Classification and Dynamic Assessment of Children with Learning Disabilities

H. Lee Swanson

The number of children classified as learning disabled has increased substantially over the last 20 years. In 1976, 783,000 children were identified with learning disabilities (LD). By 1992-93, the LD population topped 2.5 million. These children currently comprise almost half of all the placements into special education (U.S. Office of Education, 1994). Approximately 120,000 students each year are tagged as having learning disabilities, a number equal to all Americans who have contracted AIDS, hepatitis, and tuberculosis combined in 1995. Based on these figures, one could argue that the classification of children with learning disabilities is epidemic.

As Roush (1995) stated in "Arguing Over Why Johnny Can't Read," an article published in *Science*, "if learning to read and write or do math at expected levels were a disease, then American school children would be in the middle of an epidemic" (p. 1986). This is a costly epidemic because public schools spend approximately $8,000 a year to educate an LD student, compared to $5,500 for a nondisabled student. One direct means to deal with this epidemic is to improve the classification of children suspected of having learning disabilities.

This chapter focuses on the application of dynamic testing procedures to the classification of children suspected of LD. Two important assumptions about such children are

1. LD is first and foremost an information-processing deficit and, therefore, decisions about LD should be made at the processing level.
2. Current psychometric procedures misclassify several children as LD because they fail to adequately (a) assess processing potential, and (b) differentiate these children from other low achievers.

Dynamic testing provides an estimate of the stability of information-processing potential in children suspected of LD, which in turn differentiates children with LD from other low achieving children.

This article will outline problems related to defining and classifying children with LD, as well as describe the pitfalls of current assessment procedures. An understanding of these pitfalls will place dynamic assessment in a more critical light. It will review the practical outcomes of two dynamic assessment studies related to the classification of LD.

Dr. Swanson is affiliated with the School of Education at the University of California, Riverside.

WHAT IS A LEARNING DISABILITY?

Before discussing issues related to the classification of children suspected of having LD, some understanding of how to define these children is necessary. Most studies published in psychological, educational, and neuropsychological journals define a person with a learning disability as an individual with at least average ability to process and retrieve some information from their environment who *unexpectedly* have difficulty performing at their age level on specific cognitive or academic tasks.

These individuals exhibit performance on general intelligence tests comparable to children their own age, but they have problems on academic tasks (reading, mathematics) that make heavy demands on either specific low-order processes (e.g., nonword reading tasks that rely on phonological codes) or higher-order processes (e.g., problem solving tasks that rely on metacognition). The discrepancy between their average ability to retain and retrieve global information about their environment on intelligence tests and their impairment on isolated processes is assumed to be a result of a cognitive/neurological constraint or inefficiency. This constraint or inefficiency may manifest itself in specific lower-order processing skills (e.g., attention, phonological coding), specific processes of a higher-order system (e.g., awareness and monitoring of attentional resources), or in processes that facilitate the interaction between or coordination of lower- and higher-order processes (see Swanson, 1988, for a review).

Now contrast these above assumptions with a popular definition of LD. In 1981, the National Joint Committee for Learning Disabilities (NJCLD), defined a learning disability as follows:

> *Learning disabilities* is a generic term that refers to a heterogenous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual and are presumed to be due to central nervous system dysfunction. Even though a learning disability may occur commitantly with other handicapping conditions (e.g., sensory impairment, mental retardation, social, or emotional disturbance) or environmental influences (e.g., cultural differences, insufficient/inappropriate instruction, psychogenic factors), it is not the direct result of these conditions or influences. (Hammill, Leigh, McNutt, & Larsen, p. 336)

Although this definition has been evaluated positively against others (e.g., Interagency Committee on Learning Disabilities, 1987, USOE, 1977; e.g., see Hammill, 1990), it is not an operational definition. Further, an even larger epidemic in the overidentification of LD likely would emerge if school districts were to classify children solely on basis of the NJCLD definition. This is because the definition does not specify the operations or procedures by which LD can be recognized and measured.

A conservative strategy for identifying children as LD is to restrict the range of operations narrowly. These operations most likely would reflect the current assumptions about the basis of the disabilities: (a) Learning difficulties are not the result of inadequate opportunity to learn, general intelligence, or significant physical or emotional disorders but, rather, to basic disorders in psychological processes; (b) these processing deficits are a reflection of neurological, constitutional, or biological factors; and (c) the deficit depresses only a limited aspect of contextually appropriate behavior. Thus, when attempting to classify children suspected of LD, a concerted effort is made to (a) establish adequate opportunity to learn (documentation that optimal instruction has been presented but deficits in isolated processes remain), (b) establish that a specific processing difficulty is significantly inferior to general academic aptitude, and (c) establish that neuropsychological inefficiencies are reasonably correlated with cognitive processing difficulties. Further attempts in the classification

process also would include establishing a range of cut-off scores on standardized intellectual, neuropsychological, and information-processing measures.

One example of an attempt toward a concise definition is outlined by Morrison and Siegel (1991). They defined children with potential LD as those who (a) have WISC III IQ scores equal to or above 85 and (b) have one or more of the following: (1) reading subtest scores equal on the Wide Range Achievement Test (WRAT-R) to below the 25th percentile, (2) arithmetic subtest scores on the WRAT-R equal to or below the 25th percentile and WRAT reading score equal to or above the 30th percentile, and/or (3) a score greater than two standard deviations above the mean on a standardized attention rating scale (e.g., Conners Parent and Teacher Rating Scales) as LD.

This operationalization captures three high-incidence disorders within LD: reading (word recognition), arithmetic (computation, written work), and attention. No doubt, there are numbers of positive (e.g., easy identification for a broad constituency of researchers, replication of classification is assured) and negative (e.g., construct validity, conceptual integrity) aspects to these procedures in classifying children with potential LD.

CURRENT ASSESSMENT PROCEDURES

Although consensus related to the above definition is growing in popularity among researchers, the use of isolated tests and cut-off scores has not been popular in school for identification purposes. Some agreement is present, however, that efforts should be made toward "tightening" the definition of LD so the classification process is not abused as much. Concerted efforts have been made in school identification procedures to isolate only the children who exhibit a significant discrepancy between general potential and performance in an isolated domain.

Thus, the salient and distinct feature of children classified with LD within the public school systems is a disability in an academic domain, such as reading, which is highly discrepant from their general intellectual competence and various educational or social opportunities. In terms of measuring this gap, the discrepancy between IQ and achievement on standardized tests has become the prototype for representing differences between potential and actual performance.

The current use of intelligence and achievement discrepancy scores for defining LD in school practice has a historical basis. The procedure can be traced to the use of expectancy formulas that emerged within the reading literature at the turn of the century. For example, Monroe (1932) suggested a reading index in which the reading age was divided by the average

mental age, chronological age, and arithmetic computation age. Several authors in the 1960s suggested additional measures, such as that listening comprehension be given equal weight to mental age, chronological age, and arithmetic computation age in measuring the gap between potential and achievement. Around the 1960s a clear bias existed in the classification of LD to calculate aptitude (IQ)-achievement test discrepancy calculations (Bateman, 1965). The implicit assumption for including discrepancy scores in the classification of LD was that children who have reading, writing, or math difficulties, not accompanied by a low IQ, are distinct in cognitive processing from the general "run of the mill" poor, garden variety, slow, or mildly retarded learners.

To date, discrepancy scores between standardized intelligence and achievement tests have become the *sine qua non* of defining children as having learning disabilities. Therefore, most states and school districts rely on discrepancy-based formulas that include IQ and achievement as the primary procedure to classify children as having learning disabilities.

Although the discrepancy model is one of the most important concepts underlying the field of LD in theory and practice, its construct validity may be questioned. For example, when compared to nondiscrepancy-defined poor achievers, groups defined by learning disability are more similar than different in processing difficulties (e.g., Stanovich & Siegel, 1994). I will briefly review issues on the validity of discrepancy-defined groups, which in turn will serve as basis for introducing dynamic assessment to the classification process. To understand the challenges of discrepancy-based classification, however, a brief review of the important conceptual issues will be helpful.

The Problems with Discrepancy Scores

The first major study on the distinctiveness of "discrepancy-defined children" emerged in the 1970s. A study by Rutter and Yulle (1975) focused on 9-year-olds from a sample of 2,300 children. Two groups of poor readers were defined from the sample. One group of "backward readers" had IQs around 80, and the other poor-reading group, defined as a "specific reading retardation group," had IQ scores around 102. The specific reading retardation group was less likely than the backward reader group to have organic brain damage or to display neurological abnormalities. The backward readers displayed a variety of motor abnormalities and left-right confusions. Both groups, however, had similar proportions of family members with histories of delays in language development. More important, Rutter and Yule found differential growth curves for the specific reading retardation group when compared to the general retardation (backward reader) group. The backward reader group displayed more

growth in reading and *less growth* in arithmetic than the specific reading retardation group.

To date, several studies have argued against the idea that generally backward and specific reading retardation groups are distinct. Although the outcomes of some of these studies differ (see reviews by Fletcher et al. [1994], Siegel [1992], Stanovich [1991], on the equivocal findings; some studies yield null results, and others yield small but statistically significant differences between IQ-based discrepancies and low-reading groups), they reflect variations in the type of discrepancy score (no consensus seems to be present on how discrepancies should be defined).

Recent studies correcting for the flaw in the earlier studies have converged on the evidence that children classified with LD, specifically in the area of reading, reflect a normal distribution of a multifactorial trait in reading characteristic of other poor readers (Fletcher et al., 1994; Stanovich & Siegel, 1994). *In short, discrepancy- and nondiscrepancy-defined groups of poor achievers are more alike than different.*

The implication of these findings is that one of the most important constructs is ostensibly "weakened." I say weakened because the impact and implications of these findings have yet to filter down to actual diagnostic practice in the schools. In contrast, much of the discussion in practitioner-related journals on the issue of classification focuses on the significant and reliable differences between objectively measured domains, such as reading, and general intelligence (e.g., Reynolds, 1984–1985; Shephard, 1980). The discussion at the practitioner level focuses on computation (e.g., establishing formulas that control for regression and overidentification), not necessarily issues of construct validity (meaningful classification systems). Therefore, the refinement of methodology in school identification of children with LD has operated independent of construct validity issues (similar to an argument that a test may be reliable but not necessarily valid).

