THE INFLUENCE OF NON-MEASUREMENT BIAS ON THE DIAGNOSIS OF ARTICULATION IMPAIRMENT

by

Susan L. Geiger

Dissertation submitted to the Faculty of

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Educational Research and Evaluation

APPROVED:

M.V. Lichtman, Chairperson

D.E. Hinkle

R.L. McKeen

B.R. Greenberg

J. McLaughlin

August, 1983

Blacksburg, Virginia

THE INFLUENCE OF NON-MEASUREMENT BIAS ON THE DIAGNOSIS OF ARTICULATION IMPAIRMENT

by

Susan L. Geiger

(ABSTRACT)

Test and measurement bias in special education diagnosis has been well documented. Boys, linguistic and ethnic minorities, and children with behavior problems are among those overrepresented in several handicap categories, including speech impairment. Recent evidence indicates that variables associated with test interpretation or diagnostician background (non-measurement factors) may be better predictors of diagnostic bias. This study investigated the ability of non-measurement factors to predict the diangostic decisions made for 345 speech impaired children enrolled in Head Start. Computer Assisted Telephone Interviews and Case File Review were used to collect data about (1) child characteristics, (2) diagnostic criteria, and (3) diagnostician background. Hierarchical regression procedures were used to test the predictive power of these three blocks of variables and of specific variables within each block. The diagnostician's rating of articulation severity was the criterion variable. The test score was the best predictor of articulation severity rating; non-measurement factors were not found to be effective predictors of the articulation component of speech diagnosis. There was some indication, however, that non-measurement factors may be related to the language component of speech assessment. Further investigation of the language severity rating and of other non-measurement factors was suggested.

ACKNOWLEDGEMENTS

Undertaking a study and writing a doctoral dissertation are tasks which cannot be accomplished without the guidance, encouragement and cooperation of many teachers, administrators, and friends. I want to express my deepest appreciation to all those who helped me to complete this study.

I am particularly grateful to Dr. Marilyn Lichtman who served as my Student Committee Chairperson and to Dr. Bonita Greenberg who served as a member of my Committee. Their many hours of thoughtful and patient guidance have been invaluable.

I sincerely appreciate both the interest and efforts of ohter committee members, Dr. Ron McKeen and Dr. John McLaughlin. A very special thanks is extended to the Head Start administrators and speech-language diagnosticians throughout the United States who made this research possible. Their sincere interest and generous offering of time have been the backbone of this study.

Finally, I want to thank Rachel Kaye, without whose love, encouragement and personal sacrifices, this study would never have been undertaken.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EXHIBITS

CHAPTER I

BACKGROUND

Numerous court cases and research studies have emphasized the potential for discrimination that exists anytime a child is evaluated for special education. This discrimination or bias is manifested in the form of differential treatment of some minority groups. There is overwhelming evidence, for example, that children from minority ethnic groups are much more likely than their white, non-Hispanic peers to be inappropriately labeled mentally retarded or learning disabled because of their ethnicity and/or their primary language or dialect (Diana v. State Board of Education, 1973; Larry P. v. Riles, 1974; Mercer, 1970). Several studies have also suggested that boys, particularly those with aggressive or hyperactive behavior are more likely to be diagnosed as handicapped (Tomlinson, Acker, Canter & Lindborg, 1977; Walters, 1978; Weatherly & Lipsky, 1977). Age has also been identified as a potential source of bias as far more younger children are referred for special education (Bernard & Clarizio, 1981; Craig & McEachron, 1978).

Nondiscriminatory testing and diagnosis should, theoretically, result in equal treatment of students who have the same abilities on the criterion being measured. When differential treatment was observed, initial efforts to remove this discrimination were directed toward the test instruments. The instruments were felt to be the cause of what, for the purposes of this study, will be termed "measurement bias"; that is, errors in test construction or scoring were thought to result in differential treatment or discrimination. A variety of procedures were adopted by federal, state and local agencies as

safeguards against discriminatory testing. Tests which favor one group over another were revised or replaced and the use of multiple assessment sources was emphasized (BEH, 1974; Evard & Sabers, 1979; Flaugher, 1978). The effectiveness of these well-intentioned efforts, however, is highly questionable. Podilla and Garza (1975) note that "efforts to develop culture-free tests were doomed to failure from the beginning . . . no test can be experience free." Recent studies support this negative view as many groups including members of ethnic minorities, nonstandard English speakers, boys and children with aggressive behavior are still overrepresented in special education classes (Brady, Manni & Winikur, 1983; Phipps, 1982; Ysseldyke & Algozzine, 1982).

Many educators have begun to search for other sources of bias; that is, for factors that influence the test interpretation other than the test score or the test itself. The term "non-measurement bias" will be used to refer to this type of differential treatment. Reschly (1981), for example, argues that it is not the tests themselves but test use that leads to discriminatory assessment. He found that academic or behavioral problems were the "single most important determinants" in the diagnosis of handicapped children. A lack of standard criteria for diagnosing and classifying special education students allows the examiner's internal biases to affect the diagnostic outcome. More recent reforms have thus been directed toward specifying test interpretation procedures and diagnostic criteria.

Physical, health, visual, and hearing handicaps are measurable in empirical terms and are less subject to non-measurement bias. The learning disabilities (LD) category, however, is less easily defined and has been subject

to much scrutiny and redefinition (Beatty, 1977; Murai, 1982; Reynolds, 1980). Criteria for diagnosing mentally retarded children were the first to be examined and redefined (Dunn, 1968; Larry P. v. Riles, 1972; RISAC v. Board of Regents, 1973). Criteria for diagnosing emotionally disturbed (ED) students have only recently been criticized as a source of non-measurement bias (Bischoff & Shakel, 1980; Mack, 1980).

Statement of the Problem

While speech impairment (SI) is one of the most common disabilities of handicapped children throughout the nation (U.S. Department of Education, 1980), it is also one of the least consistently and operationally defined handicapping conditions. (See, for example, Administration for Children, Youth and Families, 1975). Efforts to investigate or to reduce non-measurement bias, such as those directed toward LD, EMR and BD, have generally ignored the SI diagnosis. Yet, there are no universal objective criteria for the measurement of the severity of a speech disorder. Although some local jurisdictions have established formal criteria, the diagnosis of speech and language handicaps is often considered to be a matter of clinical judgment. Emerick and Hatten (1974) have summarized this viewpoint: "The mature (speech-language) diagnostician does not look at objective scores . . . but rather aspects of an individual's communication ability. We diagnose communicators, not communication" (p. 20).

This lack of standard criteria is a particularly crucial issue in the assessment of preschool children. All children in this age group may be expected to manifest "errors" in speech and language. The diagnostician must

decide which of these children are exhibiting normal development and which represent handicapped individuals. Normative data which may have helped to settle this issue are contradictory and confusing. A number of classical studies (Poole, 1934; Templin, 1957; Wellman, Mengert, & Bradbury, 1931) were conducted with different populations and criterion measures. More recent studies indicate that earlier normative research is outdated as consistently earlier age levels have been noted for correct sound and feature production (Arlt & Goodban, 1976; Prather, Hedrick & Kern, 1975; & Sander, 1972). These later studies, however, have also produced conflicting results.

One factor which may influence not only the diagnostician's use of criteria, but the entire assessment process as well, is the training and experience of these professionals. Recent research indicates that differences in the speech diagnostician's background may result in very different diagnostic decisions (Schissel & Flourney, 1978; Wolfram, 1979).

It is not surprising, given this absence of consistent criteria and differences in diagnostician background, that charges of discriminatory classification have been leveled specifically at the speech-language diagnostic process. Charges of discrimination and overrepresentation in speech and language assessment have been made for younger children, those from low income families, ethnic and linguistic minorities and males (Craig & McEachron, 1978; Silverman & Opens, 1980; Vaughn-Cooke, 1982). There have, however been no systematic attempts to investigate non-measurement sources of bias; that is, to determine how the criteria used nationally by speech language diagnosticians or the training and experience of these diagnosticians affects their diagnosis of minorities or other subgroups.

