorithms for creating subgroups are explained in the second section.

Variables

IQ information. Full scale IQ data were based on the best test available for a child. Tests were ranked as follows: WAIS, WISC-R, WPPSI, Stanford-Binet, PPVT, McCarthy, Slosson, Detroit. A summary of IQ data is presented in Table I. At the time of initial assessment and staffing, 26.8 percent of the 790 LD pupils were placed in LD without any IQ test data; 28.5 percent had IQs below 90 (i.e., below the average range); 8.3 percent had IQs of 80 or below.

For the roughly 60 percent of cases who had a WISC-R test, 43 percent had significant verbal/performance discrepancies. This is about 10 percent more than the expected fraction, one-third, found in the normal standardization sample for the WISC-R (Kaufman, 1976).

Significant IQ and achievement discrepancies. IQ and achievement discrepancies were calculated for each child on every math and reading test and then tested for statistical significance. The data in Table II are reported separately for math and reading achievement and then for the two sets of tests combined. Two different levels of statistical significance were used. The stricter criterion, reported in the bottom panel of Table II,

TABLE I

Percentage of LD Pupils in Colorado by IQ and IQ Discrepancy (n = 790)

Distribution of IQ Scores	Percentage	Standard Error (Percent)	Cumulative Percentage of those with IQ Data
65 and below	0.9	±0.3	1.2
66–70	1.0	±0.3	2.6
71–75	2.2	±0.5	5.6
76–80	4.2	±0.9	11.3
81–85	12.0	±1.3	27.7
86–89	8.2	±1.2	38.9
90–95	10.3	±1.4	53.0
96–100	14.0	±1.5	72.1
101–105	6.9	±1.1	81.6
106–110	6.6	±1.4	90.6
111–115	3.5	±0.9	95.3
116–120	1.7	±0.5	97.7
121–125	1.2	±0.5	99.3
126 and above	0.7	±0.3	100.0
No I.Q. test data	26.8	±3.8	
	100.0		

Significant WISC-R Verbal/Performance IQ Discrepancy	Percentage	Standard Error (Percent)
Significant negative discrepancy[a] (Verbal < Performance)	19.0	±2.1
No discrepancy	33.8	±3.4
Significant positive discrepancy[a] (Verbal > Performance)	6.2	±0.9
No WISC-R administered	41.0	±3.9
	100.0	

[a] Differences greater than 1.96 standard errors of the difference (i.e., $\alpha = .05$) were coded as significant.

TABLE II

Percentage of LD Pupils in Colorado with Significant Discrepancies Between Standardized IQ and Achievement Test Scores

	On Any Reading Test	On Any Math Test	On Any Math or Reading Test
Significant Discrepancy with Weak Significance Criterion ($\alpha = .14$)			
No discrepancies	30.4 ± 2.4	34.2 ± 2.6	25.9 ± 2.0
At least 1 discrepancy	24.4 ± 2.1	16.1 ± 1.7	30.1 ± 2.6
Insufficient data	45.2 ± 3.2	49.7 ± 3.4	44.1 ± 3.2
Significant Discrepancy with Strict Significance Criterion ($\alpha = .05$)			
No discrepancies	35.5 ± 2.6	39.3 ± 2.8	32.6 ± 2.3
At least 1 discrepancy	19.2 ± 1.8	11.1 ± 1.4	23.4 ± 2.2
Insufficient data	45.2 ± 3.2	49.7 ± 3.4	44.1 ± 3.2

Note. Standard errors are reported at the right of the percentages.

is the recommended standard (Salvia & Ysseldyke, 1978). It allows a 5 percent error rate; that is, 5 percent of the IQ and achievement test pairs could be discrepant by chance. The weaker criterion in the middle of the table allows a 14 percent error rate or 1.5 standard errors of the difference between two tests. The choice of a statistical cutoff is arbitrary. The use of two different arbitrary cutoffs shows the effect on number of significant findings; that is, with the less strict criterion at least one significant discrepancy is found for 30 percent of the LD population compared to only 23 percent with the more stringent criterion. The *weak significant discrepancy* variable is used in combination with other possible signs of LD in subsequent analyses.

Forty percent of the LD pupils were either missing an IQ test or an achievement test or were given achievement tests without normative data; in these cases it was impossible for the researchers or the staffing committee to calculate a discrepancy. For example, the Key Math test is used frequently but has neither percentile norms nor standard deviations in the test manual. An additional 4.5 percent of the data were not analyzed because the researchers did not have access to the manuals of some of the more unusual tests.

For pupils with both achievement and IQ data, additional analyses were done to determine how often a significant math or reading discrepancy was confirmed by a significant discrepancy on a second test. Complete data are available in Shepard and Smith (1981). Less than 1 percent of all learning disabled students had significant ($\alpha = .05$) math discrepancies on two math tests, only 4 percent had discrepancies on two reading tests. Of course, only 12 percent and 15 percent of LD pupils had the requisite pair of achievement (and IQ) tests in math or reading, respectively. A more interpretable statistic is that only 5 percent of those who had two math tests had confirmed discrepancies; but 27 percent of those who had two reading tests had significant discrepancies on both. The percentage of replicated discrepancies, given two reading tests, suggests that a real and reliable individual deficiency is being reflected. Nevertheless the small percentage of replications, especially on math tests, indicates that there is still a considerable amount of "fitting chance." The statistical significance criterion is not stringent enough to ensure that an enduring characteristic of the individual is captured.

