

Assessment of Socio-Culturally Diverse Students: Problems in Special Educational Theory and Implications for Practice

Judith K. Bernhard

Ryerson University

digital.library.ryerson.ca/object/386

Please Cite:

Bernhard, J. K. (1990). Assessment of socio-culturally diverse students:
Problems in special educational theory and implications for practice. *The
International Journal of Dynamic Assessment and Instruction*, 1(2), 86-104.



Assessment of Socio-Culturally Diverse Students: Problems in Special Educational Theory and Implications for Practice

Judith Bernhard-Kutscher

O.I.S.E., Special Education

The present socio-cultural as well as linguistic diversity of many newcomers to Canada brings important issues in special education for critical consideration. While educators have spoken of the need to consider ethnic and cultural diversity in assessment and placement decisions, there is currently a lack of criteria for distinguishing genuine learning disabilities from the normal language barriers associated with the process of second language acquisition. A critical analysis is presented of some of the core concepts on which current assessment practices are based; these include intelligence and learning disability models and some of the most common tests and test batteries used in connection with socio-culturally diverse as well as "mainstream" students. It is concluded that the assessment tools currently used rest on dubious constructs and have questionable validity. This suggests that the segregation of children labelled LD lacks a proper rationale. Because of the many problematic areas in current special educational assessment practices, new assessment/learning paradigms are needed which accept diversity as a basic assumption and employ dynamic approaches as have been derived from the Vygotskian model.

Background and Statement of the Problem

One of the most dramatic changes affecting the Canadian educational system in the past few years has been the trend toward a culturally heterogeneous and linguistically diverse school-age population. In major areas of Canadian urban centers, the number of children with a first language other than English, and having a lack of proficiency in English, is approximately 40%, and future projections suggest that this trend will continue. In the United States the issue of the extent to which these students are being properly assessed and placed has been a major concern. In Canada, however, the field of special education has developed independently of areas such as multicultural education, and has not been that responsive to the rapidly changing demographics. Consequently, there has been surprisingly little attention given to this area.

In dealing with the problems of education in a multicultural society, one of the most striking is that certain cultural groups are highly over-represented in "Basic" or special ed-

The author wishes to acknowledge H. White, J. Cummins, L. Siegal and J. Weiner for their many insightful comments and suggestions in the preparation of this manuscript.

ucation programs. Only the latter programs will be discussed in this paper. Following a study by Mercer (1973) which documented that 80% of the group classified as educable mentally retarded (E.M.R.) originated from particular socioeconomic and ethnic sub-groups (largely Hispanic), various efforts and initiatives were designed to correct the disproportion of minority student enrollment in special education classes in the United States. For example, U.S. legislation PL 94-142, *The Education of All Handicapped Children's Act*, guarantees all handicapped students the right to testing that is not culturally-discriminatory. Nevertheless, 15 years after the drafting of PL 94-194, Ortiz and Yates (1983) reported that in Texas, Hispanics were over-represented in LD classes by over 300%. *Disproportionate representation*,¹ by itself, is not evidence of culturally-unfair or poor assessment practices, nor for that matter, of harm inflicted on the groups that are being disproportionately represented. It is possible to argue, in the manner of Jensen, that the over-representation of Blacks in EMR classes represents the actual excess number of Blacks whose IQ is measured in the vicinity of the 85 (i.e., low average) level.

The contrasting position is that disproportionate representation is *prima facie* evidence of questionable practices and/or culturally-unfair testing. In other words, the testing does not merely *discriminate* in a neutral sense, but that the testing is, in fact, *discriminatory* in the moral or legal sense. The cultural unfairness² of a given test, battery of tests, or assessment procedures, however, is a matter for empirical demonstration; more precisely, the confounding of particular psychometric measures by cultural factors is an issue to be addressed through theoretical clarification and examination of relevant data. If, in fact, it is true that over-representation is evidence of poor and/or discriminatory assessment practices, then there is much to be concerned about.

Various studies have documented a discrepancy between policy and practice in the assessment of U.S. minority students. Figueroa (1982) investigated practices related to Hispanic students in special education. Interviews with school personnel indicated that they were highly aware of the issues and believed that assessment practices were exemplary with Limited English Proficiency (LEP) students referred for special education. Nevertheless, in direct observations and file reviews, it was evident that there was minimal consideration of language or culture factors in assessment, placement, or instructional practices.

The phenomenon of disproportionate direction of students from certain minorities into special education programs has raised a number of theoretical and research issues. For example, Coles (1978) has argued strongly that there is a lack of demonstrable validity for the conventional assessment procedures used to diagnose learning disabilities. Gamlin (1989) has discussed the connection between this phenomenon and the use of static as opposed to dynamic measures of assessment. Cummins (1989) has proposed a theoretical framework for explaining academic failure partially as a function of pedagogically induced factors and has suggested an "intervention" model. He cites examples of programs where different pedagogical interventions resulted in dramatic improvements in student achievement (e.g., the Kamehameha Early Education Program (KEEP, p. 75).

Mercer (1978-1979) has argued that the traditional psychometric instruments provide a scientific rationale for the dominant group to restrict minority group members by 'demonstrating' that they do not have the language and knowledge skills to succeed in a White cultural setting. Mercer has proposed an alternate way of adding to and

¹Defined here as a wide discrepancy between the percentage of a minority group in a particular program and the percentage of that minority group in the given age group (and population at large).

²The concepts of cultural fairness or unfairness of a test are broader than those of cultural bias, as narrowly defined by Jensen (1980). The former concepts are those mainly used in the present paper.

interpreting the data of such instruments. Rueda (1989) has argued that these problems can be adequately addressed only through the development of new paradigms and educational-systemic changes. Evidence will be briefly adduced below supporting these positions, and reflecting a need for increased awareness of the issues by psychologists, educators, and teacher-trainers.

The Canadian Context

Although it is important to note that the United States and Canadian populations differ in many important ways, there is reason to believe that the same issues related to U.S. minority assessment and placement are also characteristic of the Canadian situation. In major areas of Canadian urban centers, the number of children with a first language other than English, and having a lack of proficiency in English, is approximately 40%, and future projections suggest that this trend will continue. For example the findings of a 1988 Vancouver survey (cited in Ashworth, Cummins, & Handscombe, 1988) indicates that 46.9% of the total school population were designated as ESL; two-thirds had Chinese, Vietnamese, or East Indian as first languages. Similarly, the Toronto Board of Education reported that 33% of its students were not native speakers of English (Cheng, Tsuji, Yau, & Zigler, 1987). A related finding is reported in a 1986 Vancouver School Board survey (cited in Ashworth *et al.*, 1988) which indicated that more immigrant students arrived in Canada prior to age 5 than within any other given age group. It is therefore likely in these urban centers that among children in the elementary grades, a great many would have problems with English proficiency. Such problems would bias the results on commonly used, standardized test instruments (designed for American children).