Regardless, the most serious conceptual challenge to the field is that the cognitive profile of children with LD cannot always be discriminated reliably from generally low-achieving children when using traditional assessment measures (Fletcher et al., 1992, 1994; Siegel, 1992). Quite simply, several studies suggest that children with learning disabilities cannot be separated from slow learners (garden variety poor readers, Stanovich, 1986) on psychometric or cognitive measures (e.g., Fletcher et al., 1992; Siegel, 1992). No doubt, one can garner several criticisms for the line of research that continually supports the null hypothesis. These criticisms are not powerful because they indicate only problems in the depth of our knowledge, not necessarily serious conceptual flaws. I will list these criticisms briefly, however, then proceed to the more substantive issues.

First, there is a mismatch between conventional wisdom and research outcomes. Most lay people view LD as reflect-

ing some sort of difference between what the students really can do and what they actually do in certain situations (Swanson & Christie, 1994). Although models of LD do not have to match a layperson's understanding of LD, the notion of discrepancy is a phenomenon frequently reported by parents of LD children and individuals with LD (Gerber & Reiff, 1991). [The weakness of this argument, of course, is that people who are generally slow in all academic areas also may operate with the notion, "I have more potential than I'm showing."]

Second, little empirical evidence has been available on the validity of discrepancy scores, other than between intelligence and reading. A discrepancy in discriminating between groups may have validity in the areas of mathematics, problem solving, writing, and so on.

Third, we don't know why a discrepancy emerges between IQ and achievement (e.g., reading) in a "psychological" sense (in contrast to statistical artifacts). Little research has focused directly on understanding the determinants and consequences of a discrepancy (however, see Sovik, Frostad, & Lie, 1994).

Fourth, most research studies testing the discrepancy model are conceptualized or directed toward searching for the association of deficits. Support for the null hypothesis may provide an insecure basis for theorizing. That is, the findings that differences between the nondiscrepancy and discrepancy groups are statistically comparable or that their reading scores are merely variations of the same continuum is not theoretically compelling. More compelling is the possibly of dividing discrepancy groups into subgroups in terms of children who show a disassociation in processes and those who do not. In neuropsychology, for example, the tendency has been to deemphasize associations and to place greater reliance on *disassociation* as a source of theoretical insight (Shallice, 1988).

Finally, several important questions are not answered with this line of research. For example, why have we not consistently found cognitive processes that could account for a discrepancy? Is a difference score merely an artifact, or does something mediate this difference that has not been controlled? One could make the argument that the right questions haven't been studied. That is, the question, "Do reading difficulties of the dyslexic stem from the same problems as those characterized as poor readers without an IQ discrepancy?" is not the question to ask. Rather, the question is, "What processes independent of achievement mediate the emergence of discrepancies and nondiscrepancies in low achievers?"

Substantive Issues

Although the above arguments are serious, they are not compelling. Direct answers to these questions call into question the current classification procedures for identifying chil-

dren with potential LD. Attempts to validate the use of potential-achievement discrepancies in the identification of children with learning disabilities rests on some key assumptions. Unless these assumptions are met, classification studies will continue to reflect the artifacts of difference scores discussed in the literature of the 1950s and 1960s. To progress in classifying children accurately as having learning disabilities via a discrepancy model, some key assumptions must be met. These assumptions are presented below.

1. *Construct integrity.* The beginning step in capturing the notion of a discrepancy is to define the measures and match those measures to the construct definition. A test of the construct validity of discrepancy groups stands a greater chance of being assessed if the constructs included in the classification of groups are grounded firmly in theory. Most critically, "there is little reason to believe, and certainly a lot of empirical support to disbelieve, the contention that some arbitrarily weighted function of two variables will properly define a construct" (Cronbach & Furby, 1970, p. 79).

   Important criteria for establishing construct validity include the demonstration of convergent and discriminant validity of the measures. Although schools use the WISC-III (or WISC-R) and standardized achievement tests to determine discrepancy scores, this is not an argument for conceptual integrity. Neither a theoretical rationale nor empirical evidence is available to substantiate the claim that IQ tests (e.g., WISC-III) capture the construct of "potential." Quite simply, it is not the case that individuals with comparable IQ scores on the WISC-III have the same potential. For example, a difference between an intelligence score on the Wechsler test and a serious performance deficit on the Wide Range Achievement Test (or any other achievement test) in the area of reading is not a valid test of a discrepancy model. Neither test fits into a theoretical framework of intelligence or reading.

2. *Independence of measures.* Discrepancy scores (or discrepancy-defined groups) are correlated with their component parts, and therefore the discrepancy measure will relate significantly to other variables correlated with the component parts (Cronbach & Furby, 1970). When difference scores are correlated with their component parts, there is a greater than chance tendency for them to be correlated with other variables that are associated with those component parts.

   An example of the above rule is as follows. We know that reading recognition (word skill accuracy) is highly correlated with measures of phonological awareness. Therefore, when (a) reading recognition is part of the discrepancy score, and (b) low reading ability groups are comparable on reading recognition performance, then performance is comparable between discrepancy and nondiscrepancy groups on processes (phonological awareness) related to reading. Thus, the discrepancy group is little more than a surrogate of the poor reading group.

   This problem (circularity in findings) has been recognized in the literature for some time (Bereiter, 1963; Cronbach & Gleser, 1953; Lord, 1956; Wanous & Lawler, 1972). In fact, it can be demonstrated that systematic relationships between component or correlate scores and difference scores exist even when the difference scores are generated randomly (Wall & Payne, 1973). For example, Wall and Payne demonstrated that when the components related to job satisfaction including the use of how the job should be versus the way it is, the discrepancy scores were correlated with the component measures, even when component scores were assigned randomly to subjects.

   In sum, it is reasonable to assume that a spurious relationship exists when the observed relationship between two variables (for example, intelligence and reading) is attributable largely to the relationship between both variables and a third variable. Thus, when discrepancy and nondiscrepancy groups are compared, the mediating variables shared between the two groups must be partialed out so a clear test of differences can occur.

3. *Direction of outcomes.* The direction of the discrepancy must be consequential in performance outcomes. A critical assumption in testing the discrepancy model is that differences in the direction of the profiles are important. The fact that *Child A* has a high reading score but a low intelligence score should reflect a different "set of" or "level of" processes when compared to *Child B* with a high IQ score but low reading score. A major assumption in variable selection for classification purposes that has not tested adequately is that the direction of the discrepancy is theoretically consequential (Cronbach & Furby, 1970; Johns, 1981). For example, even though phonological processing is comparable between average IQ and poor readers and low IQ and poor readers, it is not necessarily the case that low IQ and high readers have better phonological processing skills (Sparks, 1995).

   The field of LD, of course, assumes that the direction in which the discrepancy occurs is important. I have no data to argue that positive and negative discrepancies between intelligence and achievement result in the same performance on various processes between diagnostic groups. In most contexts, I suspect that children with an average IQ but very high reading ability or mathematics ability might be more likely to excel in particular processes relative to another group with average IQ that has a low reading or

math score. My point is that no study equating different uses of discrepancy scores has been done to determine if direction is consequential on cognitive performance.

It would appear to me that the information that goes into a discrepancy must have face validity in terms of the direction. If the direction is unimportant, those measures used to determine a discrepancy should be removed from the discrepancy formula.

4. *The comparable performance between groups on various measures is not the same as the relationship among variables within groups.* Few studies have been undertaken to examine the interrelationship among cognitive processing variables in predicting the academic performance of children with learning disabilities when compared to nondiscrepancy groups. The validity of discrepancy definitions seems to be enhanced if the interrelationship among processes that contribute to achievement difficulties can be shown to differ between discrepancy and nondiscrepancy groups. For example, phonological coding may contribute tremendously to word recognition in the garden variety poor readers, whereas phonological coding, working memory, and vocabulary may contribute unique variance to the reading process in discrepancy-defined groups. In addition, some cognitive deficits may co-occur in various reading groups, but that does not mean they all share the same causal link to the reading problems.

5. *Measures related to the discrepancy scores are valid only if assessed on something above and beyond their components and correlates.* Most researchers recognize the reliability problems with discrepancy scores (e.g., Zimmerman & Williams, 1982), but few recognize that the use of discrepancy scores implies that it accomplishes something beyond their component parts. As stated previously, we know that different scores are related systematically to their components, and thus one cannot test the construct validity of the discrepancy model with such scores. What has to be tested, however, is whether the discrepancy scores accomplish something beyond their components.

One obvious test is whether children defined by discrepancy scores are more likely to respond favorably to one treatment when compared to poor-achieving children without discrepancies. Responsiveness to instruction seems to be a missing test of the majority studies comparing discrepancy and nondiscrepancy groups. To date, no systematic analyses support the notion that the discrepancy model is a usable construct when it comes to intervention and prognosis to intervention.

Summary of Classification Issues

My major premise is that recent LD classification research has confirmed what has been known about the limitations of

difference scores for some time (e.g., Cronbach & Furby, 1970); and that an adequate testing of the construct has yet to occur. Research would be better served by:

- Meeting the aforementioned assumptions for testing the validity of discrepancy scores, and/or
- Choosing an alternative means of assessing discrepancies (e.g., concurrent validity on multiple discrepancy measures; profile analysis; multivariate linear modeling-analogous to growth curve analysis [Bryk & Raudenbush, 1987; Francis et al., 1991; Rogosa, 1988]) on those measures closely aligned with theoretical models of measuring potential and actual behavior.

ALTERNATIVES TO CLASSIFICATION PROCEDURES: DYNAMIC ASSESSMENT

Given the current crisis related to classifying children suspected of LD based on discrepancy scores, what are the alternatives to current practice? To identify such procedures, we must narrow down our concerns to answering two major questions.