Years below grade level. Years below grade level data are reported in Table III, which shows the number of cases meeting the criterion for "below grade level" stipulated in the left margin. The distribution of achievement spreads out in the higher grades, so it is increasingly likely that many children will be 1, 2, or even 3 years behind the median grade score. Criteria have been set to increase with grades, corresponding to the cutoffs most often used by clinicians. These typical criteria of academic deficiency do not, however, adequately compensate for increasing varia-

TABLE III

Number and Percent of LD Pupils Who Met a Years-Below-Grade-Level Criterion

Grade Level and Below Grade Level Criterion	Math Achievement n[a]	Math Achievement No. Below Criterion	Math Achievement % Below Criterion	Reading Achievement n	Reading Achievement No. Below Criterion	Reading Achievement % Below Criterion	Language Achievement n	Language Achievement No. Below Criterion	Language Achievement % Below Criterion	Spelling Achievement n	Spelling Achievement No. Below Criterion	Spelling Achievement % Below Criterion
Preschool & Kindergarten												
Criterion = .5												
Years Below Grade	19	1	5	17	1	6	9	2	22	15	1	7
Grade 1												
Criterion = .5												
Years Below Grade	102	39	38	97	27	28	52	16	31	81	21	26
Grades 2–3												
Criterion = 1.0												
Years Below Grade	149	23	15	157	44	28	74	20	27	114	33	29
Grades 4–6												
Criterion = 1.5												
Years Below Grade	161	73	45	170	109	64	68	26	38	126	84	67
Grades 7–9												
Criterion = 2.0												
Years Below Grade	84	54	64	94	69	73	50	27	54	72	60	83
Grades 10–12												
Criterion = 3.0												
Years Below Grade	24	20	83	27	22	81	11	6	55	22	22	100

[a] n's are the number of cases at that grade level who had achievement test data.

bility in higher grades. On standardized group achievement tests, being 1 year behind in grade two places students at roughly the 10th percentile of second graders, while being 3 years behind in grade 11 may be at the 25th or 30th percentile depending on the test and subject area. Thus, these apparently large deficits are not as unlikely as is often thought.

Many LD pupils were not achieving below grade level as measured by standardized tests. In preschool and kindergarten the average for LD pupils was, in fact, above grade level. In grades two and three, the average grade equivalent score was only about 4 months below grade level. In the primary years, fewer than half the cases met the below grade level criterion. In grades four through six, the mean score was about 1.5 years below grade, consistent with slightly more than half the cases meeting the criterion for academic deficit. In the junior high and high school grades, the achievement lags were greater.

Quality of evidence for processing disorders. All the tests available for assessing perceptual or psychological processing abilities are known to have low reliability (at least below the standard sometimes recommended for individual diagnosis) and to have limited evidence of validity (see Shepard & Smith, 1981; Thurlow & Ysseldyke, 1979). In the absence of psychometrically adequate tests, evidence of a processing disorder is more credible when independent confirmation of the nature of the disability is obtained from more than one fallible source. Corroboration of this sort compensates in part for the unreliability of each separate observation or measurement. Two ratings of consistency in clinicians' conclusions served as indices for the strength of evidence leading to the diagnosis of a processing disorder for each LD case.

On the first rating scale, 26 percent of the LD cases had at least some agreement between professionals on the nature of the processing disorder: that is, at least two professionals said the problem was in the same general area, such as auditory perception, although other unconfirmed problems might also have been cited. "At least some confirmation" is used in later analyses of subgroups as the *medium quality processing deficit* variable. For 11 percent of the LD cases, different processing disorders were cited by different professionals (e.g., visual problems by the psychologist and memory deficits by the LD teacher), but there were no confirmations. Five percent of the cases had contradictory evidence, that is, what was thought to be a deficit by one clinician was cited as an area of strength by another. A large percentage of cases, 39 percent, had a processing problem cited by only one clinician, which was therefore not confirmed; 19 percent did not have a processing disorder cited by any clinician.

The second rating scale is a measure of the congruence between individual professional reports and the basis of handicap cited in the staffing minutes. The ratings reflect the degree to which professionals sought confirmation of the inferred disability and attempted to reconcile and

integrate the conclusions and observations of various members of the staffing team. For example, the diagnostic conclusion received a high rating for "congruence" if the deficits listed as the basis for LD identification were those that had been confirmed by at least two professionals. This rating reflects to what extent staffing teams imposed on themselves some criterion for consistency (like our first scale) to decide what evidence to give the most weight. The three highest ratings describe only 7 percent of the LD population; in these cases the "basis of handicap," the summary diagnostic statement from the staffing minutes, reflected a coherent picture of the child's intellectual functioning put together from the several separate clinicians' reports. The more prevalent practice, however, was for the staffing minutes to include all possible deficits observed by any clinician, with no attempt to reconcile inconsistent conclusions or seek confirmation; this occurred in 23 percent of the cases. Finally, a composite variable was created from the two consistency ratings. A *high quality processing deficit* required at least some agreement on the nature of the disorder from individual reports (on scale 1) plus some congruence between the strength of evidence in the individual professional reports and the final diagnostic conclusion (scale 2). Staffing minutes, which merely listed all possible areas of cognitive dysfunction without attempting to confirm or disconfirm, were not counted as high quality.