In spite of heightened awareness of the need to consider ethnic and cultural diversity in assessment and placement decisions, a survey by Samuda and Crawford (1980) indicated that only a few Ontario Boards of Education had any policy regarding assessment and placement of minority students. Another study by Cummins in 1980 (reported in Cummins, 1984) analyzed the psychological assessment files of more than 400 students from minority language backgrounds. In examining the assumptions that psychologists and teachers made regarding the assessment of ESL students, it was found that significant knowledge gaps were evident regarding issues such as limitations of intelligence tests (the WISC-R in particular) and patterns of language and academic development among minority students. Thus Cummins concluded that psychologists' interpretation of WISC-R scores were often "reduced to intuition" (p. 51).

The issue of racially or culturally-biased testing has begun to be addressed in Ontario. Bill 82, unlike its American counterpart PL 94-142, made no specific provisions or attempts to avoid biased assessment and placement of minority students. Nevertheless, various recent Ontario provincial policy documents have addressed the issues. The most comprehensive of these is a recent report from the Provincial Advisory Committee on Race Relations entitled "The Development of a Policy on Race and Ethnocultural Equity" (Ministry of Education, 1987). One of the nine areas of concern identified by the committee summarizes the Ministry's position regarding the assessment and placement of minority students (emphasis as in original):

THE MINISTRY OF EDUCATION RECOGNIZES THAT WHILE IT IS MINISTRY POLICY THAT ALL STUDENTS SHOULD ENJOY EQUAL OPPORTUNITY TO DEVELOP THEIR INDIVIDUAL ABILITIES AND INTERESTS, CONCERN IS BEING EXPRESSED THAT YOUNG PEOPLE FROM CERTAIN RACIAL AND/OR ETHNO-CULTURAL GROUPS ARE BEING STREAMED INTO INAPPROPRIATE PROGRAMS.

The Ministry of Education encourages all school boards to put policies in place to monitor the assessment and placement of children, facilitate the involvement of parents in this process and assess the equity of student services.

Research reveals that many standardized tests (achievement, aptitude and psychological) are racially and culturally biased (p. 18).

These policy deliberations acknowledge serious problems needing to be addressed in the Canadian context.³ If solutions are to be found, it is important that they rest upon sound science and valid psychoeducational theory and constructs. Hence, there will be, in the discussion following, a critical analysis of some of the core concepts on which current assessment practices are based; in particular, the focus will be on some of the most common instruments and practices used in connection with socio-culturally diverse as well as "mainstream" students. Problems of scientific and psychological validity, of course, have implications for assessment of majority as well as minority students. In fact, the entire field of special education is affected by these issues. Specifically this paper addresses the following questions:

1. What is the construct validity of IQ measures? (or, what is the validity of the IQ construct?)
2. What is the validity of the LD construct?
3. How are attempts to measure these purported abilities, traits or conditions confounded by ethnic, cultural, and racial factors?⁴
4. How successful is one significant attempt to design a culturally-fair assessment framework for mental abilities?

The following are some important qualifications, preliminary clarifications, and limitations with respect to the present study. 1) The position taken in this paper is that the construct validity of a proposed measure of some purported ability or characteristic is to be demonstrated by the proposers. This is the position taken by many psychometric researchers (Cronbach, 1970). It follows then that it is sufficient for the purposes of this paper to determine the degree to which the proposers have succeeded in carrying out the demonstration of such validity. 2) Some selected researchers cited below have argued and presented evidence that a) labelling is harmful and b) placements and subsequent interventions made on the basis of such tests do not merely lack a proper psychometric foundation but are also harmful. While the author has a great deal of sympathy for this position, the assessment of the evidence of alleged harms or benefits, either of labelling or of specific interventions is beyond the scope of this paper. This paper does not attempt to provide definitive answers to the issues of assessment of socio-culturally diverse students, although certain promising directions and research attempts will be mentioned in the concluding portion. A goal of this paper is to reflect the complexity of the field and summarize some of the main issues with respect to the valid assessment of ethnic minority groups.

³The recommendations recognize the need a) to evaluate existing instruments for bias, b) to demonstrate an awareness and acceptance of various cognitive abilities—their multiplicity and variety of manifestations, c) to involve minority parents in the assessment procedures and make them aware of their rights, and d) to have procedures in place for the ongoing evaluation of student progress. This paper primarily focuses on a) and b); c) and d) are not discussed.

⁴Question 3 is addressed concurrently with questions 1 and 2.

Some Core Concepts in Special Education

What Is The Validity Of The IQ Construct?

The One-Many Factor Problem

The underlying assumptions in the development of instruments for intellectual assessment are based on different notions of intelligence and development. The traditional approach (e.g., Spearman, 1927; Terman, 1916) postulates that there is a general intellectual capacity called "general intelligence", "general mental ability", "general ability", or g ; and that in factor analysis of various intelligence tests, the single g factor accounts for a large proportion of the variance of scores. The rest of the variance can be accounted by one or more specific factors s . According to Spearman's theory (1927), the g factor is believed to be an index of general mental ability or intelligence. Tests with high g loadings involve complicated mental activities such as deductive thinking, reasoning, comprehension, and hypothesis-testing. The theory stated that g is a function of heredity and s is a measure of specific learning and experiences acquired by the individual. This two-factor theory has been the conceptual base for many measures of general and specific abilities.

The more recently developed intelligence tests have refined Spearman's theory. For example, Cattell and Horn (Cattell, 1959; Horn & Cattell, 1967) argue against a unitary conception of intelligence and postulate two differentiable types of intelligence; fluid and crystallized abilities. "Fluid intelligence" represents a biologically-based measure of general ability that is non-verbal and (supposedly), relatively culture-free. It involves adaptive and new learning abilities related to mental operations. Tasks such as figural analyses, matrices and pattern analysis are often used to measure fluid intelligence (e.g., Stanford-Binet Fourth Edition). "Crystallized intelligence" refers to acquired skills and knowledge that are dependent on exposure to culture. Examples of tasks that measure crystallized intelligence are tests such as vocabulary, general information, word analogies, and mechanics of language. This distinction has been widely accepted by psychologists (but see Brody, 1985) and is embodied in many tests, such as the Binet and the WISC-R. Although the specific terms are not used, the latter contains measures of both "fluid" (e.g., Block Design) and "crystallized" (e.g., Information, Vocabulary) intelligence.

Recently, Gardner (1983) has proposed that "we expand and reformulate our view of what counts as human intellect" (p. 4) and set out a somewhat similar framework.⁵ He outlines a theory of "multiple intelligences", namely, linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, and personal intelligences. Guilford (1985) has proposed a three-dimensional "structure of intellect" model and postulated the existence of 120 orthogonal independent dimensions of intellect. Other theorists have emphasized differences with respect to metacognitive abilities including the capacity to monitor cognitive processes, detect inconsistencies in reasoning, and to allocate time and intellectual resources to various components of a solution to a problem (for a review, see Sternberg, 1980).