First, how should "potential" be measured? As stated previously, the notion of potential and actual performance has played a critical role in defining LD (see Bateman, 1992, for a review). As discussed, however, a conceptual problem exists as to whether "potential" is captured adequately on traditional IQ measures (Naglieri & Das, 1988; Swanson, 1993a). The relevance of standardized intelligence measures (e.g., WISC-III, WISC-R, Stanford-Binet) in the diagnostic classification of LD has been criticized because achievement (e.g., reading) within learning disability samples is not predicted by variations (high versus low) in IQ (e.g., Siegel, 1989; Stanovich & Siegel, 1994).

Further, several authors (e.g., Campione & Brown, 1987; Feuerstein, 1980; Haywood, Brown, & Wingenfeld, 1990) have suggested that traditional intelligence tests (tests that measure unassisted performance on global measures of academic aptitude) sometimes underestimate general ability. Several authors also argue that because static or traditional approaches to assessment typically provide little feedback or practice prior to testing, failure often reflects children's misunderstanding of instructions more than their ability to perform the task (Haywood et al., 1990).

Second, can the processing potential of children with LD be separated from poor achievers whose difficulties are related to inadequate instruction or teaching? This issue is important because assessment practices that rely heavily on psychometric tests for classification of children with LD have not provided systematic procedures to separate children who have learning problems related primarily to inadequate or weak instructional support from children who have informa-

tion-processing deficits (Fletcher et al., 1994; Siegel, 1992; Stanovich & Siegel, 1994).

No systematic attempt has been made to determine whether the purported processing difficulties in children suspected of LD are stable or are amenable to instruction. Bateman (1992) stated, in relation to LD, that "in many instances assessment consists of a WISC-R, a Woodcock-Johnson, or PIAT. There may be little real diagnosis, no exploration of response to instruction, of error patterns or correlated disabilities" (p. 32). No doubt, these practices contribute to the overidentification of children as having LD, as well as inappropriate placement into special programs (Shepherd, Smith, & Vojir, 1983).

WHAT IS DYNAMIC ASSESSMENT?

One possible alternative or supplement to traditional assessment—and a procedure to partially answer the two aforementioned questions—is to measure a child's performance when given examiner assistance. Procedures that attempt to modify performance via examiner assistance in an effort to understand learning potential are called dynamic assessment (DA). Several authors consider the first area of focus in assessment to be one of improving the processing of information. For example, utilizing Vygotsky's (1978) zone of proximal development, Brown and French (1979) made a distinction between an individual's proximal potential and actual level of performance. In the area of child development, for example, they stated:

> A distinction is made between a child's actual development; i.e., his/her completed development as might be measured on a standardized test, and his/her level of potential development, the degree of competence he/she can achieve with aid. Both measures are seen as essential to diagnosis of learning abilities and the concomitant design of remedial program. (p. 255)

Traditional approaches of DA include assessment of learning potential (e.g., Budoff, 1987), testing-the-limits (Carlson & Wiedl, 1979), mediated assessment (e.g., Feuerstein, 1980), and assisted learning and transfer (e.g., Bransford, Delclos, Vye, Burns, & Hasselbring, 1987; Campione, Brown, Ferrara, Jones, & Steinberg, 1985). Although DA is a term used to characterize a number of distinct approaches (see Brown, Campione, Webber, & McGilly, 1992; Carlson & Wiedl, 1979; Haywood et al., 1990; Palincsar, Brown, & Campione, 1991; Speece, Cooper, & Kigler, 1990; for a review), two critical features are to determine the learner's potential for change when given assistance (e.g., see Campione & Brown, 1987) and to provide a prospective measure of performance change independent of assistance (Embretson, 1987).

Unlike traditional testing procedures, score changes resulting from examiner intervention are not viewed as threatening

task validity. In fact, some authors argue that construct validity increases (e.g., Carlson & Wiedl, 1979; Swanson, 1992; also see Embretson, 1987, for a review).

To obtain information about a child's responsiveness to hints or probes, dynamic approaches require the interaction of an examiner and the examinee. When a student is having difficulty, the examiner attempts to move the student from failure to success by modifying the format, providing more trials, providing information on successful strategies, or offering increasingly more direct cues, hints, or prompts. The intensity of the intervention ranges from several sessions to brief intensive prompts in one session. Thus, "potential" for learning new information (or accessing previously presented information) is measured in terms of the distance, difference between, and/or change from unassisted performance to a performance level with assistance.

Utilizing these procedures to assess potential, a number of questions can be addressed to determine the validity of the classification of some children labeled as having learning disabilities. For example, do children with LD, when given instructional support on processing tasks, reach the same level of performance as normally achieving children? An answer to this question is important because the processing difficulties of children with LD are assumed to be stable when compared to other processing abilities (see Swanson, 1988, for a review). Thus, if processing performance of children with LD can be modified and their performance is comparable statistically to normally achieving children, the "intrinsic nature" of LD has to be reexamined.

Problems with Current Dynamic Testing Models

Although DA has been suggested as an alternative or supplement to traditional assessment (see Palincsar, Brown, & Campione, 1991 for a review), no published standardized data are available on such procedures. In fact, DA has been criticized because of its highly clinical nature and poor reliability (e.g., Palincsar et al., 1991). Further, interactive testing procedures (i.e., DA) have not been operationalized in terms of test-related interventions characteristic of a school testing situation. That is, a school psychologist or diagnostician usually samples a student's processing ability during short testing sessions, not over an extended period of days or sessions.

Other difficulties arise when applying current dynamic assessment (DA) models to the classification of children with possible LD. Because of the diversity of DA approaches, I will outline only general problems. (The reader is referred to Brown et al., 1992, for a description of various DA approaches.) The limitations are as follows:

1. Several DA approaches (e.g., testing the limits) rely extensively on verbal cues or verbal comprehension, or

both. The poor linguistic processing of children with learning disabilities is well established in the literature (e.g., Shankweiler & Crain, 1986; Siegel & Ryan, 1989). Thus, children suspected of having LD would be at a clear disadvantage if the interaction (via, probes, cues, etc.) between the examiner and examinee were primarily verbal in nature. Although intervention must include verbal information (via prompts, hints, explanations, etc.), such procedures would have to allow for some compensatory processing (e.g., visual/spatial) or alternative processing, or both, to meet task demands.

2. Most DA procedures have been validated primarily on tasks that have weak correlations with academic products. Further, some tasks in which change has been shown (e.g., Raven Progressive Matrices, Block Designs) have little relevance to areas in which children with LD experience difficulties (reading and writing).

3. Some modified testing situations in DA reflect open-ended flexibility, but not necessarily consistent application (i.e., replication depends on the examiner's clinical skills). Some DA approaches fail to use a standardized approach but, rather, develop in interaction with the examinee. Such a clinical approach may hamper the development of consistent rules, principles, and procedures to assess performance accurately.

4. Most DA procedures are completely inappropriate for a school psychologist's or diagnostician's context as currently defined. Some DA procedures are not usable because they reflect interventions across several sessions rather than provide a picture of "responsiveness" in one or two testing sessions. Although this criticism is more practical than conceptual, most testing situations, at least for the school psychologist, involve classifying of children for instructional purposes in less than three testing sessions. Granted that assessment and intervention are closely aligned for classification purposes; however, one is interested in how flexible children's processing ability is during brief observation periods.

5. The "level" or "meaning of" the dependent measures is in question. Some DA approaches focus on change scores (products) rather than the processes that are changing. Although the assessment of LD certainly can be directed at how much change can occur in a score, of more interest are the cognitive processes or strategies that may influence the changes.

6. All DA approaches I have reviewed suffer from psychometric adequacy. The majority of DA approaches lack information on the validity or reliability of scores. Although changes can occur in scores during DA, we are uncertain if the change scores are better predictors of future performance in academic domains such as reading when compared to initial (static) testing conditions.

7. In some DA approaches it is uncertain whether changes in performance reflect information that is available or accessible in the "mind" of the child tested or merely reflects changes in testing format. For example, in some DA procedures, changes in performance may be related to increasing task familiarity, instructions, expectations, or individual attention. Thus, changes are not necessarily influencing any deep cognitive structures as much as they are directing the child's attention to important task parameters.

Further, it is uncertain whether examiner feedback to the child suspected of having LD activates the information that is already in the child's mind or reintroduces the concept again to a poorly stored original memory trace. For classification purposes, the differences between learners on these latter issues are important if gains in scores are to be attributed to changes in environment or changes in cognition, or both.

Recent Work Applying DA to the Classification of Learning Disabilities

To partly address some of the above issues, I have invested considerable time in developing a standardized test that includes components of DA (systematic probing, assessing the maintenance of performance after probing), and information processing (the source of problems in children with learning disabilities). Before talking about the method of DA for this measure, however, I need to discuss the *substance* of assessment.

I assume that a major process underlying all problem-solving activities (i.e., intelligent behavior) can be accounted for by individual differences in working memory. All major information-processing models on skill acquisition and learning include the component of working memory (e.g., see Baddeley, 1986; Cantor & Engle, 1993; Just & Carpenter, 1992) because it is highly correlated with performance on several academic and language-related tasks such as vocabulary (e.g., Daneman & Green, 1986), reading comprehension (e.g., Engle, Cantor, & Carullo, 1992), language acquisition (e.g., Gathercole & Baddeley, 1989), problem solving (e.g., Kyllonen & Christal, 1990), and mathematics (e.g., Swanson, Cooney, & Brock, 1993), to name a few. In fact, correlations between working memory measures and reading or intelligence measures with adult samples are in the range of .55 to .92 (e.g., see Carpenter, Just, & Shell, 1990; Daneman & Carpenter, 1980; Kyllonen & Christal, 1990). Coupled with these findings, deficits in working memory have been attributed to children experiencing learning difficulties (e.g., Siegel & Ryan, 1989; Swanson 1993b, 1993c).

Traditionally, working memory is defined as a system of limited capacity for the temporary maintenance and manipulation of information (e.g., Baddeley, 1986; Just & Carpenter,

1992). Tasks that measure working memory are those in which a person must hold a small amount of material in mind for a short time while carrying out further operations simultaneously. Everyday examples of working memory tasks include holding a person's address in mind while listening to instructions about how to get there, or perhaps listening to the sequence of events in a story while trying to understand what the story means.