Purpose of the Study

The major purposes of the study are (1) to investigate two non-measurement sources of bias - - diagnostic criteria and diagnostician background and (2) to compare the relative importance of these non-measurement sources and the child's demographic and behavioral characteristics in predicting the SI diagnosis.

This study will focus on the diagnosis of only one type of speech disorder, i.e., articulation impairment. While speech and language impairments encompass a variety of disorders including problems with articulation, language, voice, fluency, and pragmatics, disorders of articulation and language account for the vast majority of impairments. The articulation diagnosis was specifically chosen for this study because (as discussed in Chapter 2) it is measured in a more consistent manner than other speech-language problems. If variations in criteria are found in this area of diagnosis, then it is assumed that they would also be prevalent in more complex areas of testing, such as language. Articulation was also the focus of this study because it has been shown to be highly related to language diagnosis (Schmauch, Panagos, & Klich, 1978; Shriner, Holloway, & Daniloff, 1969).

The specific purposes of the study are (1) to describe the variations in criteria used nationally to rate the severity of preschool articulation impairment; (2) to describe variations in the training and experience of speech-language staff who diagnose those children, and (3) to determine the contribution of variance in criteria, diagnostician's background, and children's demographic and behavioral characteristics to the diagnosis of articulation severity.

Research Questions

Major Questions:

1. What criteria are used to rate the severity of articulation impairment in preschool children?
2. What is the training and experience of preschool speech-language diagnosticians?
3. How much variance in the articulation severity rating is predicted by the following variables:

   a. diagnostic criteria

      - percentage of articulation errors
      - articulation test used
      - additional criteria listed by the diagnostician (includes behavioral problems)

   b. diagnostician background

      - years of experience
      - highest speech-language degree earned
      - type of certification
      - recency of training
      - awareness of Head Start speech-language diagnostic criteria

   c. child characteristics

      - age
      - sex
      - ethnicity
      - primary language/dialect
      - diagnostician's appraisal of child's language disorder

- accompanying handicapping conditions

4. Are there any significant interactions among the diagnostician, diagnostic criteria, or the child variables?

Hypotheses

1. A wide range of objective and subjective criteria are used to rate the severity of articulation impairment in preschool children.
2. Speech-language diagnosticians have a wide range of training and experience.

3a. Diagnostic criteria will be better predictors of articulation severity ratings than will child characteristics or the diagnostician's background.

3b. The following single variables will be significant predictors of variance in the articulation severity rating:

- percentage of articulation errors
- diagnostician's educational degree
- child's sex
- child's ethnicity
- behavioral problems
- diagnostician's appraisal of child's language disorder

4. The following interactions will be significant predictors of variance in the articulation severity rating:

- sex and behavioral problems
- ethnicity and behavioral problems
- articulation error score and age
- articulation error score and language severity rating

Operational Definitions

For the purposes of this investigation, the following definitions will be used:

1. Speech impaired child: any handicapped Head Start child whose primary disability has been officially listed as speech impairment. This includes objective scores from diagnostic articulation instruments and subjective factors such as speech of family members and the child's behavior.

2. Articulation criteria: all factors listed by speech-language diagnosticians as factors influencing their diagnosis of articulation impairment. This includes objective scores (AES) from diagnostic articulation instruments and subjective factors such as speech of family members.

3. Articulation Error Scores (AES) will be calculated from the articulation test administered closest and previous to the date of the child's initial diagnosis of speech impairment. Bernthal and Bankson (1981) list two methods by which articulation norms are typically used to diagnose articulation problems: (1) the number of correct responses is compared to the norms for the mean number of correct responses or (2) individual responses for individual sounds are compared with the norms for those sounds. Error scores will be calculated using both methods, thus yielding two separate measures of this predictor variable:

   AES1 = the percentage of phonemes in error

   AES2 = the total percentage of errors, counting all positions in a word

   (A phoneme is technically defined as a distinctive sound family which, in combination with other phonemes, make up the spoken words of the

language (Carrell, 1968). Articulation tests typically measure the child's ability to produce a list of phonemes in 2 or 3 positions within a word (initial, medial and final). Thus a test of 20 phonemes may result in as many as 60 items or potential errors.)

4. Articulation Severity Rating: the rating of mild, mild to moderate, moderate, moderate to severe, or severe assigned by the speech-language diagnostician at the initial diagnosis.
5. Age: the child's age in months at the date of the initial evaluation.
6. Ethnicity: the child's ethnicity as determined by Head Start administrative records. The federal ethnicity categories were used:

1. White, not Hispanic	4. Asian or Pacific Islander
2. Black, not Hispanic	5. Native American or Alaskan Eskimo
3. Hispanic	

Minority ethnic groups: any group other than #1 above.

7. Language/dialect: the child's primary language or dialect as determined by the speech-language diagnostician.

Minority language/dialect: any primary language or dialect other than Standard English; this includes Spanish, Vietnamese, Black English and Appalachian dialect.

8. Behavior problems: overly aggressive, hyperactive or severely disruptive behavior as indicated at the time of the initial evaluation by Head Start records and/or the speech-language diagnostician.
9. Language Severity Rating (LSR): the rating of normal to severe assigned by the speech-language diagnostician at the time of the initial evaluation.

Significance of the Study

This is the first intensive investigation of non-measurement bias in speech-language diagnosis at a national level. The randomly selected sample includes Head Start programs in all regions of the United States and from programs of all sizes. The results may be generalized to the diagnosis of articulation impaired children in Head Start programs throughout the United States. Additionally, the majority of diagnosticians who served as study respondents are employed by other agencies such as public school systems or community health agencies. Study results will, therefore, provide some indication of procedures used in speech-language diagnosis outside of Head Start.

Study Strengths and Limitations

Some desired controls were not feasible as this study utilized secondary data analysis. This researcher, however, was thoroughly familiar with the data set. She was the Principal Investigator for the original study, responsible for the development of instrumentation and the supervision of data collection. She is thoroughly familiar with the strengths and weaknesses of the data. The major loss of control was in the measurement of the articulation severity rating. It cannot be determined what scale was used as diagnosticians were asked only to rate the severity in their own words. It is clear, however, that diagnosticians treated the severity rating as a continuous scale, ranging from mild to severe. Some respondents used only three descriptive points, i.e., mild, moderate and severe. Other respondents included two additional points, i.e., mild to moderate and moderate to severe.

The accuracy of the data was greatly enhanced by the use of Computer Assisted Telephone Interviewing (CATI) for all data collection other than document review. CATI requires that the interviewer immediately record a response in the computer before presenting another question. This procedure allows for data editing during the interview process. Out of range responses, for example, cannot be entered; the question must be repeated and the response re-entered. Additionally, no errors are made in transcribing responses from a written record to a computer form.

The population used in this study is a specialized group. Head Start actively recruits preschool children from lower socioeconomic (SES) and minority backgrounds. Generalization of study results to other age, ethnic or SES groups may not be appropriate. The sample, however, was drawn from a national population so that the results may be generalized to all articulation impaired children in Head Start.

CHAPTER II

REVIEW OF THE LITERATURE

This review of the literature has been divided into two parts. The first part reviews general studies in the diagnosis of all types of handicapped or special education students. This section is included because so little attention has been paid specifically to the diagnosis of speech impairment. Part I provides a structural framework for viewing discriminatory diagnostic procedures related to measurement and non-measurement issues in identifying special education students. Part II describes the relatively few studies of bias in the diagnosis of speech impaired children using the same framework of measurement and non-measurement bias.

Part I: Special Education Diagnosis

One outgrowth of recent efforts to protect the rights of minority groups has been a close inspection of the special education diagnostic process. Charges of overrepresentation of minority groups in certain handicapped categories have been reviewed in numerous court cases and research studies. More recent investigations have focused on additional targets of discrimination, including boys and children with aggressive behavior. The earliest criticism concerned the biases inherent in the test instruments. Part I begins with a documentation of the discriminatory effects of using biased tests in the assessment of handicapped children. The failure of efforts to eliminate discriminatory assessment through a reduction of test bias is discussed.

The search for sources of diagnostic bias has recently turned toward non-

instrument factors such as the interpretation of test data and the training of educational diagnosticians. Biases associated with these non-measurement factors are explored at the end of Part I.