It is obvious from this short review that there is currently some lack of agreement among theorists and psychometrists regarding the nature of intelligence and how it should be measured.

⁵However, Sternberg & Gardner (1982) are also in support of a general factor.

The Correlational Validation of IQ Tests

The main criterion used to validate intelligence tests is that there are substantial correlations of items with each other and with the IQ test as a whole. The word "ability" then, is used to mean whatever is being measured in the particular test. Thus the assumption that the individual items all measure the same construct, and those items that are not included do not require much intellectual ability. Nevertheless, as will be demonstrated in this section, the theory or rationale of the construction of intelligence tests actually precedes the "proof" of existence of general intelligence or *g*. It is therefore argued that these correlations are inappropriate for this sort of validation of intelligence tests.

The rationale for including particular kinds of items in a pool of items is not always obvious or explicit. Rather, it is guided by the author's own theory of what constitutes intelligence. Further, validity claims for a given test are buttressed by eliminating those items and subtests which have low correlations with the test as a whole and including items which have a high correlation with the test as a whole.⁶ Thus, only those items which correlate with the specific factors *initially consistent with the theory* are allowed to remain.

Even apart from the question of item selection, there is the question of how a given correlation matrix, derived from those items, is said to be "solved" or factored. According to Gould (1981), factor-analytic solutions of matrices are indeterminate and one can resolve the indeterminacy of the matrices in any of several ways without necessarily positing the existence of a general factor of ability. This is particularly so when using oblique or orthogonal relations.

Correlation of IQ with Various Measures of Success

The claim for the predictive validity of intelligence tests is often based on the high correlations with various measures of success, particularly with measures of school success. The argument, made by Jensen (1969, 1985) is as follows: Since success in prestigious occupations is based on intelligence, and IQ correlates with success, IQ tests must therefore be valid measures of intelligence. The implications of this syllogistic argument are that intelligence tests can be used to distinguish between those people who will do well in life and those that will not. Many psychologists believe that "giftedness", as measured by psychological tests, is a key factor in life success. Jensen, for example concluded that Terman and Oden's (1959) study proves beyond a doubt that children with high IQ were more likely to succeed in later life. Jensen stated:

One of the most convincing demonstrations that IQ is related to "real life" indicators of ability was provided in a classic study by Terman...Terman found that for the most part, these high-IQ children in later adulthood markedly excelled the general population on every indicator of achievement that was examined: a higher level of education completed; more scholastic honors and awards; higher occupational status; higher income; production of more articles, books, patents, and other signs of creativity; more entries in *Who's Who*; a lower mortality rate; better physical and mental health; and a lower divorce rate.... (Jensen 1969, p. 34).

Terman's studies do show a relationship between success and IQ scores but do not allow for sound attributions of causality according to Block and Dworkin (1976). An alternative hypothesis is that the children of the rich and educated had more

⁶According to Terman and Merrill (1960); "Tests that had low correlation with the total were dropped even though they were satisfactory in other respects" (Stanford Binet Intelligence Scale: Form L-M, p. 33).

opportunities in life, and these factors were largely responsible for their occupational success. There are many other factors (values, aspirations, money, white skin, opportunities) that could also be used as good predictors of success in life. These same factors also prevent poor children from doing well in intelligence and achievement tests. Nevertheless, these correlations would not lead psychologists to argue that values or white skin are ability factors.

Intelligence tests, it has been argued, are useful in predicting success in school-related tasks, and the relation between school grades and success is often explained in terms of intelligence. Yet McClelland (1976) is against this position and has pointed out that intelligence tests correlate more highly with proficiency in higher status jobs than with proficiency in lower status jobs. He writes:

Consider the fact that intelligence test scores correlate .08 with proficiency as canvasser or solicitor and .45 with proficiency as a stock and bond salesman.... But these two jobs differ also in social status, in the language, accent, clothing, manner, and connections by education and family necessary for success in the job (p. 49).

He concludes that "neither the tests nor school grades seem to have much power to predict *real competence* (emphasis added) in many life outcomes, aside from the advantages that credentials convey on the individuals concerned" (p. 56). He therefore advocates moving away from word skill tasks and paper and pencil problems toward direct testing of the various skills a given job may require. His rationale is summed up as follows:

It seems wiser to abandon the search for pure ability factors and to select tests instead that are valid in the sense that scores on them change as the person grows in experience, wisdom, and the ability to perform effectively on various tasks that life presents.... (p. 59).

Differential IQ Scores for Different Groups

Generally speaking, Americans labelled "black", even when equated by SES have been found to score approximately 15 points or one standard deviation lower than white Americans on standardized intelligence tests Jensen (1980; 1985) cites his own research in this regard and that of many others. The dispute has focused on whether the differences are primarily hereditary (biological) (e.g. Jensen, 1969), or environmental (Kamin, 1974; Mercer, 1978-1979; Rosenthal & Jacobson, 1968). Jensen (1969), for example, argues that the environment acts as a "threshold variable" on IQ. Extreme environmental deprivation can prevent children from performing up to their potential, but an enriching environment will not push the child beyond that potential. Thus his famous statement, "Compensatory education has been tried and it apparently has failed" (1969, p. 2). Kamin (1974), on the other side, has argued that there is no solid and conclusive data to support the belief that there is *any* genetic contribution to IQ differences. Also Gould (1981) writes on the misuse of putative psychometric and anthropometric data.

How are These Issues Confounded by Ethnic, Cultural, and Racial Factors?

The problems with test construction discussed above are exacerbated for minority groups in the sense that, members of many of these groups, tend to demonstrate inadequate performance levels on conventional intelligence tests. The results of these tests may be used to denigrate members of minority groups and to limit their educational and vocational opportunities. The problematic areas can be categorized into two distinct areas; intrinsic and extrinsic sources of bias. Each will be briefly outlined.

Intrinsic Bias

Although there have been attempts to eliminate test bias in item content, by definition, this is not possible. An example of such attempts was undertaken by the authors of the Stanford-Binet, Fourth Edition. The standardization sample comprised over 5000 cases and the stratification variables considered included geographic region, community size, ethnic group, age, gender and SES. Although there was a proportional representation of ethnic populations in the standardization sample, the relatively small number of "ethnic" subjects is arguably insufficient to remove bias—much of it built in—from the standardization process. The Composite Scores of whites are still 10–17 points higher than those of blacks and there is a 14-point difference between children in the highest and lowest SES groups.

Fluid Intelligence: The influence of culture on the fluid versus crystallized components also needs to be examined for minority students. It is claimed that, since fluid intelligence is a biologically based measure that is non-verbal, tasks such as spatial and figural analyses and matrices are more culture-fair than those tasks involving crystallized intelligence (Cattell, 1959). Nevertheless, even the results of the research of Cattell's associate, Horn have shown, according to Cattell (1985), that the lower intelligence scores of minorities are not entirely accounted for by lowered scores on measures of fluid intelligence:

Although I would tend to conclude, from other evidence [Horn, 1968], that *much* (emphasis added) of the black-white difference is located in $g(f)$, the present evidence leaves this as only a probability (Cattell, 1985, p. 227).