Described in this way, working memory differs from a related concept of short-term memory that is typically used to describe situations in which small amounts of material are held passively (e.g., digit or word-span tasks) and then reproduced in an untransformed fashion (Cantor, Engle, & Hamilton, 1991; Just & Carpenter, 1992). Several studies in the experimental literature indicate that working memory and short-term memory measures are distinct (e.g., Cantor et al., 1991; Daneman & Carpenter, 1980) and reflect independent variance in their contributions to achievement (e.g., Cantor et al., 1991; Kyllonen & Christal, 1990; Swanson, 1992).

COGNITIVE PROCESSING TEST (S-CPT)

To date, no standardized instruments have been developed specifically for use in school settings to measure information-processing ability under static and dynamic testing conditions. Thus, development of the Cognitive Processing Test (S-CPT; Swanson, 1995) was stimulated by the need for a comprehensive instrument to assess important components of processing ability under standardized dynamic testing conditions. The three stated purposes of the test are

1. To provide an approximate index of *processing potential*. Processing potential is defined as the difference between an individual's actual performance level (his/her performance as typically measured on standardized tests) and the degree of processing competence the individual can achieve with aid. Thus, processing potential is determined by assessing the examinee's ability to incorporate hints or probes to maximize test performance. The administration of probes or hints provides diagnostic information about whether the examinee's ability to access information can be facilitated and higher levels of performance can be achieved when instructional support is provided. In addition, an assessment is made of the examinee's ability to maintain a high level of performance when instructional supports are removed.
2. The test measures the examinee's declarative knowledge of the best "plan" to remember information. For certain subtests, the examinees chose one of four strategies they believe will facilitate future recall. It is assumed that some examinees who selected certain strategies do better on informa-

tion-processing tasks than those who select other strategies. The strategy choice provides diagnostic information about how the examinee may be mediating information.
3. The test serves as a diagnostic tool to help identify strengths and weaknesses in processing ability. The test is designed to identify individuals who have difficulty learning information, as well as students who excel in processing new information. Thus, performance on the S-CPT is expected to correlate with various achievement and aptitude measures.

Relevant Research

In response to some of these assumptions, I would like to briefly review some recent work I have done in standardizing a DA measure and how it facilitates the classification of children with LD (see Swanson, in press b for a description of the complete study). Two studies are reviewed, which had the following purposes.

First, a determination was made as to whether processing "potential," via DA, is related to academic achievement. Processing potential was defined as the score obtained with examiner assistance (i.e., gain score). Study 1 serves as a validation study as to whether processing potential, via DA, contributes unique variance to achievement beyond what already is contributed by a traditional intelligence measure. Linking DA with achievement, as well as determining if "potential" as measured on a commonly used IQ test differs from potential as measured under dynamic testing conditions in the prediction of achievement, is an important issue if DA is to be taken seriously as a valid assessment procedure in the diagnosis of LD.

The tasks used in Study 1 for assessing information-processing potential are related to working memory. The standardized test (N = 1610) used to measure working memory is the Swanson-Cognitive Processing Test (S-CPT, Swanson, 1995). This is an individually administered battery that is assumed to measure different aspects of working memory ability and processing potential. Working memory is defined in this test as concurrent processing and storage activities, whereas potential, via DA, is defined as (a) the learner's performance change relative to initial performance when given assistance (gain) and (b) performance change independent of assistance (maintenance).

Second, it was of interest to determine if children with LD can be discriminated, via DA, from children whose processing difficulties are highly sensitive to examiner assistance (referred to as instructionally deficient children) and children (referred to as slow learners) whose processing difficulties are poor but (a) stable (minimal change between initial and gain score) and (b) commensurate with their low achievement (i.e., minimal discrepancy in standard scores exists between processing and achievement).

As stated earlier, this is an important contrast, as several studies (e.g., Fletcher et al., 1992; Stanovich & Siegel, 1994) indicate that no clear processing distinctions exist between slow learners and children with LD. Caution, however, should prevent us from concluding that children with LD cannot be distinguished diagnostically from other learning ability groups. The fact that current practices using static measures do not distinguish children with LD from children who are low achievers does not mean it cannot be done. Thus, in Study 2, we examine whether a child's response to assisted performance provides a frame of reference for separating children who are teaching disabled (i.e., children whose processing performance is easily modified with assistance) and slow learners (children whose assisted and unassisted processing performance matches their achievement scores) from those who have learning disabilities (i.e., children who maintain a discrepancy between general processing performance and achievement under assisted and unassisted testing conditions). We also hope to provide partial answers to the five conceptual issues raised earlier related to testing the validity of discrepancy scores.

Study 1
Question: Does dynamic assessment better predict achievement than traditional IQ measures?

The question of interest is whether performance on information processing tasks, as reflected on the S-CPT, provides any unique prediction to achievement beyond what already is contributed by IQ. If achievement—in this case, reading and mathematics—is influenced by measures of information-processing potential, one would expect the correlation between process measures and achievement to be significant. Further, one would expect measures of information-processing potential, via DA, to contribute significant variance to achievement beyond what already is contributed by a traditional IQ measure.

It also was of interest to determine if information-processing measures that reflect assisted processing performance are better predictors of achievement than initial or unassisted performance on the same processing measures. It was assumed that assisted performance better reflects processing potential than unassisted performance does and therefore would contribute significant variance to academic achievement. If this is the case, measures of performance gain would provide better estimates of academic performance than initial performance measures.

To this end, the present study compared the criterion-related validity of the initial and DA scores from the S-CPT with that of the Wechsler Intelligence Test (WISC-R). Children of varying reading ability participated in Study 1. The sample of 24 females and 37 males was selected from the standardization study of the S-CPT. To ensure a wide range of achievement performance, the present sample included 10 children with poor reading skills, 40 skilled readers, and 11 highly skilled readers, based on the reading subtest of the Wide Range Achievement Test-R (Jastak & Wilkinson, 1984). The only other stipulation in the sample selection was that the children show a higher standard score for the gain (dynamic) than initial testing condition. It was assumed that if a valid test is to be made of the assumption that gain performance contributes unique variance to academic achievement, then the sample should reflect improved standard scores for the gain when compared to the initial testing condition.

Children were tested in school districts surrounding a large urban area in the Northwest. All participants were selected from middle-class to upper-middle-class schools. Ethnicity for the sample was Anglo-European. The mean chronological age for the sample was 10.63. All children were individually administered the WISC-R, reading and mathematics subtests from the WRAT-R, and the S-CPT.

S-CPT Battery. An important feature of *all* subtests in this test battery is that they require the maintenance of some information during the processing of other information. The processing of information is assessed by asking examinees a comprehension question about to-be-remembered material, whereas storage is assessed by accuracy of item retrieval. For example, on the Rhyming subtest discussed below, examinees are presented the words "lip-slip-clip" and asked if "ship or lip" was presented in the word set. They then are asked to recall the previously presented words (lip-slip-clip) in order. On the visual-spatial subtests the process question focuses on the spatial location of items. For example, on the Visual Matrix subtest, the examinee is presented a series of dots in a matrix and is allowed 5 seconds to study the matrix. After the target matrix is removed and the examinee is presented a blank matrix (a grid with no dots), the examinee is asked a process question: "Were there any dots in the first column I showed you earlier?" After answering the process question, the examinee is asked to draw dots in the correct boxes.

A previous study (Swanson, 1992), and the research reported in the testing manual (Swanson, 1995) established the psychometric characteristics (construct and criterion-related validity, reliability) for the measures (also see Swanson, in press a and b). A brief description of the various subtests administered under static and dynamic testing conditions is presented below.

Description of Subtests. The 11 subtests are described in detail elsewhere (Swanson, 1992; 1993b) and therefore are described only briefly here.

Subtest 1—Rhyming. This subtest assesses the examinee's recall of sets of words of increasing complexity that rhyme.

Subtest 2—Visual Matrix. This subtest assesses the examinee's ability to remember visual sequences within a matrix.

Subtest 3—Auditory Digit Sequence. This subtest assesses

examinee's ability to remember numerical information embedded in a short sentence. A sample sentence (item 3) is: "Now suppose somebody wanted to have you take them to the supermarket at 8651 Elm Street...." They then are presented a process question. The process question is: "What is the name of the street?"

Subtest 4—Mapping and Directions. This subtest assesses whether the examinee can remember a sequence of directions on a map that does not have any labels. The examiner presents the examinee with a street map with lines connected to a number of dots that illustrate the direction a bike would go to get out of the city.

Subtest 5—Story Retelling. This subtest assesses the examinee's ability to remember a series of episodes presented in a paragraph. The examiner reads a paragraph, asks a process question, and then asks the examinee to recall all the events that have occurred.

Subtest 6—Picture Sequence. This subtest assesses the examinee's ability to remember a sequence of shapes of increasing spatial complexity. Pictures of shapes are presented on a series of cards, a process question is presented, cards are gathered and shuffled, and then the examinee is instructed to arrange those cards in the correct sequences.

Subtest 7—Phrase Sequence. This subtest assesses the examinee's ability to remember isolated phrases. A sample sequence of phrases is: a flowing river, a slow bear, a gripping tire. A sample process question is: "Are the words about a bear or a boat?"

Subtest 8—Spatial Organization. This subtest assesses the examinee's ability to remember the spatial organization of cards that have pictures of various shapes. These cards are ordered in a top-down fashion.

Subtest 9—Semantic Association. This subtest assesses the examinee's ability to organize words into abstract categories. For example, the examinee is directed to retrieve the words that go together (e.g., shirt, pants, and shoes; apple, orange, and grapes).

Subtest 10—Semantic Categorization. This subtest assesses the examinee's ability to remember word associations within categories. A sample item is: *job*, teacher, firefighter, police officer; *season*, summer, winter, fall. A sample process question is: "Which word, 'soldier' or 'summer,' was presented?"

Subtest 11—Nonverbal Sequencing. This subtest assesses the examinee's ability to sequence a series of cards with pictures of nonsense shapes in which the examiner does not provide the organization. The examinee is asked to reproduce the cards by each row based upon their original organization.