Discrimination in special education diagnosis: Test bias

The following investigation of discriminatory testing of handicapped students has been divided into two distinct historical periods. During the first period (pre-1975) the case for nondiscriminatory assessment was mounting. Research revealed the devastating effects of test bias. In the courts, a series of precedents requiring nondiscriminatory test practices was being established. In 1975-77 the Education for All Handicapped Act (P.L. 94-142) and the accompanying Protection in Evaluation Procedures firmly established the responsibilities of the schools to provide unbiased testing. Thus, in 1975 a new era characterized by federal intervention in the handicapped assessment process began. Many test-related strategies including test revisions, changes in norming procedures and the development of new instruments were adopted in an attempt to eliminate assessment bias. Unfortunately, evidence of failure to control discrimination in the diagnosis of handicapped children still exists today.

Discriminatory test practices before 1975. According to Lloyd Dunn, a former president of the Council for Exceptional Children, about 60% to 80% of the students in the nation's mentally retarded classes in 1968 were "children from low status backgrounds - including Afro-americans, American Indians, Mexicans and Puerto Ricans; those from non-standard English speaking...(and) other non-middle class environments" (Dunn, 1968, pp. 5-6).

There is much empirical evidence to support Dunn's estimates of minority group overrepresentation in special education classes. In California, for example, Mexican American children comprised only 13% of the 1967 school population, but accounted for almost 30% of special education students in the state (Mercer, 1970). There is also abundant evidence that black children were substantially overrepresented in educable mentally retarded (EMR) classes in the same state. From 1968 to 1974, when the black student population was estimated to be 9-10%, black enrollment in EMR classes remained stable at an overwhelming 24.5 to 27% (Larry P. v. Riles, 1974). A 1971 survey of 11 Missouri school districts disclosed that learning disability (LD) programs which focus on remedial help, were predominately (97%) filled with white children from high income families (Segregation of poor, 1971). EMR programs, which focus on compensatory assistance, had disproportionate numbers (34%) of black children. In Ann Arbor, Michigan, where blacks comprised 11% of total school enrollment, they accounted for 40% of the city's EMR enrollment. Similar figures could be cited for state and local education agencies throughout the United States during this period as "disproportionate numbers of blacks and other minority students in EMR classes seem(ed) to be the rule rather than the exception" (Segregation of poor, 1971, p. 1212).

Other studies conducted during this time period began to uncover a possible sex bias in the special education diagnostic process. Blom (1971) noted the heavy overrepresentation of males in LD classrooms, based primarily on their low reading scores. Although this overrepresentation has been explained on biological grounds, Blom strongly argues that the biological

factor is disputed by data from other countries which suggests no such male-female differences in LD enrollment. Jackson and Lahaderne (1971) found that teachers and educational diagnosticians hold different attitudes toward children as a function of the child's sex. These attitudes affect the type and level of contact during the diagnostic process and, ultimately, may affect the diagnosis. Palardy (1969) studied the same differential attitudes and suggested that sex biases also affected the prognosis ascribed to the child. McLaughlin, Miller and Chansky (1970) also studied prognoses of EMR students made by special educators and found sex differences in judgments of lower class, but not of upper or middle class retardates.

Initially, the tests used in the special education diagnosis were blamed for this inequality. Dunn's criticisms of discriminatory special education testing have proven to be prophetic. In his now classic article (1968), Dunn criticized the profound effects of special education labels, requesting greater care in a diagnostic process which so heavily influenced the lives of children. Specifically, he recommended the development of nondiscriminatory tests and the use of multiple test sources with more time allotted to the diagnostic process. Dunn outlined many additional concerns in placement and other procedures, calling for "a revolution in much of special education" (p. 11).

Other members of the legal and educational profession began to support Dunn's contentions. For example, Henry Dyer, former vice-president of the Educational Testing Service, criticized the use of IQ and grade equivalent sources which were frequently biased because black schools were not included in normative samples (Segregation of poor, 1971). Schools were criticized for using intelligence tests as the sole - or at least the predominant - criterion in

labeling students mentally retarded and in placing them in special education classes (Weintraub, Abeson, & Braddock, 1971). California studies conducted to determine the racial and ethnic composition of EMR classes prior to 1975 found that measures of adaptability (while officially required for EMR diagnosis) were generally omitted. Assessment procedures were "unidimensional...focus(ing) on the narrow band of behavior sampled in the psychometric stuation (by a single IQ test)" (Mercer, 1975, p. 143).

Strategies to eliminate test bias. While the movement was initially slow to gain momentum there were efforts by the courts, Congress and test developers to eliminate discriminatory assessment through changes in the tests themselves. Once begun, however, the nondiscriminatory evaluation issue mushroomed through heavy litigation, national legislation and a proliferation of new tests.

A 1971 legal review (Segregation of poor) of the continued segregation of minority children in EMR classes, despite the protests of educators, such as Dunn and Dyer, suggests that legal action was needed to change educational policies. Several reasons are offered for the delay in bringing overrepresentation cases to court:

(1) An inordinate amount of time is required for a lawyer to pursue the structure of psychometric measures in order to determine whether the client has a case of test bias;

(2) Large expenditures of funds required for major court case are usually beyond the financial resources of parents of minority children;

(3) The courts, despite their active role in school integration, had

adopted a hands-off approach toward the student-teacher relationship, maintaining the in loco parentis powers of the schools; and

(4) Many view educational tests as "scientific", relying on their almost sacred validity.

Although legal relief from test bias in special educational diagnosis began slowly, the movement soon gained momentum, finally spawning a powerful piece of federal legislation - P.L. 94-142: The Education for all Handicapped Act. The major precedents for this law, which includes strict regulations against discriminatory testing, are reviewed in the following discussion.

The first court decision concerning the use of standardized tests and the educational placement of minority children was the landmark case of Hobson v. Hansen (1967). The "tracking system" used by the Washington, D.C. public schools was abolished by Judge Skelly Wright because (1) a disproportionate number of black and low income children were assigned to lower tracks, (2) placement tended to be permanent with little interaction between levels, and (3) track assignments were made on the basis of biased aptitude and achievement tests. This case thus questioned the legality of school classification practices and led the way for future litigation concerning educational testing bias.

A few years later a class action suit was filed in California on behalf of six black school children, alleging that they had been inappropriately classified as educable mentally retarded (Larry P. v. Riles, 1972). The complaint characterized the children as "victims of a testing procedure which fails to

recognize their unfamiliarity with the white middle class cultural background..." (p. 1306). Judge Peckham, after hearing statistics of heavy overrepresentation of blacks in EMR classes and testimony concerning the norming procedures for IQ tests, issued an injunction barring the use of IQ tests as the sole criterion for EMR placement. Following a later appeal (Larry P. v. Riles, 1974) the use of standardized IQ tests in the state of California was banned entirely.

The publicity surrounding the Larry P. case generated a series of legal battles for other minority groups which further strengthened the court's position against the use of potentially biased standardized tests. In Arizona a case was brought by the Guadalupe Organization, Inc. regarding the disproportionate number of bilingual children in EMR classes (Guadalupe v. Tempe, 1972). The school district agreed to re-evaluate all bilingual EMR children in their primary language. In Diana v. State Board of Education (1970, 1973) the plaintiffs were nine Mexican - American children who represented all minority preschool bilingual children. The schools were charged with using highly verbal, culturally biased and improperly standardized (norming population did not include rural Mexican -American children) IQ tests. The court ordered that all California schools (1) test all children whose primary home language is other than English in both their primary language and in English and (2) form a new or revised IQ test for use with Mexican American children which allows them to be compared with their peers, not with the total population. In a similar case (RISAC v. Board of Regents, 1973) which was settled out of court, the Rhode Island Board of Regents agreed to adopt a slightly different approach; they would adjust the cut off scores of IQ

tests to adjust for socio-cultural bias. In Arivisu et al. v. Waco (1973) a Texas school district was ordered to ensure that, when a standardized test was used to classify a student as mentally retarded, the test contained no bias against any cultural, racial or ethnic group.