Feuerstein (1979), has argued that the issue of spatial organization contributes to tests lacking culture fairness. More relevant to the present argument however, Feuerstein (1979) has cited other sources of cognitive impairments such as impulsive and unsystematic exploratory behavior, deficiencies in problem recognition and comparative behavior, and "a lack of orientation toward seeking and projecting relationships, grouping, organizing, and summing events" (p. 234).

Crystallized Intelligence: Lack of acculturation in the mainstream culture will often result in depressed scores on tasks supposedly tapping crystallized intelligence. Cummins (1984), for example, showed that on the information subtest of the WISC-R, minority children obtained significantly lower scores than those of the norming sample.⁷ The ability to answer the kinds of questions asked in this subtest (e.g., Who invented the light bulb? How far is it from New York to Los Angeles?) depends on having the opportunity to acquire this information through "normal" home and school experiences of the middle-class North American society. Similarly, the Comprehension subtest assumes that the examinee is familiar with the situation and understands and has sympathy for the moral percept upon which the answer is based. For example, the question, "Why is it better to give money to a well known charity rather than a street beggar?" discriminates against cultures that have different sets of conventions (Ribeiro, 1980).

Extrinsic Factors Contributing to Poor Test Performance

Minority students' performance on intelligence tests can be significantly affected by various extrinsic factors such as the nature of the testing situation, affective-motivational variables, and differential cognitive styles. These will now be very briefly reviewed.

⁷Median scaled score of 4.9 as compared with the norm of 10; more than 1/3 of the sample scored 3 or below compared with 2.5% of the test norming sample.

An intelligence test can be described as a set of standardized questions taken in a judgmental situation with an authority figure. The anxiety often experienced by examinees can be exacerbated by the characteristics of the examiner. For instance, if the examiner is aloof, rigid, or has low expectations, scores can be significantly depressed (Zigler, 1973 cited in Clarizio, 1979). Studies have documented that scores on educationally related tests can be significantly improved for minorities by such techniques as employing ethnic examiners (Bernal, 1975; Garcia & Zimmerman, 1972), and different testing procedures (Johnson & Mihal, 1973).

The lack of familiarity with a testing situation is also likely to affect performance. Minority students are likely to be unaccustomed to dealing with information in ways required by a particular instrument. Feuerstein (1979) has noted cultural differences in communication modalities or the way an individual perceives a partner in a given transaction. He notes:

"...a lack of differentiation between his own and his partner's selves. Thus, in this relationship, he does not feel the need to spell out in a detailed and clear way what he thinks, and why, since he considers this as known to the other as it is to him" (p. 68).

The perception of the need for accuracy or precision is another example of a variable that is culturally determined and that can significantly affect test performance. Other studies have documented significant cultural differences in inculcation of attitudes regarding the value of education as a way of self-improvement. Other culturally determined factors such as achievement motivation, perceptions of self-efficacy, and locus of control also tend to influence test performance. Feuerstein (1979) summarizes these variables:

The dislike for the entire situation combined with the feeling that the test lacks relevance may readily depress an IQ score, possibly by a few points, as admitted by Jensen (1969), but more likely by tens of points, as we found in our clinical work (p. 61).

Summary

This review indicates that the validity of *g* or a 2-faceted *g* construct has not been definitively established. It can be questioned whether this is the best way of construing mental ability. Even if the construct had some *usefulness* with respect to the mainstream societal group, it is highly questionable as a basis for conceptualizing and measuring the intelligence of minority students. Plausible alternative models of conceptualizing and measuring intelligence are available. It is likely that these alternative models are more appropriate for all students, but in practice they may be especially beneficial for minority students.

What is the Validity of The Learning Disabilities (LD) Construct?

The field of learning disabilities (LD) is characterized by a lack of consensus regarding conceptualization, definition, etiology, diagnostic, and treatment procedures. Identification, selection criteria, and practices differ among practitioners, boards, and geographic areas. The difficulty in distinguishing LD from other conditions is further confounded by the effects of economic or cultural influences on achievement. Although there is little doubt that a substantial number of students have serious and unexpected problems in various areas of school achievement, there are great disparities in the numbers of identified individuals depending on the criteria and instruments used.

The heterogeneity of subjects under the classification of learning disabled has been well documented in the literature (e.g., Kavale & Nye, 1981; Torgesen & Dice, 1980). Keogh (1988) cites many studies that have questioned the validity of classification decisions. For example, Ysseldyke and his associates (Algozzine & Ysseldyke, 1983; Ysseldyke, Algozzine, Richey & Garden, 1982; Ysseldyke, Algozzine, Shin & McGue, 1982) compared school-district classified learning disabled and non-classified children and found that the two groups were virtually indistinguishable (See also Shepard, Smith & Vojir, 1983).

There is however, a certain amount of consensus regarding the major components used in the definitions of LD. According to a review conducted for the Ministry of Education (McIntyre, Keeton & Agard, 1980), the following major components were found to be common among all the definitions in use: a) a disparity between achievement, as measured by a test of academic proficiency, and intellectual ability, as measured by an intelligence test,⁸ b) an assumption of an underlying dysfunctional learning process, and c) the exclusion of children whose learning problems could be primarily ascribed to other conditions such as mental retardation, visual, hearing, or motor handicaps, emotional disturbance, or environmental disadvantage. Nevertheless, as will be demonstrated in this section, there are serious difficulties with this set of criteria in that they allow for many diagnostic errors.

Discrepancy Definition

The ability—achievement discrepancy is one of the most commonly used criteria for identifying learning disabilities (McIntyre *et al.*, 1980). Children who are low achievers in school, yet of average or above average intelligence, are prime candidates for being labelled as learning disabled. Although this has often provided a starting point for more refined diagnostic efforts, there are many possible problems in the use of the discrepancy criterion.

Significant discrepancy is operationally defined in an arbitrary fashion; there is no demonstrated and proper basis for determining what amount of discrepancy allegedly indicates presence of LD. Cone and Wilson (1981) have grouped the various approaches to quantifying discrepancy into four categories: grade level deviation, expectancy formula, standard score comparison, and regression analysis. Nevertheless, as Keogh (1988) points out, there are flaws in each of the formulae, particularly grade level (age and grade level are imprecise and misinterpreted) and expectancy formulae (small normal fluctuations in ability and achievement may be misinterpreted as expressing a real discrepancy). Furthermore, an identified discrepancy can be due to a number of factors such as opportunity for schooling and motivation. Salvia and Hunt (1984) note that "if the student has not had the opportunity to acquire the skills, we are measuring the extent of disadvantage and not the constructs. Given unmotivated performance we measure apathy, not the construct" (p. 34).