Medical indicators, other handicaps, and other sources of learning problems. The remaining single variables are not by themselves essential to the definition of LD. Some of the more serious medical signs may contribute to diagnosis of LD; for example, 4.5 percent of the LD cases were identified as hyperactive by their physicians as well as classroom teachers (pupils who were called hyperactive only by their classroom teacher were not counted). Brain injury, diagnosed by physician or nurse and based on specific neurological evidence, was reported for 1 percent of the LD population. Other variables provided additional, secondary descriptors of the LD population, such as the 31 percent who have minor behavioral problems (defined as consistent reports of poor attention span, frustration with work or poor self-concept). Still other variables are important to the exclusionary clause of the LD definition; that is, they refer to characteristics that should be ruled out as the primary cause of learning problems before concluding that a child is LD (e.g., mental retardation, vision and hearing handicaps, emotional disturbance, and environmental, cultural, or economic disadvantages). A small proportion of LD cases had consistent evidence in their files that would make it more appropriate to classify them in another category of handicap; 8.4 percent had IQs below 80 or had previously been classified as educable mentally retarded (EMR); 5.9 percent had strong evidence of emotional disturbance (not secondary to the learning problem); 1.2 percent and 0.5 percent, respectively, were known to

have serious hearing or seeing handicaps. Other variables reflected fairly serious sources of learning problems but definitely did not imply a handicap (e.g., 9.5 percent of the LD cases had severe environmental learning problems such as consistently missing more than thirty days of school each year, and 3.4 percent were not English dominant).

Identification of Subgroups

The preceding comparison of LD pupil characteristics with each definitional element separately gives only a crude and potentially misleading picture of the validity of LD placements. First, the counts of cases who meet all, only one, or none of the criteria cannot be inferred from the separate analyses. At one extreme, if the cases who met one requirement were the same ones who satisfied each of them, there would be one group of highly valid (and likely severe) LD placements and another rather large group who satisfied none of the criteria. At the other extreme, if every case met only one criterion, the percentages from the separate analyses would add to very nearly 100 percent of all the cases who satisfied one or another of the eligibility criteria. The defining characteristics must be studied simultaneously to see the overlap in criteria supporting placement. Furthermore, only by considering indicators in combination is it possible to see if the pattern of signs justifies placement in LD even when no one indicator is significant by itself.

The LD population is not believed to be a homogeneous group (Hammill, Leigh, McNutt, & Larsen, 1981). The purpose of the subgroup analysis was to describe more accurately the LD population by reflecting different clinical criteria, profiles, and combinations of signs used to identify individuals as LD. Furthermore, because clinicians sometimes report that they ignore both legal and clinical definitions to obtain services for a particular child (Smith, 1982), the purpose of the subgroup analysis was also to identify the salient characteristics of non-LD cases currently served in the LD category. These are the cases that have confounded previous efforts to study the collateral symptoms of LD.

Hierarchical Algorithm. An algorithm was devised to sort the 790 coded cases into mutually exclusive subgroups on the basis of each child's most salient characteristics. The decision rules which governed the analysis are described with the results for each group. The decision rules were derived from several sources: the professional literature on LD identification, a survey of identification practices among a representative sample of LD teachers, school psychologists, speech language specialists, social workers, and principals in Shepard and Smith (1981), and a reading by the first authors of a randomly parallel sample of 200 LD cases (obtained when the original sample of 1,000 cases was subdivided for quantitative and qualitative analyses). By using combinations of variables and hierarchies of

inclusion and exclusion rules, it was possible to "model" the reasoning and combination of signs used in clinical interpretations.

The analysis was termed "hierarchical" because some variables were considered more important than others in deciding the category to which a case should be assigned. For example, a child could have a highly significant discrepancy between ability and achievement but also have several reports that he is emotionally disturbed with severe behavior problems. If the emotional disorder were sufficient to explain the discrepancy and there were no other signs of a learning disability, the emotionally disturbed classification would take precedence.

The hierarchical algorithm is illustrated in Figure 1. The order of categories represents the strength of the defining characteristics. LD pupils were assigned to the highest category first if they met the decision rule; only the remaining cases were considered for inclusion in subsequent clusters. Therefore, some of the cases in the early subgroups may also have some characteristics like pupils in later groups, but they are distinguished by a more prominent variable on the basis of which they were assigned to the prior category.

Most of the identifiable clusters within the LD population were created by the combination of two or three variables.

Language Interference. First, pupils were assigned to this group if they were *not English dominant.* This variable was coded only if the child's lack of fluency in English was judged severe enough to cause the child's academic problems. Mere existence of a Spanish surname or evidence of some fluency in a language other than English was not sufficient to conclude that a child was not English dominant.