As court battles were being waged for changes in discriminatory test practices, hearings were being conducted in Congress and throughout the nation concerning the rights of handicapped students. In 1975 Public Law 93-380, the Education Amendments of 1974, went into effect. This legislation clearly ordered the states to establish procedures to ensure that tests were not racially or culturally discriminatory. The Bureau of Education for the Handicapped (BEH) interpreted these procedures more specifically, recommending the following guidelines:

> A procedure also should be included in terms of a move toward the development of diagnostic-prescriptive techniques to be utilized when for reasons of language differences or deficiencies, non-adaptive behavior, or extreme cultural differences a child cannot be evaluated by the instrumentation evaluated by the instrumentation of tests. Such procedures should insure that no assessment will be attempted when a child is unable to respond to the tasks or behavior required by a test because of linguistic or cultural differences unless culturally and linguistically appropriate measures are administered by qualified persons. In those cases in which appropriate measures and/or qualified persons are not available, diagnostic-prescriptive educational programs should be used until the child has acquired sufficient familiarity with the language and culture of the school for more formal assessment. These evaluation procedures should also assure that persons interpreting assessment for information and making educational decisions are qualified to administer the various measures and qualified to take cultural differences into account in interpreting the meaning of multiple sets of data from both the home and the school. (BEH, 1974, p. 29)

Although, BEH did not recommend any specific tests, there is a broad statement concerning the selection of evaluation measures:

> The various evaluation materials and procedures used for purposes of classification and placement of handicapped children should meet a test of reasonableness in the eyes of competent professional persons and informed laymen; and such procedures should be administered by qualified persons under conditions which are conducive to the best performance of the child (BEH, 1974, p. 26).

This new law did not begin to satisfy all the issues raised during four years of Congressional hearings. It was merely a stopgap measure. In April, 1975 Congress began a series of legislative hearings to extend and amend P.L. 93-380. By this time over half of the states had either been through or were going through litigation related to the rights of handicapped students (Sabatino, 1979). In August of 1975 the Office of Civil Rights (OCR) of the U.S. Department of Health, Education, and Welfare combined the provisions of P.L. 93-380 with the intent of other civil rights legislation (Title VI of the Civil Rights Act). The OCR summarized these acts in a fairly detailed directive to the nation's chief state school offices which attempted to further curb the use of biased tests and evaluation procedures (Abeson & Ballard, 1976).

Finally on November 29, 1975 P.L. 94-142, the Education for All Handicapped Children Act, "representing a major new commitment by the federal government..., the most important education legislation enacted since the Elementary and Secondary Education Act of 1965" was passed (Congressional Quarterly Service, 1976, p. 651). The majority of the P.L. 94-142 provisions were scheduled to be in effect by October 1, 1977.

One important provision of this landmark legislation was the 1977 "Protection in Evaluation Procedures" clause. Included in this section were regulations underscoring Congress' philosophical position that:

- professional and legal requirements are necessary to ensure that children are not misclassified due to inappropriate selection, administration, or intrepretation of assessment instruments;
- tests and other evaluation materials include those tailored to assess specific areas of educational needs and not merely those designed to provide a single IQ score; and
- information from sources other than ability or achievement tests, including physical, sociocultural and adaptive behavior assessment, be considered in the assessment.

In conjunction with the passage of the Protection in Evaluation Procedures, many new and revised tests and testing procedures were developed to meet the new federal requirements. A number of strategies were adopted in an attempt to ensure that tests were racially, culturally and linguistically fair to all students. The Federal Government, through its now renamed office of Special Education Programs (formerly the Bureau of Education for the Handicapped) was a leader in efforts to develop effective strategies. A successful planning conference, "With Bias Toward None" resulted in a document of non-biased assessment activities which was later published by Oakland and Matuszek (1977).

Oakland, DeLuna, and Morgan (1977), Olmedo (1981), and Samuda (1975)

have pointed to the area of assessment of linguistic minorities as a particularly difficult one. This has proven to be a complex area as the primary language or dialect must be determined before accurate educational assessment may begin. Olmedo (1981) notes that bilingual students differ widely in terms of their receptive and expressive language dominance. Some speak one language exclusively, while understanding two languages; others use two systems equally well. The combination of possibilities is staggering. The basic goal should be "to use language in a manner that maximizes the opportunity of examinees to understand the testing situation and to respond according to their best abilities in either or both languages" (Oakland & Matuszek, 1977, p. 98). Despite the importance of primary language determination, however, available language assessment measures are disappointing. For example, in a review of 27 language dominance measures, Oakland, DeLuna and Morgan (1977) found only four which provided both validity and reliability information.

For speakers who must be tested in a language other than English, a variety of tests have been made available. Samuda (1975) listed nearly 100 tests in Spanish alone. These include Spanish translations of conventional tests (Stanford-Binet, WISC) and tests developed in terms of parallel English-Spanish forms (Test of Auditory Comprehension). Olmedo (1981) notes the magnitude and complexity of problems associated with these instruments. Tests may be written in formal Spanish, for example, which is inappropriate for dialect speakers.

When appropriate tests are unavailable in the speaker's primary language, as is often the case with Vietnamese and Chinese students for

example, English tests are typically translated during the test session. These tests, however, have not resulted in technically equivalent forms and the test contents remain culture-bound (Samuda, 1975).

Another approach to eliminating test bias has been to develop criteria for assessing discriminatory items and to revise test administration for those items only (Flaugher, 1978). Revisions have taken the form of different cut-off scores, elimination of discriminatory items, acceptance of alternate responses or other revisions of those items (Evard & Sabers, 1979; Vaughn-Cooke, 1982). Some basic problems, however, have been noted with this approach. Criteria for determining item bias have been difficult to establish. Sandoval and Mille (1980) found subjective interjudge agreement among minority persons concerning which items were biased to be very low. On the other hand, objective criteria generally result in little or no evidence of item bias for commonly used tests (Cole, 1981, Reschly, 1981; Sandoval, 1979).

Reschly (1981) notes that an additional "popular nonsolution to test bias" is to ban the use of discriminatory standardized tests altogether. There has been a ban on IQ tests in California since early 1975 (Larry P. v. Riles, 1974), yet there was no appreciable change in minority representation in special education classes until quotas were set in 1979 (Reschly, 1981). Reschly further argues that IQ tests may be useful in preventing inappropriate classification.

A fifth strategy in eliminating discriminatory assessment has been the development of specialized culture-fair tests (Jirsa, 1983). Perhaps the most ambitious and well-known culture-fair test is the system of Multicultural Pluralistic Assessment (SOMPA) (Mercer, 1979). The SOMPA is an entire

measurement system which utilizes conventional instruments with a complex set of pluralistic norms. It is based on Mercer's research (1975) which found a low and racially different correlation for classroom grades vs. WISC scores. The correlation was .20 for blacks and .46 for whites. An even more fundamental problem of standardized tests, according to Mercer, was their differential validity; that is, blacks scored more errors than whites regardless of their aptitude. As Mercer (1979) states, "the greater the sociocultural distance between the individual and the dominant core culture, the lower his or her score will be" (p.14). Mercer assumes that there are different normal distributions of behavior for each type of sociocultural life style. Thus she has developed the "ELP (Estimated Learning Potential) (which) has multiple normal distributions, one for each pattern of Sociocultural Scale scores" (Mercer, 1979, p. 137).

In response to legislation and research findings, measurement specialists have attempted to develop a number of alternate approaches to the nonbiased testing of linguistic, ethnic and cultural minorities. Although sex was also mentioned as a possible source of assessment bias, it was not a specific focus of the new federal handicapped legislation; therefore, the elimination of sex bias has not been a priority. The nondiscriminatory strategies discussed thus far have all been related to changes in the test instrument. A variety of different approaches including language dominance measures, test translations, parallel measures, revised items, test banning, and new tests were discussed. The following section discusses the effectiveness of these test-related strategies.