Many theorists have objected to the requirement of average or above average intelligence as a necessary condition for learning disabilities classification. All these issues lead one to basic questions as to the validity of the present LD construct. It is therefore concluded that discrepancy-based criteria in themselves are problematic and an

⁸There is often the additional provision that the latter score be at least "normal". The inadequacy of the use of IQ in the definition of LD is demonstrated in the work of Siegel (1988), who argues that since the IQ and achievement tests are both measuring abilities that are dependent on the specific problems of the reading disabled child (memory, expressive language), the use of a discrepancy criterion is questionable.

adequate theoretical basis has not been shown. There are, of course, alternatives to discrepancy-based definitions (See for example, Siegel & Heaven, 1986).

Underlying Dysfunctional Learning Process

Most definitions of LD presume a neurological basis or etiology. For example, Kirk and Gallagher's (1983) definition, "A learning disability is a *psychological or neurological impediment* (emphasis added) to spoken or written language or perceptual, cognitive or motor behavior" (p. 368). At the moment, however, there is no valid test (or battery of tests) that permits the operationalization of terms such as, "disorder in one or more basic psychological processes" or "due to central nervous system dysfunction". Although these terms may refer to some real phenomena, they are difficult to define and identify; the existence of the phenomena is, therefore, merely assumed or postulated by most researchers.

Further, the construct cannot be separated from instruments used to measure it. Coles reviewed validation studies of the ten most frequently used procedures for diagnosing learning disabilities, and analyzed commonly used instruments such as the Bender Gestalt (Visual-Motor) Test and the Frostig Developmental Test of Visual Perception. Speaking of the tests he noted: "The evidence against them is mounting" (p. 355). He concluded that most of these tests have dubious psychometric validity. (See also Davis & Shepard, 1983; Thurlow & Ysseldyke, 1979).

Exclusion Criteria

The usual exclusion criterion is stated as follows:

"the definition does not include children who have learning problems which are *primarily* (emphasis added) the result of impairment of vision or hearing; motor handicaps; mental retardation; primary emotional disturbance; or environmental, cultural, or economic disadvantage" (Ministry of Education, Ontario, 1978, p. 2).

In the following discussion, we will consider only that part of the exclusion criterion which refers to environmental, cultural, or economic disadvantage.

In established fields such as psychology, psychiatry, and medicine, definitions of syndromes/disorders/disabilities do not contain socio-economic exclusions because of the obvious absurdity involved. For example, autism as a disorder, given in DSM-III-R (American Psychiatric Association, 1987), is a diagnosis based on certain signs (e.g., marked lack of awareness of the existence or feelings of others, no or abnormal seeking of comfort at times of distress, no mode of communications, stereotyped body movements, etc., pp. 38-39). In DSM-III-R, the questionable directive is *not* given that the diagnosis is set aside if the patient's social disadvantage has been a primary cause of the signs. In the case of medical conditions, the situation is even more obvious. One might consider the case of rickets. Signs such as weak bones and deformed skeletal structures cannot be dismissed as indications of rickets if poverty has caused the conditions. In agreement with this position is that of McIntyre *et al.*, (1980) who state that "in the extreme, this type of definition would inevitably produce LD classes composed only of children of average intelligence from middle-class families who are not doing well in school. The emphasis on defining a clinical entity without proper attention to the defining process and the role of the definers (Mercer, 1971) is a serious and basic flaw in the major existing definitions" (p. 13-14). Vellutino (1978), also has raised serious theoretical questions about exclusionary criteria and identification of primary causal factors.

The Prospects of Establishing Adequate LD Constructs

The confounding of learning disabilities with other achievement-related conditions indicates that there are no agreed upon conceptual or operational criteria for classification. It is likely that this problem will exist as long as one continues to search for a single definition. Many theorists have argued that what is needed is a classification scheme and rationale based on a theoretical framework that helps distinguish between those who do and do not have learning disabilities and places LD in proper perspective with other learning problems. In her paper on the future of the LD field, Keogh (1986) emphasized the importance of "rationally defined subgroups", marker variables, and empirical subtyping. The recent work of various researchers (e.g., Petruskas & Rourke, 1979; Siegel & Heaven 1986; Stanovich, 1986) is contributing to the needed task of differentiating and validating relatively homogeneous subtypes subsumed under this label. Rourke (1985) has proceeded further with efforts to define subtypes of LD.

Summary

Based on the preceding discussion, it is concluded that the psychometric validity of a unitary LD construct has not been satisfactorily demonstrated. The possibility of demonstrating more limited, valid LD constructs has been mentioned, but the consequences for diverse groups are not known. It follows that educational assessments based on the tests mentioned may be questionable with respect to members of racial and cultural minority groups.

How Successful is one Significant Attempt to Design a Culturally-Fair Assessment Framework for Mental Abilities?

There have been various attempts to remedy the inadequacies of the use of standardized tests with minority students. Some efforts have been aimed at minimizing the cultural and verbal components of testing (Cattell, 1959). Nevertheless, these approaches have failed to solve the problems identified above. They have seemingly failed satisfactorily to address the problem that all tests measure learned behavior (for a review of culture-fair tests, see Cummins, 1984, pp. 191-197). According to many researchers (Anastasi 1961; Samuda, 1975 cited in Agard, 1982), it is impossible to develop a test which is free from cultural influence. Other efforts to translate present tests have also met with failure.⁹ It has been proposed that since LD, theoretically should show up in both first and second language, that the child be tested in both first and second languages. The practical difficulties of such double assessments, as well as availability of equivalent tests has led to a situation where almost no research has been based on this approach. Nevertheless, this research direction would seem to be very promising. Rather than attempting comprehensively to review the literature on "culture-fair" test batteries, the discussion here will focus on one of the best known and best elaborated attempts as an example—Jane Mercer's (1973) System of Multicultural Pluralistic Assessment.

The System of Multicultural Pluralistic Assessment (SOMPA) consists of a battery of measures that incorporates medical, social, and pluralistic information in the assessment of the cognitive, perceptual-motor, and adaptive behavior of Black, White and American-Hispanic children between the ages of 5-0 and 11-11 years. The purpose of

⁹For example, studies of Spanish translation of the WISC-R report difficulties in maintaining item difficulty and that words used in one language may be used rarely in another (Chandler & Plakos, 1969; Coyle, 1965). See Wilen & Sweeting (1986) for a discussion of Spanish versions of such tests as the Woodcock Language Proficiency Battery.

the SOMPA is to provide a comprehensive assessment that will allow educational decisions to be made in a racially or culturally non-discriminatory way.

Description of the SOMPA

The standardization sample of the SOMPA consisted of over 2,000 California public school children between the ages of 5-0 and 11-11. Three racial ethnic groups were represented: Black, White, and Hispanic children. No attempt was made to match the U.S. census distribution for California or for the U.S.