A complete description and example of probing (DA) are provided in Swanson (1992). Hints (probes) are presented to the examinee after he or she fails an item in each subtest. The hints are ordered in terms of explicitness; the general hints are given first, and more explicit hints later. Hints are administered based on the type of error made (i.e., whether the error is related to recency, primacy, or middle item), and prompting procedures continue until all targeted items cannot be recalled. The "bow-shaped curve," commonly found in episodic memory studies, provides the basis for ordering a series of hints from implicit to explicit information.

For example, consider the Rhyming Task (Subtest 1), which includes the sample item "car-star-bar-far" and the process question, "Which word did I say—jar or star?" Based on where the child omits, inserts, or orders the words incorrectly, a series of probes might include the sequence: (1) The last word in the sequence was "far"; now can you tell me all the words in order? (2) The first word in the sequence was "car"; now can you tell me all the words in order? (3) The middle words in the sequence are "star" and "bar"; now can you tell me all the words in order? and (4) All the words in order are "car-star-bar-far"; now can you tell me all the words?

The examiner provides a series of hints based on the error that is closest to Probe 1. For example, suppose the child for the example responds *car-bar-far*. The child obviously left out a word in the middle, so the experimenter would provide a hint related to the middle words (Probe 3 in this case). If Probe 3 does not provide the correct response, the experimenter then would move to Probe 4. In contrast, if a child responds initially by saying only *car*, the sequence would begin with Probe 1 and proceed through all probes until the correct response is given. If a correct response does not occur after the probing, the task is discontinued and the next task is administered. If the examinee responds correctly, the next set of items of increased difficulty is presented.

Maintenance performance also was established by readministering the highest item on the gain condition but without prompts. Maintenance performance was tested after the first six subtests had been administered (session 1). Maintenance performance for the next five subtests (subtests 7 through 11) was determined after administration of subtest 11 (session 2). Approximate time to assess maintenance was 15 minutes in each session.[1]

---

[1] The reader may question whether systematic cuing represented in the current procedures reflects dynamic assessment; that is, the current procedures emanate from information-processing research, whereas some readers associate dynamic assessment with mediated learning (e.g., Feuerstein, 1980) or as a derivative of Vygotsky's (1978) assumptions on proximal development (e.g., Ferrara, Brown, & Campione, 1986). The present approach attempts to *test the limits* of children's performance by providing brief standardized prompts. It is assumed that such testing conditions on the S-CPT reduce the number of inefficient strategies for accessing previously presented information, thereby directing children to use comparable strategies. The systematic cuing procedures used in the present study improve access to previously stored information by emphasizing sequential processing strategies.

What should we measure? An unresolved issue in DA concerns the nature of scores necessary to measure cognitive modifiability or processing potential (see Embretson, 1987; Palincsar et al., 1991, for a review). For example, Campione and Brown (1987) measure modifiability as the number of hints required to solve a problem that has been failed. The fewer the hints, the more modifiable is the examinee's processing.

Embretson (1987) has suggested, however, that this score "merely provides a better estimate of initial processing ability." She suggested that a more appropriate measure is to bring scores to an asymptotic level (under the probing conditions) and then obtain a second measure on the task performance after the probes have been removed. The basic rationale for this measure is to eliminate performance differences that result from different processing strategies or unfamiliarity with the laboratory procedures.

Another measure of cognitive modifiability is a simple difference score (gain score minus initial score). Such a score, however, is subject to regression toward the mean. In short, no agreed-upon measure of cognitive modifiability is available as yet.

To address these issues, as well as to capture the breadth and depth of processing ability, the S-CPT has six composite scores, in addition to scores during the *initial* (static) testing condition. The first measure, the *gain* score or asymptotic level, is defined as the highest score obtainable under probing conditions. A second measure, *probe* score (referred to in the test as instructional efficiency), is determined as the number of prompts necessary to achieve the asymptotic level. A third measure, *maintenance* score, is the stability of the asymptotic level without the support of probes or hints. The maintenance measure is scored dichotomously so that if the gain score is not maintained beyond the initial score, the initial score is assigned to the examinee. Thus, modifiability for maintenance conditions is measured in an absolute sense.

A fourth measure captures the difference between the gain and initial scores. This *processing difference* score measures the difference between the actual performance level and the level of potential performance as determined under guided assistance. Vygotsky (1935/1978) considered this in-between

It is assumed that by directing all children to use these strategies, the number of competing strategies employed is reduced.

It also is important for the reader to note that no current dynamic testing procedures to date have produced evidence related to the psychometric qualities of the measures. Further, interactive testing procedures have not been operationalized in terms of a test-related intervention characteristic of a school psychologist's or a teacher's testing situation. Most studies of dynamic assessment have involved intervention over several sessions, which prohibits their use by school psychologists in normal testing situations. Thus, the purpose of this study was to measure the sensitivity of a test-related intervention that was sufficiently brief to administer within a normal testing period.

state "the zone of proximal development." A fifth score captures *processing stability*. This score is determined by subtracting the initial score from the maintenance score. A high stability score suggests that some internalization of previous guided instruction has occurred.

A final score, *strategy efficiency*, relates to selecting a strategy prior to recall. As previously stated, six of the 11 subtests required subjects to select a picture of someone using a strategy that best reflects how they plan to remember. The strategy efficiency score reflects the examinee's procedural knowledge prior to DA. Picture choices are ranked in terms of efficiency, based on the standardized sample. It is assumed that the degree of modifiability in performance may be related to individual differences in procedural knowledge.

Results. All raw scores from the S-CPT were transformed to standard scores, based on the standardization sample. These scores were standardized by age.

The results clearly indicated that Gain scores are better predictors of achievement than WISC-R scores in the current sample. Further, Gain scores are the only variable that significantly predicted mathematics performance.

What does this study suggest about classification of children suspected of LD? *First*, when contrasted with the WISC-R, the results suggest that DA, via the S-CPT, contributes significant variance in predicting reading and math achievement. Although the usefulness of intelligence scores has been questioned in the past in terms of their value in predicting reading (Siegel, 1989), the present study does suggest that DA measures contribute important variance. *Second*, the results suggest that reading and math performance can be better predicted from scores on the S-CPT than the WISC-R. Finally, scores that reflect processing potential (gain) are better predictors of achievement than initial and WISC-R scores, suggesting that criterion-related validity is enhanced under dynamic testing conditions. In general, the results support the validity of using DA measures to predict achievement.

Study 2
Question: Does the classification of learning disabilities change as a function of dynamic assessment?

Study 2 gets to the heart of the classification issue. This study addressed the issue of whether children with LD can be separated from other children who are seriously deficient in reading, via DA procedures. A focus on reading is important, as 80 percent of the children served in special education settings have reading disabilities.

What does a child with learning disabilities look like from a dynamic testing perspective? Assuming that *general* working memory performance (based on Study 1) is a better "stand-in" for potential than a Full-Scale WISC III score (or any other traditionally administered intelligence test), an LD child's *general*

processing potential is in the average range. This general processing score (gain score) is discrepant from the achievement score in at least one academic domain (e.g., reading).

Further, under dynamic testing conditions, LD children's *general* overall processing score is stable when compared to nondynamic testing conditions. That is, although some children suspected as having LD will experience increases in processing performance on some isolated domains of working memory (e.g., visual-spatial versus verbal), their general scores will remain the same. This is because their deficit area (e.g., verbal working memory) loses ground relative to the standardization group, whereas their strong processes (e.g., visual-spatial processing) relative to their deficit processes maintain their rank relative to the standardization group. Thus, general scores (similar in computation to a Full Scale score) capture both the losses and gains in processing performance under dynamic testing conditions. In short, the general processing score of children suspected of having LD remains fairly stable from static (i.e., initial scores) to dynamic testing conditions (gain and maintenance scores).

Given that children with LD exhibit general information-processing performance that is (a) stable across initial and dynamic testing conditions and (b) their general processing potential score is discrepant from their academic achievement, these children might perform differently on working memory measures than children whose performance is more susceptible to change, via examiner assistance.

In Study 2, two diagnostic groups are of interest in determining discriminant validity of our classification scheme. One group, instructionally deficient, or teaching-deficient, children, was defined as poor readers who demonstrate substantial improvement in information-processing performance during dynamic testing conditions relative to their initial test performance. For the present study, instructionally deficient children were defined operationally as children who yield high Probe and Gain scores when compared to other subgroups within a poor reading sample.

The second group, slow learners, was defined operationally as having below average performance across an array of cognitive measures under static and dynamic testing conditions when compared to the other groups. It also was assumed that their potential score, as measured by Gain performance, is not discrepant from their achievement scores (e.g., see Fletcher et al., 1992; Siegel, 1992; for a related discussion on IQ and achievement scores). This group is analogous to current classification procedures that define nondiscrepancy groups as having low Full Scale and reading scores.

To this end, school-identified LD and skilled readers were compared on an array of subtests from the S-CPT administered under nonassisted (static) and assisted (dynamic) testing conditions. LD readers' and skilled readers' perform-

ances then were cluster-analyzed on Gain and Probe scores (the number of probes necessary to establish gain performance). Although a cluster analysis was expected to identify distinct subgroups, it was uncertain whether these empirically derived subgroups match diagnostic classifications reflective of slow readers and instructionally deficient children.

It also was unclear whether the empirically derived subgroups present unique profiles that differ from one another on variables *independent* of the classification variables. Thus, we tested the assumption that, although groups of children with reading deficiencies can score similarly on initial processing tests, they may perform significantly different from one another when provided examiner assistance. Further, these performance differences may prove useful in discriminating operationally defined LD children from other children who are poor readers.

All 155 participants were selected from public and private schools surrounding large urban areas. Consistent with the recent literature, children with LD met a minimum level or basal level on standardized intelligence tests IQ and performed below a standard score of 85 on a word-recognition measure. For LD readers, ethnicity was 57% Anglo-European, 8% Black, 31% Hispanic, and 4% other. For skilled readers, ethnicity was 71% Anglo-European, 5% Black, 21% Hispanic, and 3% other. Chronological age was 10.55 for poor readers and 10.66 for skilled readers.