Effects of attempts to eliminate assessment bias. Some investigators have concluded that the Protection in Evaluation Procedures and similar regulations have produced positive results. The focus of these evaluations has been the representation of cultural and ethnic minorities in special education. Jones (1981) reviewed enrollment data from five states with culturally diverse populations. After finding a significant reduction in the number of children classified as handicapped from December 1, 1978 to December 2, 1978 be concluded that the "new standards may be accomplishing their intended purpose" (p.116).

Matuszek and Oakland (1979) used fictitious case records to investigate recommendations for handicapped placement by teachers and psychologists. They found that teachers and psychologists relied on test data more often than on ethnic factors. In a similar study using simulated files of children with the same low test scores but different backgrounds, psychologists selected only 60% of the low socioeconomic blacks and 100% of the high income children of either race as eligible (Frame, Clarizio, Porter & Vinsonhaler, 1982). This reverse discrimination suggests that factors other than test bias may be operating in discriminatory assessment.

Other researches have investigated referral - the initial step in the diagnostic process. Giesbrecht and Routh (1979) found another form of reverse discrimination. The elementary teachers in their sample tended to expect more favorable educational prognosis and less need for special education help for black students than for whites. Other more recent research (Low & Clement, 1982; Tobias, Cole, Zibrin, & Bodlakova, 1982) has uncovered no evidence of differences in referral rates related directly to the student's

ethnic background. Teachers, however, tended to refer children from their own ethnic backgrounds significantly less often than they referred children from different backgrounds. In a similar study (Lietz & Gregory, 1978) black children were referred to the office for disciplinary action more often than white children, but no similar difference rate was found for referrals to the multidisciplinary diagnostic team.

While these studies suggest that some attitudes may be changing, they do not establish a situation of nondiscrimination in the special education diagnostic process. Indeed, there is much more evidence which attests to a continuing problem. Studies conducted since the 1977 nondiscriminatory legislation went into full force indicate that assessment bias continues to operate against ethnic minorities, males, linguistic minorities, and children with behavior problems. There is even some evidence of age discrimination in the handicapped diagnosis.

The majority of studies have continued to focus on ethnic biases. After reviewing enrollment figures in a number of districts throughout the nation, the Education Advocates Coalition (1980) concluded that the new laws "had not proven adequate to perform the task" (p. 2). They found, for example, that black children were being misclassified as EMR at a rate over three times that of white children. School districts were found to be diagnosing linguistic minority children as handicapped using incomplete and/or inappropriate evaluations. Other evidence (Manni, Winikur & Keller, 1980) has demonstrated a similar problem of ethnic proportion in classes for the emotionally disturbed. In a longitudinal study Tucker (1980) followed enrollment patterns in 50 southwestern school districts for eight years. He concluded that the

proportion of minority ethnic enrollment in EMR dropped following the P.L. 94-142 legislation. The increased minority enrollment in LD classes was so great, however, that minority children are now overrepresented in a category which had been almostly exclusively white. Having reviewed the results of "substantial" national efforts, members of the New Jersey State Department of Education concluded that "the concern over disproportionate representation of (ethnic) minorities in special education programs remains a reality and likely will continue to be so well into the next decade" (Brady et al., 1983, p. 53).

A recent court decision (Martin Luther King Jr. Elementary School children v. Ann Arbor, 1979) provided more specific guidelines for nondiscriminatory linguistic minorities, indicting that the federal legislation had not been sufficient to prevent this type of bias. At issue was not bilingual assessment, which has been discussed previously, but the evaluation of English dialect speakers. The suit was brought on behalf of 15 preschool children whose primary language was a nonstandard dialect of English; that is, black English. The school system was charged with having failed to determine whether the children's learning disabilities stemmed from cultural, social, economic, or academic factors during assessment for special education classes. Expert witnesses testified that:

(1) Students who speak black dialect experience difficulty hearing and making certain sounds used discriminantly in standard English.

(2) The unconscious but evident attitude of teachers causes a psychological barrier to student learning (p. 1381).

As a result of this testimony the Ann Arbor school system was ordered to propose a plan to help teachers identify black English speakers and to use that knowledge in testing and in teaching these students. Vaughn-Cooke (1981) points out that this will be a difficult process as many standardized tests are invalid for dialect speakers, although they have been repeatedly used with this group. There are no figures concerning the overrepresentation of dialect speakers in special education; however, given the attitudes of teachers described above and the paucity of valid assessment measures, such overrepresentation appears likely.

Another area which has not responded to nondiscriminatory strategies has been sex bias. Males continue to outnumber females in elementary and secondary classes by a 2:1 ratio (Tomlinson, et al., 1977). Those girls who are identified as LD typically have much more severe academic problems than boys (Owen, 1978). The same 2:1 ratio has also been found at the preschool level for LD, speech impaired (SI) and emotionally disturbed (ED) categories in Head Start (Walters, 1978). More recent research (Phipps, 1982) indicates a similar trend for older children. Approximately 85% of the LD, EMR, and ED children in one school district were males.

Investigation of the reasons for disparate enrollment figures of males and females have generally focused on student behavior. Phipps (1982) found that teacher perception of conduct and behavior problems in boys "clearly played a major role in the identification of children who ultimately participated in special education programs. (This is an example of) the 'squeaky wheel' concept" (p. 425). Teachers are often aware of this tendency sometimes stating that they refer the loudest children first (Weatherly &

Lipsky, 1977). A random sample of 100 North Carolina elementary teachers reported that they were most likely to make special education referrals of children who had behavior problems, passing up non-behavior problem children with similar academic test results (Giesbrecht & Routh, 1979). Diagnosticians were found to ignore standardized test information which indicated no behavior disorder, retaining an earlier stereotype of problem behavior indicated by teacher referral forms (Ysseldyke & Algozzine, 1982).

While not clearly established as a factor in discriminatory assessment, age and geographic location have recently been shown to be influential factors. In a national study of over 14,000 students, Craig and McEachron (1978) found that teachers tended to identify far more younger elementary children (grades 1-3) than any other age group. Regional differences in referrals were also found in this national study, although this presented a "puzzling," inconsistent pattern. An analysis (Bernard & Clarizio, 1981) of planning and placement decisions for approximately 1,000 students in the state of Michigan revealed that age, sex, IQ and geographic location contributed substantially to the special education diagnostic decision.

Clearly the present legislation against discriminatory testing and the strategies adopted to comply with the law have not had the desired effect. Despite some data to the contrary, there is overwhelming evidence that sex, ethnic, linguistic, behavior, age and regional bias still exists in special education assessment.

Thus far, the discussion of nondiscriminatory strategies has focused on the elimination of bias in tests and testing procedures. Some of the studies cited above which investigated the referral process have provided a clue to

other possible sources of bias. Reschly (1981) and others have begun to search for other causes of assessment bias beyond those associated with measurement instruments. These non-measurement sources are considered in the following section.

Discrimination in Special Education Diagnosis: Non-measurement Bias

Controversies about the discriminatory effects of tests in special education assessment have begun to take a new form. New evidence indicates that it is not the actual test instruments which are responsible for misclassification. It is certainly clear from the previous discussion that test reforms have produced little effect. More importantly, test reforms may not even have been needed in the first place. This section focuses on a review of recent evidence which suggests that it is test usage, not the instrument itself which results in discriminatory assessment. Two factors associated with test usage - diagnostic criteria and diagnostician's background - have been found to be very influential and will be reviewed in depth.

Test reform is not the answer. A number of recent studies have found little or no evidence of bias in standardized tests even for the WISC-R which is the most commonly used instrument in special education diagnosis (Jirsa, 1983). Oakland and Feigenbaum (1978) found no consistent patterns of WISC-R bias on any basis including sex, race, socioeconomic status, age, birth order, health, family size, family structure or geographic region. Sandoval (1979) stated that "test bias against minority children (has) not been found in the WISC-R" (p. 926). Clarizio (1979) and Vance and Wallbrown (1978) have documented similar findings. In a highly rigorous factorial study, Gutkin and

Reynolds (1981) concluded that "the WISC-R is not biased when used to assess the intellectual functioning of black children" (p. 230).

Similar evidence exists for other tests and for other minority groups (Gutkin & Reynolds, 1981; Jirsa, 1983; Reschly, 1978). In an Illinois Federal District court case, which is now in the appeal process, the court concluded that IQ tests are largely unbiased and that overrepresentation per se does not violate legal or constitutional protections (Bersoff, 1981).