In the SOMPA, nine measures from three different models are used: the medical model, the social system model, and the pluralistic model. The medical model "focuses on identifying biological anomalies, disease processes, sensory or motor impairments, or other pathological conditions in the organism" (Mercer & Lewis, 1979, p. 42).

In the social system model "normal" behaviour is defined as behavior that meets the social system expectations for role performance. Behaviour that does not conform to the expectations of group members is considered to be deviant or abnormal. The measures used in the social system model are the WISC-R and the The Adaptive Behavior Inventory for Children (ABIC) (Mercer & Lewis, 1979). The WISC-R, when used with the standard, published norms is interpreted merely as the student's School Functioning Level (SFL) and is viewed as a measure of language proficiency and achievement—not as an intelligence measure. The SFL provides an indication of the amount of difficulty a particular student is likely to experience in the typical mainstream classroom. The Adaptive Behavior Inventory for Children (ABIC) consists of 242 multiple choice questions that are asked of the parent in order to obtain an indication of the students' emotional status.

Mercer and Lewis' (1979) pluralistic model uses multiple norms to estimate a child's "learning potential". The interpretation of the WISC-R is guided by the four socio-cultural scales in the normative framework; urban acculturation, socio-economic status, family structure, and family size. Through the use of regression scores, the standard WISC-R IQ scores are used to arrive at an Estimated Learning Potential (ELP). These ELP scores are viewed as estimates of a child's probable school success when judged in comparison to children from similar socio-cultural backgrounds. Based on this description, one can see that an ELP may be just a revised or disguised version of the classical "g".

Comment on the Adequacy of the SOMPA

Although the work of Mercer and her associates has been invaluable in pointing out the many alternative approaches that may be necessary in correctly assessing linguistically and culturally diverse students, there has been considerable controversy over the SOMPA. The various areas of concern include: a) its reliance on the WISC-R and the ABIC, thus its disguised measurement of some capacity very much like g; b) the use of separate norms and the questionable validity of ELP's as predictors of achievement; c) the lack of representativeness of the norm sample; and d) the inappropriateness for students whose dominant language is other than English. Each of these will be briefly discussed in turn.

Reliance on the WISC-R and the ABIC

As has been discussed at length in earlier sections, the use of the WISC-R with minority students is problematic. Plata (1982) has argued: "Ultimately, the procedure of adding points is demeaning to the culturally different student who has been subjected to a testing situation with an invalid instrument" (p. 6). Further, adding points arguably

yields a mere adjusted *g* measure. Additionally, Luther and Quarter (1990) feel that ELP is a "misnomer" because there is no "learning" on the SOMPA.

According to Mercer (1978-1979), the validity of the Adaptive Behavior Inventory for Children (ABIC) should be "judged by its ability to reflect accurately the extent to which the child is meeting the expectations of the members of the social systems covered in the scales...." (p. 109). Thus the implication that in order to judge the child's adaptive behavior, we need to know the expectations that the specific community has for a child. The ABIC, however, does not provide such information. Several studies (cited in Sattler, 1988, p. 383) have documented that low-income Mexican American children score lower than other groups on the ABIC.¹⁰

The Use of Separate Norms and the Questionable Validity of ELP's as Predictors of Achievement

The use of separate norms has been criticized by some (Clarizio, 1982) as being against sound psychometric practices. Another argument against separate norms has been that these children will ultimately need to manifest their academic potential in mainstreamed instructional programs geared toward middle-class Anglo norms. The ELP's, therefore, represent an inflated estimate of students' potential abilities and have little predictive validity under the conditions of actual educational environments. There are studies which indicate that ELP's are no more effective than WISC-R IQ's in predicting school-related performance, as measured by learning tasks (e.g., Wurtz, Sewell, & Manni, 1985).

The Lack of Representativeness of the Norm Sample

The ELP's yielded by SOMPA may be regionally specific. For example, Oakland (1979) has argued that the regression equations for students in Texas do not match those of the California sample. He reports average differences of 2 points for Blacks, 3 for Hispanics, and 10 points for Whites. Agard (1982), on the other hand, attempted to determine the appropriateness of the SOMPA, ELP norms with a Toronto population. He found that "in spite of different patterns of loadings, mean predicted WISC-R IQ's generated by Toronto regression equations were very similar to those of the California normative sample" (p. i).

The Inappropriateness for Students Whose Dominant Language is Other Than English

A limitation of the SOMPA is that it is inappropriate for students whose dominant language is not English. Mercer (1983) has admitted this limitation and points out that "Presently, there is no comprehensive system for assessing language disorders in students whose primary language is not English" (p. 52).

Summary

Although there are many problems with the SOMPA, it should be recognized as a substantive attempt to develop a comprehensive assessment process for sound educational decisions. Nevertheless, it does not overcome the cultural and linguistic biases embedded in the more conventional assessment procedures; it attempts post hoc corrections (albeit empirically derived) of the results of such procedures. Arguably, SOMPA is merely "g" in a slightly revised form (See also Wilen & Sweeting, 1986).

¹⁰For a positive recommendation however, regarding the ABIC with certain Hispanic students see Wilen & Sweeting, 1986.

Concluding Remarks

Implications of the Above Analysis for Special Education

The issues discussed above have direct implications for special educational theory and practice. It has been argued that the assessment tools currently used rest on dubious constructs and have questionable validity. This suggests that certain practices, such as the segregation of children labelled "learning disabled", lack a proper rationale. Further, the psychometric foundation of special programs is likewise subject to question. To put the same point a little differently one could say that special education intervention succeeds because of the sound general educational principles which constitute the core of these programs.

As a result of the many problematic areas in current special educational practices, new assessment/learning paradigms have been proposed accepting diversity as a basic assumption. Some of these will be briefly outlined in the following section, although no attempt will be made to evaluate, but merely to indicate briefly the nature and direction of these promising alternative paradigms.

Possible Alternative Assessment Models

Some research, based on Vygotsky's (1962) "zone of proximal development", has attempted to move away from "static" assessment procedures to assess what the child can learn. One example is Feuerstein's (1979) dynamic assessment procedures which employ a test-teach-test model to assess learning potential. In contrast to conventional assessment procedures which attempt to assess what the student knows, Feuerstein advocates inferring learning potential on the basis of direct assessment of learning in the assessment situation. In this paradigm, intelligence is defined as

the capacity of an individual to use *previously acquired experiences* to adjust to new situations. The two factors stressed in this definition are the capacity of the individual to be modified by learning and the ability of the individual to use whatever modification has occurred for future adjustments (1979, p. 76).

Similarly, Savron (1989) developed a Similarities Reasoning Test, which measures the level of generalizability in thinking. The tester collaborates with the child in developing "weak" cognitive structures through a restructuring process. It was found that children can attain "higher levels of thinking" when given this type of training. The measures were related to Gamlin's (in press) theory of similarity reasoning which is derived from the theory of Vygotsky (1962).