No significant differences were found between ability groups in terms of chronological age or gender, but clear differences did emerge between groups in terms of ethnic representation. All ethnic groups spoke English as their primary language. The high-ethnic representation in the poor reading (LD) sample reflects the geographic region (Southwest) of the sample selection.

The sample of children reflected the diversity of children classified with learning disabilities found in the majority of schools (i.e., IQ scores in the average to low-average range, but achievement scores significantly below average). Further classification of children with LD followed state and provincial guidelines that closely matched the 1977 *Federal Register* definition. The definition reflects: (a) a learning problem that is specific, generally confined to one or two academic areas; (b) the child's poor achievement is not commensurate with his/her ability as in other academic areas, which are average or above based on the child's chronological age; and (c) the learning difficulty is not primarily the result of retardation, poor teaching, or cultural deprivation.

All children with LD were administered the reading and math subtests from the Wide Range Achievement Test (Jastak & Wilkinson, 1984). Intelligence scores were measured on the K-ABC, the Slosson Intelligence Test-R (SIT-R), the WISC-R, or the Peabody Picture Vocabulary-R Test.

Results. All S-CPT scores were significantly higher for skilled than LD readers. To determine which measure best predicted reading ability group classification, the six composite scores (initial, gain, maintenance, probe, processing difference, and stability score) from the S-CPT were entered into a forward stepwise discriminant analysis. The analysis indicated that Gain scores best predicted ability group classification.

A closer look at the school-based LD classification. Although the above analysis indicates that Gain scores best predict ability group classification, the discriminant validity was weak. The weak discriminant validity is attributable to the heterogeneity within each reading group. To address this problem, subgroups were derived empirically, via a cluster analysis. A nearest centroid sorting procedure created four subgroups within each sample. Variables used to cluster the groups were arithmetic, gain, and responsiveness to instruction (probe scores) performance. Arithmetic was used as a clustering variable, as it has served as a major variable in the subtype classification (Siegel & Ryan, 1989; also see Rutter & Yulle, 1975). Gain scores were used because these scores best predicted reading classification, and Probe scores because they reflect responsiveness to hints.[2]

To classify the subgroups, it was necessary to analyze discrepancy scores between processing potential (i.e., initial, gain, maintenance) and achievement. Such an analysis provides some external validity for interpreting the above clusters, because two of the processing scores (initial and maintenance scores) used to compute a discrepancy were *not* used in the cluster analysis. The standard score differences between the three main S-CPT scores (initial, gain, maintenance) and achievement scores (reading and arithmetic scores) are shown in Figure 1.

Because all scores are standardized, it is a rather straightforward procedure in reanalyzing the classification of children labeled as having LD.[3] For the LD sample, the first subgroup showed minimal discrepancy between S-CPT and
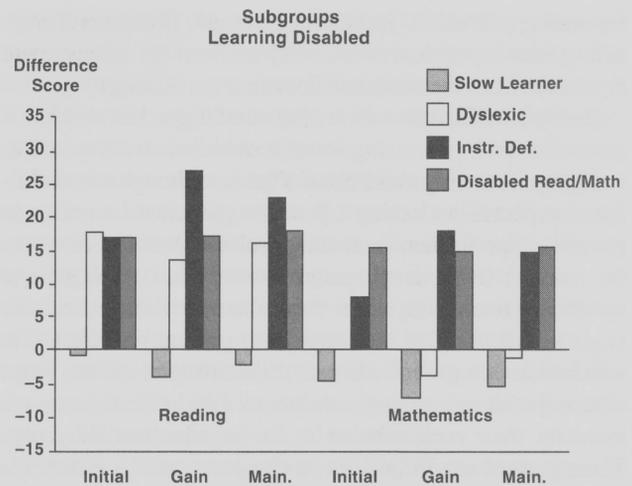

FIGURE 1
Discrepancy Scores as a Function of Subgroups

achievement scores. This group performed poorly on all classification variables. Most important, the mean potential score (Gain score) was at the same level as the reading and math scores. This group was defined as *slow learners*.[4]

Subgroup 2 showed no important discrepancies between mathematics performance and S-CPT scores, but clear discrepancies existed related to reading. Subgroup 2's achievement scores are characteristic of dyslexic children identified in the literature (Siegel & Ryan, 1989). With their average performance in math coupled with average performance in reading, this group was defined as dyslexic or children with *specific LD in reading*.

The third subgroup showed a *large* achievement discrepancy between achievement and gain and maintenance scores when compared to the discrepancy between achievement and initial scores. This group also had the highest Gain and Probe scores when compared to the other subgroups in the school-

---

[2] The reader should note that we did not cluster the groups in terms of discrepancies between component scores (verbal versus spatial; episodic versus semantic, retrospective versus prospective errors) as a function of static versus dynamic testing conditions. This was because we were trying to determine global patterns within in a school-identified LD sample. Such a microanalysis, however, should be done for a more refined classification of learning disability subgroups (e.g., the criteria of an LD subgroup under dynamic testing conditions are that some component scores are in the average range and discrepant from the poor achievement score, whereas the deficit component matches the deficit achievement score.

[3] Remember the classification criteria for LD using dynamic classification criteria: (a) processing score that is stable across initial and dynamic testing conditions, and (b) general processing potential (gain and maintenance) score that is discrepant from academic achievement. No doubt, a classification scheme also can be argued at the specific process level (e.g., verbal processing score is below average and stable across initial and dynamic test-

ing conditions, whereas the visual spatial score is average or above average under both testing conditions, and/or (c) a processing potential score at a component level (visual-spatial) is discrepant from an academic achievement and verbal processing score). For a representation of how dynamic scores can be included in discrepancy formula as a measure of potential, however, global scores (similar to the Full-Scale WISC-III) were used to illustrate discrepancies in performance.

[4] One also may create the argument that because all children have IQ scores > 85, S-CPT standard scores should *match* the low-achievement standard scores for the LD classification to hold (i.e., the slow-learning group). The problem with this definition, of course, is that the discrepancy model (difference between processing potential and academic achievement) is completely eliminated from the classification (i.e., a key construct of LD). In addition, if IQ is used only as a basal measure to separate LD children from those with mild retardation, then approximately 91% of the LD sample were misclassified—a situation that may severely restrict access to special services.

identified LD sample. These children were defined operationally as *instructionally deficient* because of their high responsiveness to probes when compared to other subgroups (as well as their above average scores on the S-CPT, based on the standardization sample).

Subgroup 4 showed a large discrepancy (similar in range to subgroups 2 and 3) in S-CPT scores and achievement in both reading and mathematics. Their Gain and Probe scores, however, were in the low-average range. This subgroup was defined as having *LD in reading and mathematics*.

Thus, based on the cluster analysis, approximately 17% of the sample's (subgroup 3) processing performance is easily modified and therefore reflects instructionally deficient children, whereas 9% of the poor-reading sample (subgroup 1) were best classified as slow learners. Further, the discrepancies in subgroup 2 and 4 between potential (Gain scores) and achievement scores clearly are distinguishable from slow learners (i.e., no significant discrepancy emerged in this sample).

Heterogeneity in performance also existed within the skilled reading sample. A cluster analysis was computed on the skilled reading sample, which yielded four subgroups. The four subgroups that emerged were classified as gifted (*high* Gain scores), low math achievers (*low* scores in math), skilled math achievers (*high* math scores), and instructionally responsive children (*high* probe scores).

Achievement and processing. The subsequent analysis compared the subgroups across the two reading samples. To simplify the reporting of results, children with LD were referred to as subgroup 1 through subgroup 4 (slow learner, LD reader, instructionally deficient, and LD reading/math, respectively), and skilled readers were referred to as subgroups 5 through 8 (gifted, low math, skilled math, and instructionally responsive, respectively). Also as expected, significant differences were found between the eight subgroups in reading and mathematics. Because of the classification criteria, none of the subgroups from the LD sample scored comparably to skilled readers on the reading measure. Two subgroups, however, did score comparably to skilled readers on the mathematics measure. For the two classification variables for subgroup determination from the S-CPT, some statistically comparable scores were found between subgroups of skilled readers and LD children on scores from the dynamic testing condition: Gain and Probe performance.

Demographic differences. No significant differences were found between subgroups related to gender. In terms of ethnicity, however, clear differences emerged between subgroups. This is a critical finding because some children of minority descent are placed inappropriately in LD classrooms. Minority children are highly represented in the instructionally deficient group in the LD sample, and in the instructionally responsive subgroups in the skilled reading group. This find-

ing implies that some minority children are classified incorrectly as having learning disabilities. Further, some minority children respond readily to examiner assistance, providing a better estimate of their information-processing performance. In essence, a high representation of children are labeled incorrectly as having learning disabilities.

External validity. The external validity of the above classifications was analyzed by comparing the subgroups on variables not used in the original classification. The four composite scores not related to classification variables (initial, maintenance, processing difference, stability) yielded significant differences between subgroups. These differences are shown in Figure 2. Differences were found between subgroups on initial, maintenance, processing difference, and stability scores. As shown in Figure 2, these results indicate that subgroup 1 is consistently low and subgroup 5 is consistently high on measures *external* to the classification variables. Subgroup 3 clearly was discriminated from the other subgroups in the LD sample on processing and stability scores, but *not* initial scores.
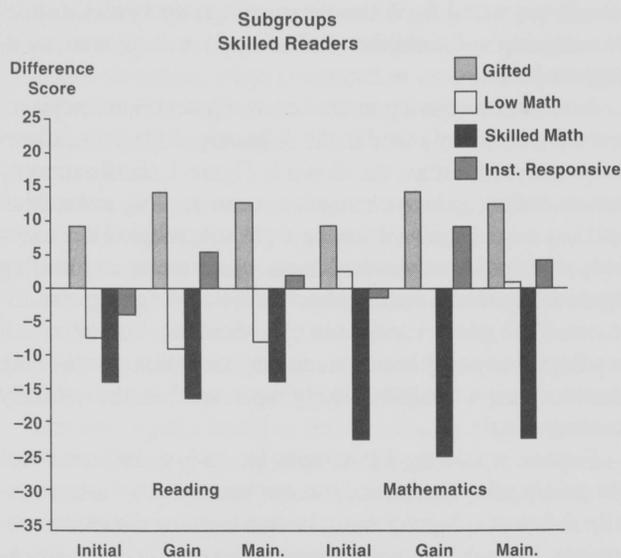

FIGURE 2
Processing Scores as a Function of Subgroups

In sum, some subgroups were easily discriminated on measures independent of the classification variables, and some subgroups were not. Three important findings related to classification issues emerged. *First*, except for mathematically gifted children (subgroup 5) and slow learners (subgroup 1), *no differences emerged among the other subgroups on general scores from the initial testing condition.* This is important because the static or initial score condition characteristic of most assessment approaches did not clearly differentiate

the subgroups of poor readers (a finding analogous to the notion that children with and without LD are in the normal range of intelligence). On measures of general potential (Gain scores), however, clear differences emerged between subgroups.