Pupils were also classified as LANGUAGE INTERFERENCE if they were Chicano or Indian and had a significant *verbal/performance IQ discrepancy.* Many Chicano and Indian children in the sample did not meet this criterion; but those who had a verbal IQ score significantly below their performance IQ were believed to have a language or cultural background that was sufficient to explain their learning difficulties in the school setting.

Finally, however, any of the above children were excluded from this category if they had a high quality processing deficit. For most children who met the above criteria (for non-English dominance or depressed verbal performance due to cultural and language background), language interference was thought to be a better description of the nature of their learning problems than LD. LANGUAGE INTERFERENCE was held to be the appropriate classification even if a child had a significant IQ/achievement discrepancy thereby satisfying the federal criteria for LD. For these cases, the discrepancy would plausibly be caused by the second language or cultural influences. In fact, one-fifth of the cases eventually counted in this

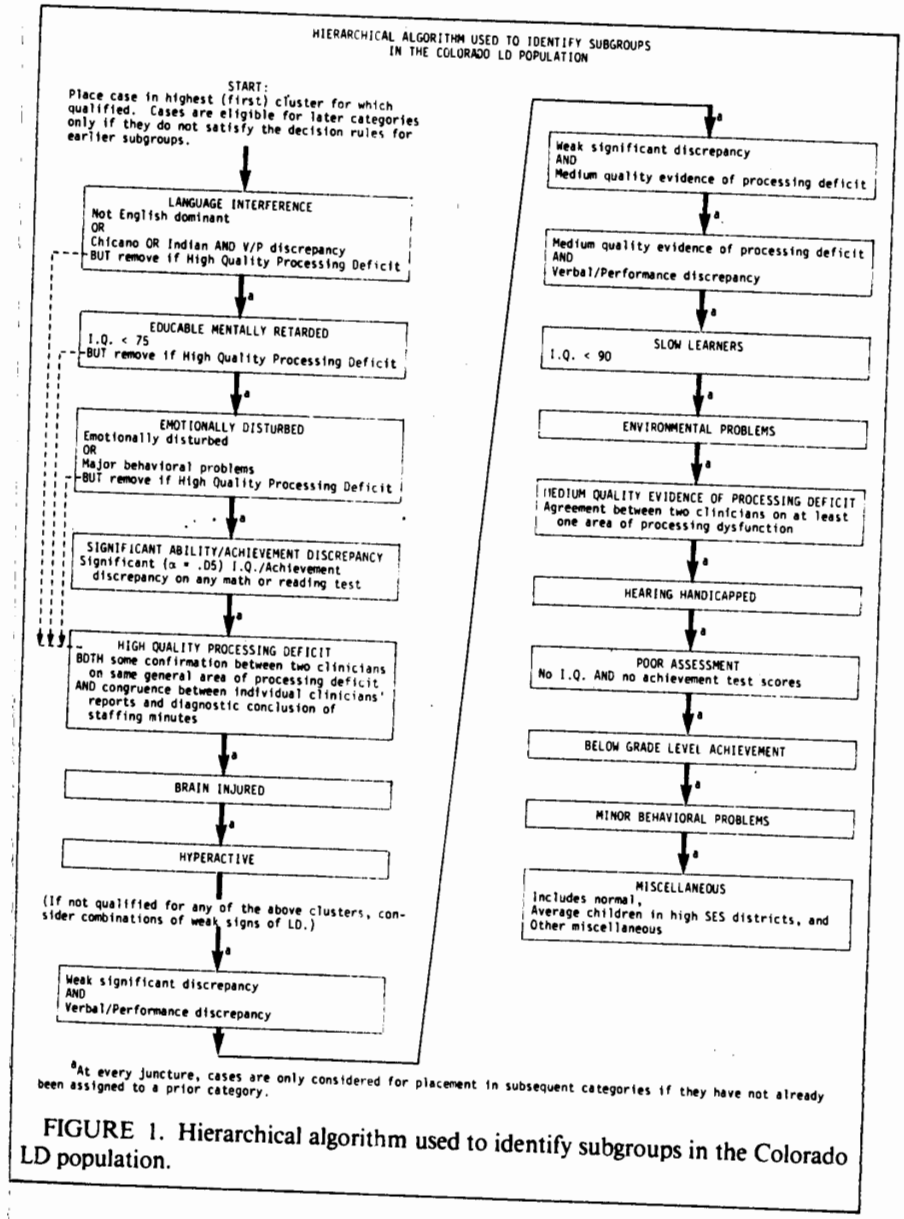

FIGURE 1. Hierarchical algorithm used to identify subgroups in the Colorado LD population.

group had a significant discrepancy on at least one achievement test. However, the authors did not rule out the possibility that a child might not be English dominant but also have a learning disability. Hence, the algorithm excluded from this cluster children with confirmed evidence of a processing disorder.

The percentage of the LD population estimated to be in the LANGUAGE INTERFERENCE cluster was 6.6 percent. This figure is probably an underestimate since children's ethnic group was often not known and because stringent criteria were used for the coding of non-English dominance.

Educable Mentally Retarded (EMR). Children whose full-scale IQ scores were 75 or less were placed in this cluster. However, because a learning disability caused by a perceptual or processing dysfunction could substantially reduce performance on an IQ test, cases were removed from this category if they had a high quality processing deficit and were assigned to a later learning disabled category.