McClelland (1976) has proposed alternatives, previously discussed, such as testing for diverse competencies instead of using traditional intelligence tests. His advocacy of criterion sampling and tests designed to reflect changes in what the individual has learned is in accord with the argument of this paper.

Other efforts have been motivated by a desire to merge special and regular education (e.g., Rueda, 1989; Stainback & Stainback, 1984). Rueda (1989), has argued that the problems regarding minority students can only be addressed in the larger framework of the field of special education as a whole. He has, therefore, proposed basic structural changes and revision of most of the major fundamental operating assumptions used in special education. Rueda advocates moving away from the current classification systems toward more functional approaches.

Summary and Conclusions

The validity of the IQ and LD constructs has been discussed. First, with respect to the concept of 'g' (general intelligence) the discussion leads this author to oppose the

view, held by only a few psychologists today, that the validity of the IQ construct has been definitively demonstrated. Reasonable questions may still be raised about this well-known construct. Higher IQ scores are apparently found more among mainstream students, but from this writer's perspective, the existence of a genetic or biological basis has not been demonstrated. Hence, the use of the IQ construct has questionable implications, especially for socio-culturally diverse students.

Secondly, at the present time, and contrary to the opinion of many special educators, the present author holds, on grounds already given, that the usefulness of the LD diagnosis also has not been adequately demonstrated. Simple discrepancy-based definitions are inadequate and confound ability with poor achievement, which can be the result of several different causes. A more limited sort of reading disability construct, as found in Stanovich's (1986) or Rourke's (1985) work may have demonstrable validity, but measures have not been thoroughly tested across different populations and the implications for minority student assessment have not been extensively investigated and are unclear.

The SOMPA was designed because of an apparently genuine concern for culturally diverse students; however it makes substantial use of scales, the validity of which has been seriously questioned. Furthermore, as already mentioned by Mercer (1983), the SOMPA does not overcome certain *linguistic* barriers; so there is reason to suspect it may not overcome certain *cultural* barriers insofar as these may be language-related. Clearly there is a need for alternative models of assessment based on sound theories of learning, and taking into account the diversity of different linguistic and cultural environments.

References

- Agard, R. (1982). *Sociocultural intelligence and achievement scores in a black Toronto population: An analysis for programming decisions*. Unpublished Doctoral Dissertation: University of Toronto.
- American Psychological Association (1987). *Diagnostic and Statistical Manual of Mental Disorders* (Third Edition-Revised) Washington DC: APA.
- Anastasi, A. (1961). Psychological tests: Uses and Abuses. *Teachers College Record*, 62, 389-393.
- Algozzine, B. & Ysseldyke, J. (1983). Learning disabilities as a subset of school failure: The oversophistication of a concept. *Exceptional Children*, 50 (3), 242-246.
- Ashworth, M., Cummins, J., & Handscombe, J. (1988). *Report of the external review team on the Vancouver School Board's ESL programs*. Unpublished manuscript.
- Bernal, E.M., Jr. (1975). A response to "Educational uses of tests with disadvantaged subjects." *American Psychologist*, 30, 90-92.
- Block, N.J. & Dworkin, G. (1976). *IQ, heritability, and inequality*. In N.J. Block and G. Dworkin (Eds.), *The IQ controversy*. N.Y.: Pantheon Books.
- Brody, N. (1985). The validity of intelligence tests. In B.B. Wolman (Ed.) *Handbook of Intelligence: Theories, measurements, and applications*. New York: John Wiley & Sons, Inc.
- Cattell, R. (1959). *Handbook for the Culture-Fair Intelligence Test: A measure of "g"*. Champaign, Illinois: Institute for Personality and Ability Testing.
- Cattell, R. (1985). Intelligence and g: An imaginative treatment of unimaginative data. *The Behavioral and Brain Sciences*, 8 (2), 227-228.
- Chandler, J.T., & Plakos, J. (1969). Spanish-speaking pupils classified as educable mentally retarded. *Integrated Education*, 7 (6), 28-33.
- Chen, M., Tsuji, G., Yau, M., & Ziegler, S. (1989). *The every secondary student survey*. Research Section, Library Services Department, Toronto Board of Education.
- Clarizio, H.F. (1979). In defense of the IQ test. *School Psychology Digest*, 8 (1), 79-88.
- Clarizio, H.F. (1982). Intellectual assessment of Hispanic children. *Psychology in the Schools*, 19, 61-71.
- Coles, G.S. (1978). The learning disabilities test battery: Empirical and social issues. *Harvard Educational Review*, 48, 313-340.

- Cone, T.E., & Wilson, L.R. (1981). Quantifying a severe discrepancy: A critical analysis. *Learning Disabilities Quarterly*, 4 (4), 359-371.
- Coyle, F.A. (1965). Another alternate wording on the WISC. *Psychological Reports*, 16 (3), 1276.
- Cronbach, L. (1970). *Essentials of psychological testing*. N.Y. Harper
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, England: Multilingual Matters.
- Cummins, J. (1989). A theoretical framework for bilingual special education. *Exceptional Children*, 55.
- Davis, W.A. and Shepard, L.A. (1983). Specialists' use of tests and clinical judgment in the diagnosis of learning disability. *Learning Disability Quarterly*, 6 (2), 128-138.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments, and techniques*. Baltimore: University Park Press.
- Figueroa, R. (1982). *Diana revisited: A review of psychological testing policies and practices with Hispanic children in five selected school districts in California*. Davis, CA: University of California, Davis.
- Gamlin, P. (1989). Issues in dynamic assessment/instruction. *The International Journal of Dynamic Assessment and Instruction*, 1 (1), 13-25.
- Gamlin, P. (in press). Promoting the generalization of knowledge: A developmental intercultural approach to teaching metaphorical thinking. Cited in B. Savron (1989). Modifying similarity thinking: Implications for the generalization of knowledge. *The International Journal of Dynamic Assessment and Instruction*, 1 (1), 27-47.
- Garcia, A.B., & Zimmerman, B.J. (1972). The effect of examiner ethnicity and language on the performance of bilingual Mexican-American first graders. *Journal of School Psychology*, 3-11.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.
- Guilford, J.P. (1985). The structure of intellect model. In B.B. Wolman (Ed.) *Handbook of Intelligence: Theories, measurements, and applications*. New York: John Wiley & Sons, Inc.
- Horn, J.L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75 (3), 242-259.
- Horn, J.L., & Cattell, R.B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107-129.
- Jensen, A.R. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8 (2), 193-263.
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Johnson, D.F., & Mihal, W.J. (1973). Performance of blacks and whites in computerized versus manual testing environments. *American Psychologist*, 28, 694-699.
- Kamin, J. (1974). *The science and politics of IQ*. Potomac Md.: Lawrence Erlbaum Associates; distributed by Halsted Press.
- Kavale, K.A., & Nye, W.C. (1981). Identification criteria for learning disabilities: A survey of the research literature. *Learning Disability Quarterly*, 4, 383-388.
- Keogh, B.K. (1986). Future of the LD field: Research and practice, *Journal of Learning Disabilities*, 19, 455-460.
- Keogh, B.K. (1988). Learning disability diversity in search of order. In M.C. Wang and M.C. Reynolds, *Handbook of special education: Research and practice: Volume 2. Mildly handicapped conditions*. Oxford: Pergamon Press.
- Kirk, S.A., & Gallagher, J.J. (1983). *Educating exceptional children: Fourth edition*. Boston: Houghton Mifflin Company.
- Luther, M.G. and Quarter, J. (1990). *The genie in the lamp: Intelligence testing reconsidered*. North York: Captus Press.
- McClelland, D.C. (1976). Testing for competence rather than for "intelligence". In N.J. Block and G. Dworkin (Eds.), *The IQ controversy*. N.Y.: Pantheon Books.
- MacIntyre, R.B., Keeton, A., & Agard, R. (1980). *Identification of learning disabilities in Ontario: A validity study*. Toronto: Ministry of Education.