Even if culture free or culture fair tests were the answer to assessment problems, many have argued that there has been little success in developing such instruments (Cole, 1981; Nazzaro, 1976; Reschly, 1981). According to Henry Dyer of the Educational Testing Services (Purvin, 1975), a test can be culture fair only if (1) the information required to perform acceptably or (2) the stimulus materials are available to all people of all cultures. He feels that neither of these conditions is possible. Podilla and Garza (1975) note that "efforts to develop culture-free tests were doomed to failure from the beginning. All human experience is modulated by human society, and no test can be experience free. The materials used in the test, the language of the test, the manner of getting the testee to respond, the criteria for choosing which responses to record, the categories into which responses are classified, the test's validity criterion - all are culture bound" (p. 54).

Given these uncertainties about culture fair testing, some diagnosticians have turned to other aspects of the assessment process as possible sources of bias. Studies cited earlier, for example, pointed to discrimination in the referral process -before the child is seen for diagnostic testing (Matuszek & Oakland, 1979; Reschly, 1981; Ysseldyke & Algozzinne, 1982). Reschly (1979)

suggests that IQ tests at worst have only a neutral effect on minority overrepresentation, while it is possible that they actually reduce the degree of overrepresentation that would result from the discriminatory referral process.

Non-test bias may be introduced not only at the referral stage, but during the diagnostic/classification process as well. A number of investigations (Cleary, Humphreys, Kendrick & Wesman, 1975; Harber, 1981b; and Reschly, 1981) have argued that it is not the tests themselves but the interpretation and use of test and referral information that leads to discrimination. Two major factors associated with discrimination in test use have been identified in the literature; these are, (1) criteria for applying test results and (2) diagnostician's training and background. The following two sections are devoted to a review of research related to these two non-measurement factors.

Criteria for Special Education classification. There can be little doubt that diagnostic and eligibility criteria for special education classification have an important effect on the assessment process. Some have concluded that criteria are the crucial factor in nondiscriminatory assessment (Duffey, Salvia, Tucker & Ysseldyke, 1981; Evans, 1979; Malgoire, Craig & Kaskowitz, 1979; Martin, 1979). Consistent, objectively stated criteria are thought to be the key to fair and accurate evaluation. "Biased decision making may be attributed to incomplete and inconsistent operationalizing of definitions and/or criteria on which decisions are based. The inherent problem is that decisions are made on the basis of situation specific (and perhaps even assessor specific) definitions and rules" (Duffey et al., 1981). The obvious question is "How complete and standard are eligibility criteria for special education

classification?" The answer to this question involves a consideration of how criteria are developed, what criteria are presently in existence and how those criteria are actually applied.

Although the Federal Government has traditionally exercised relatively little direct authority over school assessment practices, P.L. 94-142 widely expanded the sphere of federal control and influence on assessment practices. Consequently, the federal definitions of handicapping conditions must form the basis for state and local eligibility criteria. The federal definitions for each handicap are presented in Exhibit 1. Each state is required to develop an annual program plan which includes diagnostic procedures and eligibility criteria based on the federal standard. All school districts and other public agencies are then required to submit similar plans which include assessment procedures to the state (Martin, 1979).

State education agency (SEA) officials have accepted varying degrees of responsibility for developing criteria. Some have simply restated all or some of the federal definitions and left more specific criteria to individual districts or local education agencies (LEA's) (Alabama Department of Education, 1980; Oklahoma State Department of Education, 1979; and Pennsylvania Department of Education, 1980). The majority of SEA's have adopted a limited role by including a slightly more thorough, but not operationalized description of each handicap (Kansas State Department of Education, 1980; Rhode Island Board of Regents for Education, 1978; and West Virginia Department of Education, 1980). Officials from the State of Washington (Department of Public Instruction, 1980) have elected to take a much more active role. Their definitions and eligibility criteria for each handicapping condition list (1) the

EXHIBIT I

Federal definitions of handicapping conditions (U. S. Congress, 1975)

1. "Deaf" means a hearing impairment which is so severe that the child is impaired in processing linguistic information through hearing, with or without amplification, which adversely affects educational performance.
2. "Deaf-blind" means concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational problems that they cannot be accommodated in special education programs solely for deaf or blind children.
3. "Hard of hearing" means a hearing impairment, whether permanent or fluctuating, which adversely affects a child's educational performance but which is not included under the definition of "deaf" in this section.
4. "Mentally retarded" means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period, which adversely affects a child's educational performance.
5. "Multihanidcapped" means concomitant impairments (such as mentally retarded-blind, mentally retarded-orthopedically impaired, etc.), the combination of which causes such severe educational problems that they cannot be accommodated in special education programs solely for one of the impairments. The term does not include deaf-blind children.
6. "Orthopedically impaired" means a severe orthopedic impairment which adversely affects a child's educational performance. The term includes impairments caused by congenital anomaly and impairments from other causes (e.g., cerebral palsy, amputations, and fractures or burns which cause contractures).
7. "Other health impaired" means limited strength, vitality or alertness, due to chronic or acute health problems such as a heart condition, tuberculosis, rheumatic fever, nephritis, asthma, sickle cell anemia, hemophilia, epilepsy, lead poisoning, leukemia, or diabetes, which adversely affects a child's educational performance.
8. "Seriously emotionally disturbed" is defined as follows:
   (i) The term means a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree, which adversely affects educational performance:
      (A) An inability to learn which cannot be explained by intellectual, sensory, or other health factors;
      (B) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers;
      (C) Inappropriate types of behavior or feelings under normal circumstances;
      (D) A general pervasive mood of unhappiness or depression; or
      (E) A tendency to develop physical symptoms or fears associated with personal or school problems.
   (ii) The term includes children who are schizophrenic or autistic. The term does not include children who are socially maladjusted, unless it is determined that they are seriously emotionally disturbed.
9. "Specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. The term does not include such conditions as perceptual handicaps, brain injury, minimal brain disfunction, dyslexia, and developmental aphaisa. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, or of environmental, cultural, or economic disadvantage.
10. "Speech impaired" means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, which adversely affects a child's educational performance.
11. "Visually handicapped" means a visual impairment which, even with correction, adversely affects a child's educational performance. The term includes both partially seeing and blind children.

types of tests to be administered, (2) regulations to encourage culture-fair testing which also does not discriminate on the basis of sex and (3) specific cut-off scores. The Louisiana Board of Education (1980) offers the same degree of specificity and recommends the SOMPA as a nondiscriminatory estimate of choice for evaluating intellecutal potential. The criteria outlined by the Nebraska Department of Education (Campbell, 1979) for preschool children are brief but offer operationalized definitions. Regulations published by the Georgia Department of Education (1978) indicate the powerful role of this agency. Each type of handicap is first operationally defined in terms of student characteristics and types of assessments. A more limiting set of criteria, however, is implied as quota restrictions are also supplied for each handicap category.

There is a similar variation in acceptance of responsibility at the LEA level. Many districts choose to reiterate SEA policy with no further comment. (See e.g., Bibb County Board of Education, 1979; Guntersville Board of Education, 1980). For some districts this policy results in no more than a copy of the federal guidelines (Allegheny County Public Schools, 1980; Oklahoma City Education Department, 1980). Some districts have experienced particular difficulties with eligibility criteria for learning disabled children and have published additional clarifying guidelines which extend beyond state guidelines. (Arlington County Board of Education, 1981; Half Hollow Hills District Board of Education, 1979.) Some larger districts particularly those with large minority populations have developed very comprehensive eligibility criteria. Dade County Public Schools (1979) in Miami, Florida devotes 15-20 pages for <u>each</u> handicap in order to provide a complete list of procedures and criteria.

Specific tests or types of tests are suggested in conjunction with operationalized criteria. The Fresno Unified School District (1981) lists similar specific guidelines.