Second, instructionally disabled (subgroup 3) children in the poor reading sample scored similarly to the instructionally efficient subgroup (subgroup 8) in the normal reading sample on Probe and Gain scores, as well as measures not used in the classification of subgroups: maintenance, processing difference, and stability scores. This finding is important because some children school-classified as LD are comparable to normally achieving children in general information processing ability and, therefore, are classified incorrectly.

Finally, scores between children with reading disabilities and children with both reading and math disabilities (subgroups 2 and 4) were comparable to low and high math performers (subgroups 6 and 7) in the skilled reading group across all processing conditions under both dynamic and static testing conditions. These children, however, were clearly separated from slow learners (group 1) and instructionally responsive children (subgroup 3) under dynamic testing conditions.

Because discrepancy scores between potential and achievement are commonly used in the diagnosis of LD, these scores were analyzed further. As shown in Figure 1, six discrepancy scores (initial, gain, maintenance minus reading and mathematics) were calculated for the eight subgroups. The results indicated that discrepancies between gain scores and reading (gain scores minus reading) best predicted subgroup classification (45% of the variance in classification). For the initial/reading discrepancy scores, three subgroups from the learning disabled sample could be clearly separated from the normally reading sample.

Further, within the LD sample, an analysis indicated that the gain/reading discrepancy scores separated the instructionally deficient subgroup from the two learning disability subgroups. Thus, a *larger discrepancy emerged between teaching-deficient students and learning disability subgroups that were deficient primarily in reading (subgroup 2) or reading and math (subgroup 3)*, even though these subgroups were comparable in reading performance. No clear discrepancies between achievement and S-CPT scores occurred for the slow-learning sample during the initial or dynamic testing conditions.

In general, Study 2 yielded two important findings. *First*, the findings support the notion that processing differences between some children with or without LD are not primarily a function of instruction or environmental factors but, instead, reflect processing variables intrinsic to the child. Further, some evidence indicates that skilled readers lose some of their processing advantage when compared to some subgroups of children with reading difficulties on some processing conditions (Gain conditions).

Second, the S-CPT was able to identify children who fit an *a priori* definition of LD. This definition focuses on (a) the *stability* of general processing from the initial to gain conditions, and (b) the *discrepancy* between general-processing potential, as reflected in gain condition, and academic achievement. In terms of processing stability, the results indicated that processing stability was average for three subgroups (dyslexic, slow learners, LD reading/math) in the learning disability sample, whereas one subgroup (instructionally deficient) exhibited high or above-average modifiability in processing performance.

In terms of performance discrepancies, the results indicated that the discrepancies between potential (Gain scores) and achievement were high for three subgroups (dyslexic, instructionally deficient, LD reading/math) in the learning disability sample, whereas one subgroup (slow learners) exhibited minimal performance discrepancy.

Thus, *only* two subgroups (dyslexic and LD reading/math) met both criteria for classification of LD. These two subgroups also were distinct from subgroups in the normal reading sample. That is, although two subgroups in the skilled reading sample had stability scores comparable to the two learning disability subgroups, their discrepancy scores reflected *higher* achievement than potential scores. In sum, the results suggest that Subgroup 2's and 4's performance is relatively stable and less amenable to brief intervention when compared to other groups. Further, their potential scores are clearly discrepant from their achievement.

Back to the Substantive Issues of Discrepancy Scores

Does Study 2 help address the issues we raised earlier about validating the use of discrepancy between potential and achievement as a basis for classifying children as having learning disabilities? Based on the previous discussion, five substantive issues about the validity of discrepancy scores in the classification of LD were raised. How does Study 2 help address these issues?

1. *Construct integrity.* Because the S-CPT is geared to gather scores that estimate learning potential, the S-CPT is more likely than traditional IQ measures to "tap" stable as well as "real" discrepancies between potential and actual achievement (Study 1; also see Swanson, in press a). That is, Study 1 suggested that because scores from the dynamic testing condition better predicted achievement than general IQ scores, they may provide an appropriate measure of "potential" in discrepancy formulas.

In addition, as shown in Study 2, the S-CPT makes a distinction between examinees who are receptive to minimal intervention (gainers) versus those who are not. Children who have low reading scores and also show minimal variation in discrepancies for initial, gain, and maintenance versus academic performance would be better candidates for LD classification than those who are "clear gainers" under dynamic testing conditions.

As stated earlier, the beginning step in capturing the notion of a discrepancy is to match the measure to the construct definition. If *potential* is best defined as gain or maintenance processing performance, such information seems to have more meaning than a static measure in classification procedures.

Thus, what we are proposing is an alternative measure of potential. Rather than using traditional IQ tests, we suggest that scores based on gain performance be included in the classification of LD. This alternative addresses three problems with traditional assessment procedures for classifying children as learning disabled.

First, examinees are assumed to have been provided instruction systematically. Unfortunately, current classification approaches make no attempt to test directly, with standardized instruments, an examinee's responsiveness to instruction; therefore, it cannot be assumed that the discrepancy reflects a stable processing disorder rather than a lack of exposure to instruction or systematic instruction.

Second, IQ measures are included in discrepancy formulas on the assumption that they are measures of potential. In many cases, however, IQ measures reflect the content of achievement tests, although the processes on such measures may be indirect (e.g., arithmetic subtest on the WISC-III and arithmetic tests on various achievement measures). Finally, discrepancies based on IQ and achievement scores in many cases are not stable (e.g., Shaywitz, Fletcher, Holahan, & Shaywitz, 1992) or reflect normally occurring events in the general population.

2. *Independence of measures*. Previous research has indicated that the commonality in profile between a discrepancy group and a nondiscrepancy group on a number of processes is related to shared low scores in reading. Although reading scores and initial scores were comparable among the subtypes with poor reading, it does not appear that performance related to responsiveness (probe scores), processing potential (gain scores), and maintenance were. Thus, distinct profiles in processing performance emerged between the poor reading groups (i.e., school-identified children with LD).

Dynamic assessment views performance potential as modifiable rather than constant. The general form of assessment represented in the S-CPT is: test, provide help, test, provide help, and so on, which is much like saying "assess, instruct, assess, instruct." Therefore, the assessment focus is *not* on what individuals can display when working alone but, rather, what they are capable of doing with escalating levels of assistance. This range from a high level performance to a partial level of performance is what Vygotsky (1978) termed the "zone of proximal development" (ZPD). It should be obvious that instruction within the ZPD is desired for classification purposes. Although obvious in principle, it has been ignored in traditional assessment practice.

3. *Direction of difference scores*. A critical assumption in testing the discrepancy model is that differences in the direction of the profiles are important. The fact that *student A* has an above-average reading score but an average Gain score (negative discrepancy) should reflect a "different set of or level of" process variables in terms of performance than *student B* with an average gain but low reading score (positive discrepancy).

As shown in Study 2, readers with high discrepancy scores between reading/math and Gain scores had higher processing scores on measures independent of the classification measures, when compared to examinees who had higher reading/math scores compared to their Gain scores. This finding suggests that the direction of the discrepancy is important.

4. *Comparable performance between groups is not the same as the relationship among variables within groups*. We cannot address this assumption from the information present in Study 2. In a *post hoc* analysis, we looked at the best predictors (all processing scores) of reading within each subgroup. Clear differences emerged between subgroups (also found in Swanson, 1993b; 1993c). Although several previous studies failed to find processing differences between discrepancy and nondiscrepancy groups, it is possible that a distinct functional relationship among processes exist between the groups. There will always be the possibility of co-occurrence (both groups are low and statistically comparable on various measures), but how the processes are shared within each ability group must be distinct to validate the use of discrepancy scores.

5. *Measure something beyond component parts*. Although the subgroups in Study 2 differed in their discrepancy between initial and reading/math measures, the subgroups clearly reflect a different profile in responsiveness to probes. This is important because it suggests a link between assessment and instructional procedures. That is, the S-CPT attempts to combine assessment with instruction, for while students are being evaluated, they also are being taught strategies about the targeted information.

The S-CPT provides partial answers to questions such as: "How much and what kind of information is needed to improve or help an examinee improve her or his learning performance?" The S-CPT assumes that the key link between assessment and intervention is an understanding of the child's sensitivity to change. Thus, the best way to learn about a child's responsiveness to instruction is to assess the effects of experimentally induced change on performance of processes strongly related to instruction.

The assessment approach used in the S-CPT also has much in common with instruction by the classroom teacher. There is (a) an active role for the examiner, (b) a collaborative interaction between the examiner and the child, and (c) a deliberate effort to change what is being assessed. The assessment process yields important information about the kinds of help needed to be successful on various tasks.

To my knowledge, no data suggest that a certain profile on DA measures would warrant a certain type of instructional approach. It seems unlikely, however, that children completely unresponsive to probes and cues would be responsive to methods and materials that emphasize active or constructed learning strategies (e.g, reciprocal teaching, learning strategies) or highly interactive teacher/student approaches. These children would most likely benefit from procedures that place minimal demands on constructing strategies. The approaches would emphasize relevant conceptual knowledge (via drill and practice), motivation, and programmed (e.g., computer-mediated) instruction.