- Mercer, J. (1971). *The labelling process*. Paper presented at the Joseph P. Kennedy, Jr., Foundation International Symposium on Human Rights, Washington, D.C., 1971.
- Mercer, J. (1973). *Labelling the mentally retarded*. Los Angeles: University of California Press.
- Mercer, J. (1978-1979). Test "validity", "bias", and "fairness": An analysis from the perspective of the sociology of knowledge. *Interchange*, 9, 1, pp. 1-16.
- Mercer, J. (1983). Issues in the diagnosis of language disorders in students whose primary language is not English. *Topics in Language Disorders*, 3, 46-56.
- Mercer, J., & Lewis, J.F. (1979). *Systems of multi-cultural pluralistic assessment (SOMPA)*. New York: The Psychological Corporation.
- Ministry of Education, Ontario. (1978-1979). *Re: The education of students with learning disabilities: Statement to school boards*. December 21, 1978.
- Ministry of Education, Ontario. (1987). *The development of a policy on race, and ethnocultural equity*. Report of the Provincial Advisory Committee on Race Relations. September.
- Oakland, T.D. (1979). Research on the adaptive behavior inventory for children and the estimated learning potential. *School Psychology Digest*, 8, 63-70.
- Ortiz, A.A., & Yates, J.R. (1983). Incidence of exceptionality among Hispanics: Implications for manpower planning. *NAEB Journal*, 7, 41-54.
- Petrauskas, R.J., & Rourke, B.P. Identification of subtypes of retarded readers: A neuropsychological, multivariate approach. *Journal of Clinical Neuropsychology*, 1, 17-37.
- Plata, M. (1982). *Assessment, placement and programming of bilingual exceptional pupils: A practical approach*. Reston, Virginia: The Council for Exceptional Children.
- Ribeiro, J.L. (1980). Testing Portuguese immigrant children: Cultural patterns and group differences in response to the WISC-R. In D.P. Macedo (Ed.), *Issues in Portuguese Bilingual Education*. Cambridge, Mass.: National Assessment and Dissemination Center for Bilingual Education.
- Rourke, B.P. (1985). *Neuropsychology of learning disabilities*. New York: Guilford Press.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt Reinhart and Winston.
- Rueda, R. (1989). *Defining handicapping conditions within linguistic minority populations*. Paper presented at the Convention of the American Education Research Association: San Francisco.
- Salvia, J., & Hung, F.M. (1984). Measurement considerations in program evaluation. In B.K. Keogh (Ed.), *Advances in special education: Volume 4. Documenting program impact*. Greenwich CT: JAI Press.
- Samuda, R.J. (1975). From ethnocentrism to a multicultural perspective in educational testing. *Journal of Afro-American Issues*, 3 (1), 4-18.
- Samuda, R.J., & Crawford, D.H. (1980). *Testing, assessment, counseling and placement of ethnic minority students*. Toronto: Ministry of Education.
- Sattler, J. (1988). *Assessment of children: Third edition*. San Diego: Jerome M. Sattler, Publisher.
- Savron, B. (1989). Modifying similarity thinking: Implications for the generalization of knowledge. *The International Journal of Dynamic Assessment and Instruction*, 1 (1), 27-47.
- Shepard, L.A., Smith, M.L., & Vojir, C.P. (1983). Characteristics of pupils identified learning disabled. *American Educational Research Journal*, 20 (3), 309-331.
- Siegel, L.S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology*, 42 (2), 201-215.
- Siegel, L.S., & Heaven, R. (1986). Categorization of learning disabilities. *Neurological Aspects of Learning Disabilities*, 1, 95-121.
- Spearman, C. (1927). *The abilities of man*. New York: MacMillan.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21 (4), 360-406.
- Stanovich, K.E., Cunningham, A.E., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: issues of task comparability. *Journal of Experimental Child Psychology*, 38, 175-190.
- Stainback, W., & Stainback, S. (1984). A rationale for the merger of special and regular education. *Exceptional Children*, 51 (2), pp. 102-111.

- Sternberg, R.J. (1980). Sketch of a componential subtheory of human intelligence. *Behavioral and Brain Sciences*, 3, 573-614.
- Sternberg, R.J., & Gardner, M.K. (1982). A componential interpretation of the general factor in human intelligence. In H.J. Eysenck, *A model for intelligence*, Springer.
- Terman, L.M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and extension of the Binet-Simon Intelligence Scale*. New York: Houghton Mifflin Co.
- Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin.
- Terman, L.M., & Oden, M.H. (1959). *The gifted group at mid-life*. Stanford, California: Stanford University Press.
- Thurlow, M.L., & Ysseldyke, J.E. (1979). Current assessment and decision-making practices in models LD programs. *Learning Disabilities Quarterly*, 2 (4), 15-24.
- Torgensen, J.K., & Dice, C. (1980). Characteristics of research on learning disabilities. *Journal of Learning Disabilities*, 13 (9), 531-535.
- Vellutino F.R. (1978). Toward an understanding of dyslexia: Psychological factors in specific reading disability. In A.L. Benton and D. Pearl (Eds.). *Dyslexia: An appraisal of current knowledge*. New York: Oxford University Press.
- Vygotsky, L.S. (1962). *Thought and language*. Cambridge, Mass.: M.I.T. Press.
- Wilen, D.K., & Sweeting, C.M. (1986). Assessment of limited English proficient Hispanic students. *School Psychology Review*, 15 (1), 59-75.
- Wurtz, R.G., Sewell, T.E., & Manni, J.L. (1985). The relationship of estimated learning potential to performance on a learning task and achievement. *Psychology in the Schools*, 22, 293-302.
- Ysseldyke, J.E., Algozzine, B., Richey, L., & Garden, J. (1982). Declaring students eligible for learning disability services: Why bother with the date? *Learning Disability Quarterly*, 5 (1) 37-43.
- Ysseldyke, J.E., Algozzine, B., Shinn, M.R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education*, 16 (1), 73-85.