Just as SEA and LEA regulations differ in the level of specificity, they also vary widely in the content of the criteria. One handicap which has been particularly hard to define has been learning disabilities (LD). The original federal definition (see Exhibit 1) was suggested in 1974 by the National Advisory Committee on the Handicapped (Beatty, 1977). This guideline proved to be too general as enrollment in learning disability classes skyrocketed from 1975 to 1977 (Savage, 1977). Congress demanded that the Bureau of Education for the Handicapped (BEH) develop specific procedures and diagnostic criteria for determining what constitutes a learning disability (42 Federal Register 65081). One trial formula offered by BEH used chronological age (CA), IQ score and grade placement score (GR) to define LD:

LD if CA ((IQ/300) + .17) - 2.5 GR (Reynolds, 1980).

This formula was criticized as having "managed to capture in one procedure the difficulties of basing an estimate of educability on an IQ score plus all the problems of difference scores" (Reynolds, 1980, p. 603). The discrepancy formula was withdrawn but not replaced (Martin, 1979). After consulting experts in education, psychology and medicine, BEH concluded:

1. The state of the art in the field of specific learning disabilities and its associated fields is such that it is not presently possible to specify exactly all of the components of each specific learning disability. There remain strong opposing professional opinions as to the validity of certain

behavioral manifestations as being indicative of a specific learning disability. At present, the only generally accepted manifestation of a specific learning disability is that there is a major discrepancy between expected achievement and ability which is not the result of other known and generally accepted handicapping conditions or circumstances.

2. There exists no hard research data collected on a large enough sample in order to state, with certainty, which are the common characteristics of all learning disabled children.
3. There are several theories as to what causes children to have specific learning disabilities.
4. There appear to be no generally accepted diagnostic instruments which can be singly and appropriately utilized . . . with all such children. (Savage, 1977, p. 75)

After reviewing each state's criteria for LD, EMR and Emotionally Disturbed (ED), Hallahan and Kaufman (1977) concluded that the regulations were confusing, contradictory and not specific. The range of definitions was so wide that it was difficult to distinguish one handicap from another. The range of prevalence figures among states (LD, e.g., ranged from 1% to 30%) was further evidence of the confusion arising from these nonspecific guidelines. Murai (1982) found LD criteria to be arbitrary and based mainly on fiscal concerns.

A survey of definitions of LD offered by 42 state departments of education was equally disappointing (Mercer, Forgnone, & Wolking, 1976). Among the SEA's the following differences were found:

- lack of consensus about IQ range

- wide variation in the flexibility and autonomy given to LEA's setting criteria
- types of disorders included varied widely
- only 62% excluded visual or hearing impairment
- wide variation in other types of characterisitic included.

A similar review of definitions of preschool handicap revealed that only 57% of the SEA's had specific definitions or any provisions for identifying this population (Lessen & Rose, 1980). Additionally, there were wide variations in the age for onset of services, criteria for demonstrated need and categorical definitions.

Although the category, ED, has not received the attention afforded the definition of LD and EMR disorders, a few recent investigations have suggested that criteria for this disorder are equally vague. Bischoff and Shakel (1980), for example, have expressed concern about Alaska's current state definition of ED. This is the only category in Alaska's guidelines which specifies noneducational treatment as part of the eligibility criteria. They further criticized the ED criteria for being vague, relying on noneducational experts and specifying treatment as part of the criteria. In a survey of 49 state definitions of ED, Epstein, Cullinan and Sabatino (1977) discovered wide discrepancies between states. They criticized most definitions for being vague, ambiguous, and self-contradictory. Mack (1980) conducted a similar study a few years later and found the situation no better. States did not even agree on the term for the handicap. Labels ranged from "emotionally conflicted" in Alabama to "significant identifiable emotional or behavioral disorder" in Colorado. Two states, Massachusetts and Vermont, had no

specific criteria as they offer only noncategorical special education programs. Surveys conducted by Gresham (1981) and Reynolds (1980) support these findings. Bischoff and Shakel (1980) conclude that eligibility criteria and definitions for other handicaps may be equally problematic. Recent emphasis on EMR and LD regulations has merely hidden these problems.

Given the vague, confusing criteria at the federal, state and local level, it is not suprising that the responsibility for diagnostic decision making generally filters down to the individual diagnostician or diagnostic team. Several sources have suggested that even when criteria are available, they are often so confusing and contradictory that they are ignored. More than 45% of the students enrolled in LD classes in Colorado, for example, did not meet the state standards (Shephard & Smith, 1981). In one district, no psychometric differences were found between students in LD classes and low achievers in regular classrooms (Ysseldyke, Algozzine, Shinn & McGue, 1982). In another school district, diagnostic decisions did not correlate with decisions based on the application of federal definitions (Algozzine, Ysseldyke, & Shinn, 1982). Ysseldyke and Algozzine (1982) found that educational staff often based diagnostic decisions on referral information. They relied on their own internal criteria rather than applying test results to legal definitions. The researchers conclude that "even before a child utters a response to a test item, the assessment cards may already be stacked" (p. 228).

There is overwhelming evidence that vague definitions are associated with a very wide range of application of criteria. Li (1977) found that each educational diagnostician had his or her own definition and eligibility criteria for LD. A large national research project (Malgoire, Craig, & Koskowitz,

1979) revealed a wide variation in criteria for several handicap categories as well as different indicator sets within a particular handicap. Another study (Johnson, 1980) suggests that severity is a factor in standardizing criteria. When IQ-achievement discrepancy scores were severe, a group of 356 school psychologists made predictable EMR, LD and ED placements. When scores were less severe or other complex combinations of factors were present no predictable criteria emerged. Johnson found that the study participants could not agree on criteria for diagnosis or placement. Similar results were obtained in a study of special education directors. Although the directors have been found to be the best informed staff concerning the legal and practical issues of diagnostic decision making (Fenton, Yoshida, Maxwell, & Kaufmann, 1979), their descriptions of the assessment sequence and procedures varied greatly (Poland, Ysseldyke, Thurlow, & Mirkin, 1979). Some special education directors, for example, clearly differentiated between placement and eligibility decisions while others did not.

Another method of investigating the universality of criteria has been to study inter-rater reliability for diagnosis of the same cases. When 18 educational diagnosticians were asked to diagnose 100 children simply as LD or non-LD they "were extremely inaccurate in their classification and in little agreement with each other" (Epps, McGue & Ysseldyke, 1982, p. 209). Frame and others (1982) found interclinician agreement to be only about 30% for classification according to handicap category. Case studies of handicapped children whose families moved across school districts caused Ysseldyke and Thurlow (1980) to conclude that "the decision to declare a student eligible for service may be entirely situation specific" (p. 104). A child who was declared

ineligible in one district, for example, was found eligible in another. The researchers concluded that "were these students to be referred to another school, it is highly probably that a different outcome would result again" (p. 104).

Other investigations have focused on the measurement procedures used to make diagnostic decisions. Educational diagnosticians (LD teachers, psychologists and speech-language clinicians) often choose tests with poor reliability and validity, even when superior instruments are available (Davis & Shephard, 1982). Even when adequate instruments were chosen 25-50% of these diagnosticians did not interpret scores correctly. Similarly, Montgomery County Public Schools in Maryland were judged to be in compliance with all federal and state legislation except that related to test selection and interpretation (Silversmith, 1980). Thurlow and Ysseldyke (1979) reached similar conclusions in a study of Child Service Demonstration Centers, funded by BEH to develop and use exemplary practices.

One reason offerred for the failure of diagnosticians to choose adequate instruments or to interpret them correctly has been poor training (Davis & Shephard, 1982; MacMillan & Meyers, 1979). In the following section diagnostician background and training is considered as a factor in discriminatory assessment.

Diagnostician background and training. Reschly (1981) examined a number of issues related to discriminatory assessment and concluded that the examiner's background was a highly influential variable in the assessment of minority children. He states that "it is essential that the examiner be specifically trained to test these special examinees. Some of the factors to be

considered are the examiner's background, sex, testing style, and acculturation" (p. 1083). Bennet (1981) concurs that the examiner's training may be as important in discriminatory assessment as test bias or inadequate criteria. His research uncovered "serious problems with regard to the competence of professionals involved in the assessment of exceptional children" (p. 437). Some of these problems included:

- misuse of tests
- inadequate preparation and training
- interrater variation over test scoring and interpretation
- mechanical scoring error.