If pupils satisfied both the LANGUAGE INTERFERENCE and the EMR criteria, they remained in the LANGUAGE INTERFERENCE group because this group had logical priority (i.e., the language problem could also depress the IQ score; 2.2 percent of the LD population were in this relatively extreme EMR category).

Emotionally Disturbed. When the individual files were read and coded, cases were identified as EMOTIONALLY DISTURBED if reports from various clinicians indicated that the child's emotional or behavioral problems were severe enough to qualify for placement in the emotionally or behaviorally disturbed handicapped category. Coders tended to be conservative in judging a case to be emotionally disturbed. Evidence had to be clear-cut, otherwise the assumption was made that the clinicians believed that the emotional problem was secondary to the learning disorder. In addition, pupils who had *major behavioral problems*, sufficient to account for their poor achievement, were counted in this category.

Again, as illustrated in Figure 1, pupils who satisfied one of these criteria but had a high quality processing deficit were excluded and placed in a later LD category, because in that case the observed behavioral disorder could well be the result of the learning disability. In the absence of this type of evidence, however, it is more plausible to attribute the poor academic functioning and even a significant discrepancy between ability and achievement to the emotional or behavioral disorder. One-fifth of the cases eventually counted as emotionally disturbed had a significant discrepancy on at least one achievement test.

Of the LD population, 7.5 percent were classified as EMOTIONALLY DISTURBED. An additional 1.0 percent of the population had severe behavioral or emotional disorders but were already counted in either the EMR or LANGUAGE INTERFERENCE categories.

Significant Ability/Achievement Discrepancy. After eligible pupils had been assigned to the above categories, the remaining cases were searched and any pupil with a significant IQ/achievement discrepancy ($\alpha = .05$) on any math or reading test was assigned to the SIGNIFICANT ABILITY/ACHIEVEMENT DISCREPANCY category. The level of statistical significance used is the standard criterion proposed in the literature (Salvia & Ysseldyke, 1978) and is one which in the long run would allow 5 percent of the cases to have this great a difference just by chance.

Of the LD population, 20.5 percent were in the SIGNIFICANT DISCREPANCY group. A small proportion (4 percent) of those pupils with a strict significant discrepancy also had high quality evidence of a processing deficit. They remain in this category, however, because the conjoint characteristics simply mean these cases have two strong signs of a learning disability.

High Quality Processing Deficit. The definition of a high quality processing deficit derives from the two consistency ratings described as single variables in the preceding section. Ratings that contributed to the HIGH QUALITY designation were those that reflected any attempt to reconcile reports from various professionals or the tendency to include statements of deficits in the placement decision that had some confirmation.

Of the LD cases, 4.7 percent were classified as HIGH QUALITY PROCESSING DEFICIT. Because this group was created after the SIGNIFICANT DISCREPANCY group in the analysis hierarchy, it is clear by definition that none of these cases has a significant IQ/achievement discrepancy. Nevertheless, it is very likely that they are validly identified LD cases because the criteria for independent clinical confirmation of a processing deficit were stringent. Some clinical alternative to the statistical decision rule has to be permitted because it is likely that a learning disability could depress measured IQ and preclude a significant discrepancy. It should also be noted that only half the total LD cases could be rated for congruence between individual professionals' reports and the final staffing diagnosis because staffing minutes or rationale were missing for half the cases. These pupils were then automatically ineligible for the HIGH QUALITY PROCESSING evidence category because of missing data. They were, however, counted in a subsequent group, MEDIUM QUALITY PROCESSING DEFICIT, if two clinicians confirmed an area of deficit.

Brain Injured. LD children who had hard signs of brain injury reported by a physician and had not met the criteria for the two previous LD categories were placed in a special BRAIN INJURED category of learning disabilities. This group accounted for 0.6 percent of the LD population.

Hyperactive. Children who did not qualify for assignment in any of the above clusters and who had medical reports of hyperactivity were identified in this group. Often this meant they were taking Ritalin. In a very few instances, confirming reports from two or more clinicians would qualify

in place of a physician's diagnosis of hyperactivity. However, classroom teachers' reports of hyperactive behavior were not counted without a physician's corroboration.

Of the LD population, 2 percent fell in this HYPERACTIVITY category. This percentage may be contrasted with the 4.6 percent who were coded as hyperactive in the single variable analysis. More than half of cases with this characteristic have already been assigned to prior subgroups (e.g., EMR, other LD categories, or emotionally disturbed). The 2 percent assigned to the hyperactive subgroup did not meet the criteria for any of the prior clusters and are therefore best characterized as hyperactive.

Weak Significant Discrepancy and Verbal/Performance IQ Discrepancy. All the above clusters are based on fairly strong indicators that the children were or were not learning disabled. Up to this stage in the analysis these highly interpretable and reliable variables accounted for 44.5 percent of the LD population, placing them in either the learning disabled categories, other handicaps, or the language interference cluster. After all possible assignments had been made to these subgroups, the remaining cases were examined for combinations of weaker indicators that would account for their LD placement.