Given these caveats, the S-CPT informs the teacher about instruction for the examinee in a least five ways:

1. Effectiveness of simple feedback on examinee performance
2. Examinee's general knowledge of strategies (procedural knowledge)
3. Degree to which instruction is maintained after hints are removed
4. Examinee's flexibility to simple intervention
5. Examinee's preference for verbal or visual-spatial information, which in turn may inform the teacher about the type of cues needed to bolster the maintenance of information.

SUMMARY

Results of the two DA studies reviewed raise an important question as to why DA procedures are able to separate out slow learners and instructionally disabled children from children operationally defined as having learning disabilities. A second question is why children with LD were generally inferior in performance to average readers under dynamic testing conditions. To answer these two questions, let's consider the goals of DA.

Embretson (1987) described three goals of DA: "(1) improving ability estimates; (2) assessing new constructs, and (3) improving true ability" (p. 167). These assumptions may be plausible in evaluating the sensitivity of DA measures in predicting performance in children with LD.

The first explanation is that the DA measures simply provide a better indicator of ability group differences than the initial testing conditions. As shown in Study 1, the DA measures were simply a more sensitive indicator of achievement than other measures, and therefore the ability group differences in Study 2 reflect true differences.

An alternative explanation, however, is that the performance differences were an artifact of spreading out the scores. It is unlikely, however, that the ability group differences in Study 2 were simply an artifact of inducing variance. For example, Gain scores for examinees operationally defined as LD readers were not that different from initial scores. Likewise, a significant number of normal achievers improved their gain scores, whereas other normal readers did not. Therefore, instead of inducing variance or spreading out the scores, the measures were sensitive indicators of processing potential in some students and not others. Thus, a more plausible explanation is that the initial (static) testing conditions *prevented* many students from achieving greater success on the processing measures.

The second explanation is that DA measures tap new abilities: modified performance. To determine if probing procedures in Study 2 modified performance, a post-hoc analysis of effect size was calculated from the raw scores in the learning disability and skilled reading sample. Effect sizes were computed for gain (gain mean score-initial mean score/standard deviation of initial scores) and maintenance conditions (maintenance mean score-initial mean score/standard deviation of initial scores). The effect size for raw scores averaged approximately 1.5 standard deviations for Gain conditions and about .5 standard deviation for maintenance conditions in both ability groups. These post-hoc findings suggested that responsiveness to probes was not simply an artifact of "reading ability," thereby suggesting that a modification in processing ability occurred for both ability groups.

Finally, DA influences children's information-processing ability. That is, dynamic testing procedures were expected to produce changes in ability group classification because it is assumed that many psychological entities are not static (e.g., Carlson & Wiedl, 1979; Haywood et al., 1990). The results clearly support the notion that changes in processing ability occurred across some ability groups. For example, the results indicate that some generally "inefficient" information processors (such as LD) were influenced by procedures that facilitate access to previously stored information. Further, some DA measures were more likely to yield high R^2 values than

the initial scores in discriminating among ability group classifications, suggesting that new abilities were being tapped.

In summary, measures of processing must be developed to capture the subtle processing differences between children with and without LD. The two studies we considered, however, suggest the applicability of DA to the measurement of learning potential and provide further evidence regarding the relationship between performance on information-processing tasks and the classification of LD.

REFERENCES

Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.

Bateman, B. (1965). An educator's view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.) *Learning Disorders* (Vol. 1, pp. 219-239). Seattle: Special Child.

Bateman, B. (1992). Learning disabilities: The changing landscape. *Journal of Learning Disabilities, 25*, 29–37.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison: University of Wisconsin Press.

Bransford, J. C., Delclos, J. R., Vye, N. J., Burns, M., & Hasselbring, T. S. (1987). State of the art and future directions. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 479-496). New York: Guilford Press.

Brown, A., Campione, J. C., Webber, L. S., & McGilly, K. (1992). Interactive learning environment: A new look at assessment and instruction. In B. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 121-211). Boston: Kluwer Academic Publishers.

Brown. A., & French, L. (1979). The zone of potential development: Implications for intelligence in the year 2000. *Intelligence, 4*, 255–273.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic assessment: Foundations and fundamentals* (pp. 173–195). New York: Guilford Press.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: Foundations and fundamentals* (pp. 82-115). New York: Guilford Press.

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence, 9*, 297–315.

Cantor, J., & Engle, R. W. (1993). Working memory as capacity in long term memory activation: An individual difference approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 9*, 1101–1114.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*, 229–246.

Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing the limits employing the Raven Matrices. *Intelligence, 3,* 323–344.

Carpenter, P. A., Just, M., & Shell, P. (1990). What one intelligence test

measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

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

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 74,* 68–80.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450–466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25,* 1–18.

Embretson, S. E. (1987). Toward development of a psychometric approach. In C. Lidz (Ed.), *Dynamic assessment: Foundations and fundamentals* (pp. 141-172). New York: Guilford Press.

Engle, R., Cantor, J., & Carullo, J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology, 18,* 972–992.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies in proximal development. *Child Development, 57,* 1087–1099.

Feuerstein, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore: University Park Press.

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.

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

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Davidson, K. C., & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. *Journal of Consulting & Clinical Psychology 59,* 27–37.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory & Language, 28,* 200–213.

Gerber, P. I., & Reiff, H. B. (1991). *Speaking for themselves: Ethnographic interviews with adults with learning disabilities*. Ann Arbor: University of Michigan Press.

Hammill, D. D. (1990). On defining learning disabilities: An emerging consensus. *Journal of Learning Disabilities, 23,* 74-84.

Hammill, D. D., Leigh, J. E., McNutt, G., & Larsen, S. C. (1981). A new definition of learning disabilities. *Learning Disability Quarterly, 4,* 36–42.

Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psychoeducational assessment. *School Psychology Review, 19,* 411–422.

Jastak, J. J., & Wilkinson, S. (1984). *Manual: The Wide Range Achievement Tests*. Wilmington, DE: Jastak Associates.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior & Human Performance, 27,* 44–46.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working-memory. *Psychological Review, 99,* 122–149.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more) working-memory capacity ?!. *Intelligence, 14,* 389-433.

Lord, F. M. (1956). The measurement of growth. *Educational & Psychological Measurement, 16,* 421–437.

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

Morrison, S. R., & Siegel, L. S. (1991). Learning disabilities: A critical review of definitional and assessment issues. In J. E. Obrzut & G. W. Hynd (Eds.), *Neurological foundations of learning disabilities: A handbook of issues, methods, and practice* (pp. 79–95). San Diego: Academic Press.

Naglieri, J. A., & Das, J. P. (1988). Planning-arousal-simultaneous-successive (PASS): A model for assessment. *Journal of School Psychology, 26,* 35–48.

Palincsar, A., Brown, A., & Campione, J. (1991). Dynamic assessment. In H. L. Swanson (Ed.), *Handbook on the assessment of learning disabilities: Theory, research and practice* (pp. 75-95). Austin, TX: Pro-Ed.

Reynolds, C. R. (1984). Critical measurement issues in learning disabilities. *Journal of Special Education, 18,* 453–476.

Reynolds, C. R. (1985). Measuring the aptitude-achievement discrepancy in learning disability diagnosis. *Remedial & Special Education, 18,* 453–476.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-210). New York: Springer-Verlag.

Rousch, W. (1995). Arguing over why Johnny can't read. *Science, 267,* 1896–1898.

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

Shallice, T. (1988). *From neuropsychology to mental structures.* Cambridge, MA: Cambridge University Press.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition, 24,* 139–168.

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.

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

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

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

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

Siegel, L. S., & Ryan, E. B. (1989). The development of working-memory in normally achieving and subtypes of learning disabled. *Child Development, 60,* 973–980.

Sovik, N., Frostad, P., & Lie, A. (1994) Can discrepancy between IQ and basic skills be explained by learning strategies. *British Journal of Educational Psychology, 64,* 389–405.

Sparks, R. L. (1995). Phonological awareness in hyperlexic children. *Reading & Writing, 7,* 217–235.

Speece, D., Cooper, D., & Kigler, J. (1990). Dynamic assessment, individual differences, and academic achievement. *Learning & Individual Differences, 2,* 113–127.

Stanley, J. C. (1967). General and special formulas for the reliability of differences. *Journal of Educational Measurement, 4,* 29–252.

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

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26,* 7–29.

Stanovich, K., & 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.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* Cambridge: Cambridge University Press.

Swanson, H. L. (1988). Toward a metatheory of learning disabilities. *Journal of Learning Disabilities, 21,* 196–209.

Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology, 84,* 473–488.

Swanson, H. L. (1993a). An information processing analysis of learning disabled children's problem solving. *American Educational Research Journal, 30,* 861–895.

Swanson, H. L. (1993b). Individual differences in working memory: A model testing and subgroup analysis. *Intelligence, 17,* 285–332.

Swanson, H. L. (1993c). Working memory in learning disability subtypes. *Journal of Experimental Child Psychology, 56,* 87–114.

Swanson, H. L. (1995). *Swanson-cognitive processing test.* Austin, TX: Pro-Ed.

Swanson, H. L. (in press a). Effects of dynamic testing in the classification of learning disabilities: The predictive and discriminant validity of the S-CPT. *Journal of Psychoeducational Assessment.*

Swanson, H. L. (in press b). Using the S-CPT to assess processing ability: The development of a dynamic assessment measure. *School Psychology Review.*

Swanson, H. L., & Christie, L. (1994). Implicit notions about learning disabilities: Some directions for definitions. *Learning Disabilities Research & Practice, 9,* 244–254.

Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The influence of memory and classification ability on children's word problem solution. *Journal of Experimental Child Psychology, 55,* 374–395.

USOE. (1994). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42,* 65082–65085.

Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds. and Trans.), *Mind in society: The development of higher psychological processes* (pp. 79-91). Cambridge, MA: Harvard University Press. (Original work published 1935)

Wall, T. D., & Payne, R. (1973). Are deficiency scores deficient? *Journal of Applied Psychology, 58,* 322–326.

Wanous, J. P., & Lawler, E. E. III (1972). Measurement and meaning of job satisfaction. *Journal of Applied Psychology, 56,* 95–105.

Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement, 19,* 149–154.