The statistical criterion for determining the reliability of the difference between IQ and achievement was relaxed from 1.96 to 1.5 standard errors of the difference, corresponding to a shift in the percentage of differences occurring by chance from 5 percent to 14 percent. Pupils who met this lowered cutoff were said to have a *weak significant discrepancy.*

Because this more liberal level of significance is not generally recommended in practice (Salvia & Ysseldyke, 1978; Thorndike & Hagen, 1977), it was used here in conjunction with a second indicator, a significant verbal/performance IQ discrepancy, which is also not considered a valid sign of a learning disability in and of itself (Kaufman, 1976). Together, however, these two indicators help to rule out the possibility that either discrepancy occurred by chance, and they begin to suggest a pattern of irregularities or discrepant strengths and weaknesses in learning processes. Those satisfying this pair of requirements were 3.6 percent of the LD cases.

Weak Significant Discrepancy and Medium Quality Processing Deficit. Following the same reasoning as above, a weak significant discrepancy was also combined with a medium quality processing deficit. Cases were said to have medium quality evidence of a processing deficit if there was some confirmation of any of the deficits cited by various professionals, but not necessarily a correspondence between the confirmed deficits and what was reported to explain the placement decision. In other words, cases that were not eligible for the HIGH QUALITY PROCESSING DEFICIT category but did have a medium quality processing score were considered for this group. Cases with medium quality evidence of a processing disorder plus

a weak significant discrepancy accounted for 1.1 percent of the LD population.

Medium Quality Processing Deficit and Verbal/Performance IQ Discrepancy. A third category of weaker indicators was created by requiring both a medium quality processing index and a significant verbal/performance IQ discrepancy. Again, the two indicators together more strongly suggest that a problem exists, where a single indicator would be less compelling. For example, the 12-point difference required to ensure that the verbal/performance discrepancy is reliable does not guarantee its validity as a sign of learning disabilities. Fully one-third of the normal population in the standardization sample for the WISC-R had a difference this large or larger (Kaufman, 1976). However, this discrepancy plus two or more clinicians agreeing on an area of deficit make the diagnosis of LD more plausible. The percentage of the Colorado LD population in this group was 6.6 percent.

Slow Learners. The raw data reported in Table I show that 28.5 percent of the LD population have IQ scores below 90, suggesting that they do not have average ability and hence may be behind in school simply because of their generally lower intellectual ability. We did not call this entire group slow learners, however, because we wanted the computer sorting to reflect the reasoning that obtained IQ scores could have been lowered by the presence of a learning disability. Therefore, the slow learner category was only entered in the hierarchy after all the strong and weak indicators of learning disorders had been entered; a child who met both criteria would be counted in the LD group. SLOW LEARNERS were defined as cases with IQs less than 90 who also had none of the preceding signs of a processing disorder or significant discrepancy. The decision rule to exclude, from the SLOW LEARNER cluster, those with processing deficits gives the benefit of the doubt to clinical judgments and underestimates the number of true slow learners. Clinicians often agree that a child has a processing deficit when scores are below age norms; they do not, however, consider whether the low processing scores are uniformly low and commensurate with low IQ, therefore making it more likely that the child is a slow learner and not learning disabled. We did not try to introduce this consideration into the analysis. Using the above decision rules, 11.4 percent of the LD population were placed in the SLOW LEARNER category.

Environmental Problems. When LD files were originally read and coded, cases received an "environmental" code if the number of times the children had changed school or the number of absences was so extreme as to be a cause of serious learning problems. In the single variable analysis, 9.5 percent of the LD population were identified with severe problems of this type. At this stage in the hierarchy, a case with detrimental environmental circumstances that had not been selected for any previous category was placed in the ENVIRONMENTAL PROBLEMS group. For these children

without other signs for learning disabilities or other handicaps, who also were not slow learners, the very best explanation of their learning problems is a lack of opportunity to learn. Identified in the ENVIRONMENTAL PROBLEMS subgroup were 2.2 percent of the LD population.

Medium Quality Processing Deficit. Finally, after other stronger and more reliable indicators had been used, remaining cases that had only MEDIUM QUALITY PROCESSING DEFICIT evidence were placed in a cluster. These cases had no ability-achievement discrepancy but were cited by at least two clinicians as having a particular processing deficit. This MEDIUM QUALITY PROCESSING group was 3.5 percent of the LD population.

Hearing Handicapped. Children who were reported to have severe hearing loss and who had not been selected for any previous category were placed in this group; they represent only 0.2 percent of the LD population. An additional 1.5 percent of the LD population had been identified as *hearing* or *vision handicapped* but these cases also had other characteristics that allowed them to be placed in previous categories.

Poor Assessment. Cases in this group were missing both IQ tests and achievement tests. This meant that neither we nor the clinicians involved in staffing could judge whether achievement was significantly behind expectancy. Furthermore, if perceptual or processing tests had been given, the clinical reports did not have enough consistency even to qualify for the medium quality processing deficit rating used to create a prior cluster. The POOR ASSESSMENT category accounted for 6.4 percent of the LD cases.

Below Grade Level Achievement. The BELOW GRADE LEVEL category is a hodgepodge of cases that did not fall into any of the previous clusters. Using the cutoffs reported in Table III for below grade level performance, cases with lagging achievement were counted in this category. This subgroup accounted for 6.1 percent of the LD population.

To understand the characteristics of this cluster, one has to recall the criteria for previous categories. These pupils all had at least one achievement test on which they were below grade level. One-fourth of this group had IQ tests and did not qualify for any of the significant discrepancy clusters or the SLOW LEARNER group. Most of these cases had IQ scores in the 90 through 95 range, so their below average achievement was consistent with their slightly below average IQ. It should be remembered that, although the below grade level cutoffs were selected to reflect typical decision rules followed by clinicians, on a typical test (the PIAT, for example) these cutoffs correspond to the 10th percentile (for second graders on PIAT math), 12th percentile (for fifth graders on PIAT math), 26th percentile (for eighth graders on PIAT math), and 26th percentile (for 11th graders on PIAT math). An IQ of 90 is at the 25th percentile; an IQ of 95 is at the 37th percentile. Computations that allow for unreliability in the

scores would require either that achievement be far below the cutoff or that IQ be well above 90; hence these cases did not have significant discrepancies.

Three-fourths of the BELOW GRADE LEVEL group did not have IQ test scores. By comparing the distribution of achievement scores for this subgroup to the distribution for those with IQ scores, it is possible to estimate how many cases are like the first group, below grade level but not significantly below ability. This accounts for most of the subgroup. The remaining cases, accounting for less than 1 percent of the LD population, are those who might have had significant discrepancies if a measure of ability had been administered. All together the BELOW GRADE LEVEL group accounts for 6.1 percent of the LD population.

Minor Behavior Problems. Children who did not have any of the above indicators of learning disabilities or academic problems and who had minor behavior problems were counted in this category. During the coding of case files, researchers coded *minor behavioral problems* if the classroom teacher and clinicians had consistently noted problems such as poor attention span, inability to stay in seat, frustration with work. Because this category was entered last in the hierarchy, this 3.7 percent of the LD population have minor behavior problems as their only defining characteristic.

Miscellaneous. A residual group, 10.6 percent of the LD population, could not be classified by the above decision rules. Some of these children might have been classifiable if the data had been better or if the analysis had been more sensitive. The weight of evidence suggests, however, that these LD cases could not be categorized because they did not have any of the signs of the disorder, not because the analysis was inaccurate. These 11 percent of the LD population had no significant discrepancy, were not below grade level, had no consistent evidence of a processing disorder, did not have IQs below 90, could not be described as emotionally disturbed or hearing handicapped, and did not have even the weaker indicators of learning disabilities. Furthermore, these cases were not generally "missing data" cases. If they had been missing IQ and achievement test scores, they would have been classified in the POOR ASSESSMENT group. Evidence was available for these cases but was not consistent with a learning disabilities diagnosis. When cases in the miscellaneous category were listed and inspected, most looked like normal children. One hypothesis was that these were children of average ability or slightly below average ability (IQs above 90) who appeared deviant in their particular school context. Of the five high socioeconomic status (SES) districts in the study, three had disproportionately high percentages in this last category. These three districts alone accounted for one-third of all the miscellaneous cases (although their enrollments are only 17 percent of all sample enrollments)

or 3.4 percent of the total LD population. All five high SES districts together accounted for 55 percent of the category or 5.8 percent of the LD population.

Summary of Identifiable Subgroups in the LD Population. The percentages of LD cases in each of the computer-identified clusters are summarized in Table IV. The data have been reorganized into four major categories: other handicapping conditions, which includes 10 percent of the LD population; the learning disabled subgroups, 43 percent of the LD population; children with other learning problems, 30 percent of the LD cases; and an "other" category that includes poor assessments, normal children, and other miscellaneous, 17 percent of the LD cases.

CONCLUSIONS

This research was aimed at describing the characteristics of pupils whom educators had identified as learning disabled. A representative sample of a defined population was studied. Fewer than half the sample had characteristics that are associated in federal law and professional literature with the definitions of learning disabilities. Included in this category were pupils characterized as hyperactive, brain injured, those with statistically significant discrepancies between ability and achievement and both strong and weak signs of perceptual processing disorders. This list may comprise the first approximation for a system of subcategories within the general construct of learning disabilities.

Besides this meaningful subgrouping of learning disabilities, researchers must attend to the large number of pupils who have been identified as LD but who do not have characteristics conforming to the definitions. Included are those with mild mental retardation, emotional disturbance and other miscellaneous handicaps, those whose low achievement is due to language interference, and those with low achievement due to other causes than psychological disabilities. Though most of these pupils have learning problems, they are incorrectly called learning disabled. No logical or linguistic analysis of the latter term can be stretched to cover them. The incidence of misidentification revealed in the quantitative analysis of pupil cases was confirmed in a separate, qualitative analysis and in the survey of professionals (Shepard & Smith, 1981) and corroborates the research of Ysseldyke et al. (1979).

The implication of these results for basic research on learning disabilities is that the label applied for the purpose of providing services cannot be assumed valid. If the label is taken as a dependable sign of the disability, then research on LD heritability patterns, prevalence rates, and the effectiveness of interventions will be confounded. The meaning or meanings of learning disabilities will remain elusive.

Educators identify pupils as learning disabled for a variety of reasons,

TABLE IV

Identification of Subgroups in the Colorado LD Population Presented in Major Categories

	Percentage of LD Cases	Standard Error
Other Handicaps		
EMR	2.6	±0.6
Emotionally Disturbed	7.5	±1.0
Hearing Handicapped	0.2	±0.2
	10.3	
Learning Disabilities		
Significant Ability/Achievement Discrepancy	20.5	±2.0
High Quality-Processing Deficit	4.7	±0.8
Brain Injured	0.6	±0.3
Hyperactive	2.0	±0.6
Weak Significant Discrepancy and Verbal/Performance Discrepancy	3.6	±0.6
Weak Significant Discrepancy and Medium Quality Processing Deficit	1.1	±0.4
Medium Quality Processing Deficit and Verbal/Performance Discrepancy	6.6	±1.2
Medium Quality Processing Deficit only	3.5	±0.8
	42.6	
Other Learning Problems		
Language Interference	6.6	±1.0
Slow Learners	11.4	±1.4
Environmental Causes	2.2	±0.6
Below Grade Level Achievement	6.1	±1.0
Minor Behavioral Problems	3.7	±0.8
	30.0	
Other		
Poor Assessment (no IQ *and* no achievement tests)	6.4	±1.3
Miscellaneous (including normal)		±1.3
From high SES districts	5.8	
Other miscellaneous	4.8	
	17.0	

some of which are unrelated to the traits of the pupils. Schrag and Divoky (1975) argued that educators invented the term to provide a medical-sounding excuse for their failure to educate all pupils. Our survey (Shepard & Smith, 1981) suggested that professionals identify those pupils in need of remedial and support services as LD. There appeared also to be a lack of understanding among professionals about definitions and diagnostic indicators. Smith (1982) presented a list of influences on identification practices. Among them were ambiguities of definitions, unreliable and invalid measures used in diagnosis, bureaucratic pressures to find and serve all possibly handicapped pupils, parental pressure to secure remedial services, and the pressure to adjust demand to supply of services. These

social forces may contribute to misidentification and need to be taken into consideration by the researchers seeking validation of the construct of learning disabilities.

REFERENCES

COCHRAN, W. G. *Sampling techniques.* New York: Wiley & Sons, 1963.

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

HAMMILL, D. D., LEIGH, J. E., MCNUTT, G., & LARSEN, S. C. A new definition of learning disabilities. *Learning Disabilities Quarterly*, 1981, *4*, 336–342.

HOPKINS, K. D. Regression and the matching fallacy in quasi-experimental research. *Journal of Special Education*, 1969, *3*, 329–336.

KAUFMAN, A. S. Verbal-performance IQ discrepancies on the WISC-R. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 739–744.

MCCARTHY, J. J., & MCCARTHY, J. F. *Learning disabilities.* Boston: Allyn and Bacon, 1969.

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

SALVIA, J., & YSSELDYKE, J. E. *Assessment in special and remedial education.* Boston: Houghton-Mifflin, 1978.

SCHRAG, P., & DIVOKY, D. *The myth of the hyperactive child.* New York: Dell Publishing, A Laurel Edition, 1975.

SHEPARD, L., & SMITH, M. L., WITH DAVIS, A., GLASS, G. V, RILEY, A., & VOJIR, C. *Evaluation of the identification of perceptual-communicative disorders in Colorado.* Boulder, Col.: Laboratory of Educational Research, University of Colorado, February 1981.

SMITH, M. L. *How educators decide who is learning disabled: Challenge to psychology and public policy in the schools.* Springfield, Ill.: Charles C Thomas, 1982.

THORNDIKE, R. W., & HAGEN, E. P. *Measurement and evaluation in psychology* (4th ed.). New York: Wiley & Sons, 1977.

THURLOW, M. L., & YSSELDYKE, J. E. *Current assessments and decision-making practices in model programs for the learning disabled* (Research report No. 11). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

WARNER, M. M., ALLEY, G. R., DESHLER, D. D., & SCHUMAKER, J. B. *An epidemiological study of learning disabled adolescents in secondary schools: Classification and discrimination of learning disabled and low-achieving adolescents* (Research Report No. 20). Lawrence, Kan.: The University of Kansas Institute for Research in Learning Disabilities, 1980.

WERNER, E. E. Environmental interaction in minimal brain dysfunctions. In H. E. Rie & E. D. Rie (Eds.), *Handbook of minimal brain dysfunction.* New York: Wiley & Sons, 1980.

YSSELDYKE, J. E., ALGOZZINE, B., SHINN, M., & MCGUE, M. *Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers* (Research Report No. 13). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities, 1979.

AUTHORS

LORRIE A. SHEPARD, Associate Professor and Chairman, Research and Evaluation Methodology, School of Education, University of Colorado, Boulder, CO 80309. *Specializations:* Measurement and evaluation.

MARY LEE SMITH, Associate Professor, Department of Research and Evaluation Methodology, School of Education, University of Colorado, Boulder, CO 80309. *Specializations:* Naturalistic research and evaluation; research integration; psychotherapy outcome research.

CAROL P. VOJIR, Department of Research and Evaluation Methodology, School of Education, University of Colorado, Boulder, CO 80309. *Specializations:* Medical technology; research and evaluation in allied health professions; data analysis.