Most research concerning diagnostic background has focused on Bennet's concern of inadequate training. Garrison and Hammill (1977) concluded that a minimum of 25% of all EMR children in Philadelphia had been misplaced due directly to the diagnostician's poor test interpretation and inadequate knowledge of cut-off scores. Outright charges of "inappropriate and unprofessional behavior" have been leveled against diagnosticians for improper test usage (Kirp & Kirp, 1976, p. 85). Ysseldyke (1978) concluded that many school personnel administer and interpret tests without the training to do so. He believes that the major problems are a failure to consider the technical limits of the test and a lack of awareness of the nature of standardized samples. Bennet (1980) found that a group of 95 educational diagnosticians averaged only 50% correct responses on a 64 item test of measurement competence. Gary McDaniels, former director of the Office of Special Education's Division of Assistance to States, admitted that professional competence in the selection and administration of tests is a major concern

(1979). He noted that the two most common problem issues in P.L. 94-142 were (1) measurement practices and (2) personnel training - not improvements in the technology of measurement.

The effects of the diagnostician's experience as well as assessment training, were investigated in one study which used simulated case files of EMR and "normal" children (Amira, Abramowitz & Gomes-Schwartz, 1977). The investigators found no race or socioeconomic bias in decision making. They also found no differences in diagnostic decisions with years of experience, years of training or types of training.

Many have called upon state and local governments to tighten certification requirements for special education personnel. MacMillan and Meyers (1979) feel that this is the only safeguard to prevent irresponsible, incorrect test scoring and interpretation. Meyers, Sundstrum and Yoshida (1974) concluded that one reason professionals on assessment teams were inadequately prepared was that they had been "grandfathered in". Due to the recency of changes in legally specified diagnostic standards, many staff who did not meet those new standards were allowed to stay with no provisions for updating their skills. During the same period, training programs were criticized for being particularly deficient in teaching skills related to the analysis and interpretation of data, particularly in regard to the evaluation of minority children (Bransford, 1974).

A national study of certification requirements for special education personnel noted a wide variation among states; e.g., (1) special education preschool staff had to meet stringent requirements in some states, while other states had no requirements; (2) some had established freestanding

requirements for special education, while others treated it as an endorsement attached to regular education; and (3) there were no general terms for certification of those involved in diagnostic decisions (Barresi & Bunte, 1979).

Summary. Discriminatory assessment has persisted in special education diagnosis despite the intensive efforts of legislators, test developers and professional educators. These efforts, however, were focused primarily on revising test instruments. Recently, the search for other sources of discrimination has led to an investigation of non-instrument factors, that is, not the tests themselves but how the tests are used. Eligibility criteria and the diagnostician's background and training have been identified as two important potential sources of bias. Eligibility criteria were shown to vary widely throughout the United States and to be incomplete and often not operationalized. Much attention was given to criteria for three special education categories deemed difficult to define - ED, LD, and EMR. It is highly likely, however, that other handicaps suffer from the same problems of inconsistent and inadequate definition of eligibility criteria.

Part II: Speech-Language Diagnosis

Speech impairment[1] is one of the major special education categories as it includes almost half of all handicapped children (Kakalik, Brewer,

[1]The term "speech impairment" will be used to refer to that category of disorders which includes all forms of communication disorders. Other common terms for the same handicap include "speech-language disorder" and "communication impairment".

Dougharty, & Fleischauer, 1973; Sabatino, 1977; U. S. Department of Health and Human Services, 1980). Speech impaired children thus represent a large portion of those included in the minority and sex overrepresentation figures cited earlier. That is, despite the Protection in Evaluation Procedures and other strategies to eliminate bias cited earlier, there are still more ethnic minorities, linguistic minorities and males among speech impaired (SI) children than would be expected, given their representation in the population. Studies limited specifically to the SI population provide further evidence of overrepresentation for younger children, those from low income families, ethnic and linguistic minorities, and males (Craig & McEachron, 1978; Emery, 1973; Granger, Matthews, Quay, & Verner, 1977; and Walters, 1978).

While some research has been conducted in the area of speech test bias, non-measurement sources of bias have not received the same attention devoted to LD, ED and EMR categories. This may be due to the fact that the SI label is considered less stigmatizing (Randolph, 1977) and to the lower costs associated with serving speech impaired children when compared to other handicaps (Kakalik et al., 1973). SI diagnosis has the same potential for both test and non-measurement bias that exists for other handicaps. In fact, there are some sources of bias specific to the speech and language diagnostic process. Part II is devoted to a review of test and non-measurement sources of bias as they relate to the diagnosis of SI.

Discrimination in Speech Impaired Diagnosis: Test Bias

A particular concern in the testing of children with potential speech impairment is the child's primary language or dialect. Until 1973 all speech

tests seemed to have been standardized only for standard English speaking populations (Evard & Sabers, 1979). Some research, however, has indicated that dialect patterns incorporate a fully developed linguistic system with predictable rules (Baratz, 1968; Labov, 1970). Consequently, test developers have begun to include speakers of dialects as well as speakers of other languages in the norming populations. As linguistic minority speakers are also generally members of an ethnic minority, attempts have also been made to develop culturally fair items (Evard & Sabers, 1979). The problems in eliminating this type of bias are reviewed below.

Grill and Bartel (1977) investigated the Grammatic Closure test of the Illinois Test of Psycholinguistic Abilities (ITPA). Their findings indicate that at least one third of the test items discriminate against black dialect speakers; that is, responses that were appropriate in the dialect for those items but scored as incorrect standard English accounted for 52-100% of all errors. Grill and Bartell conclude that this subtest is inappropriate for those who speak a nonstandard dialect, even infrequently. Vaughn-Cooke (1982) noted that two-thirds of the items on this subtest were potentially discriminatory and that this measure should be renamed the "Standard English Grammatic Closure Subtest" (p. 31). Vaughn-Cooke also reviewed three other widely used speech and language tests (Utah Test of Language Development, Peabody Picture Vocabulary Test and the Bankson Language Screening Test) and found them inappropriate not only for nonstandard dialect speakers but for standard English speakers as well due to invalid assumptions about language and poor reliability.

The Test for Auditory Comprehension of Language was developed in both Spanish and English and specifically constructed for different ethnic/racial groups (Carrow, 1973). Test norms, however, were collapsed into a single norm when no differences among groups were found. This procedure has been criticized as no low socioeconomic populations were included. Evard and Sabers (1978) suggest that group differences would have appeared if a low income population had been sampled. Similar tests have been developed for black and Spanish dialect speakers (Drumwright, Van Natta, Camp, Frankenburg & Drexler, 1973; Toronto, 1976) and for specific ethnic and racial groups (Arnold & Reed, 1976; Foster, Giddan & Stark, 1973). These procedures have been criticized for being based on inadequate samples, containing insufficient information or being difficult to obtain (Evard & Sabers, 1979).

Additional evidence indicates that several other speech-language measures suffer from a number of psychometric inadequacies. Empirical studies have consistently demonstrated that the Peabody Picture Vocabulary test (PPVT) not only discriminates against ethnic and linguistic minorities, but that it also overestimates IQ by about six points when compared to the WISC. The Revised PPVT has also been found to show poor correlation with other placement IQ tests (Bracken & Prasse, 1981).

Miller and Prutting (1979) found a similar inconsistency among three measures of language comprehension; i.e., the Test of Auditory Comprehension of Language, the Northwestern Syntax Screening Test, and the Bellugi Comprehension Test. Longhurst (1977) obtained similar results using three other comprehension measures. Young subjects correctly identified a grammatic form on one test but not on another, indicating that pictorial

ambiguity, linguistic content, sequencing or other factors associated with the tests were the source of bias.

At least one expressive measure, the Developmental Sentence Scoring procedure has proven to be unreliable, producing significantly different scores when administered to the same Head Start children in different structured situations (Longhurst & File, 1977). Less structured situations (play or adult-child conversations) produced significantly higher scores than more structured environments (response to pictures). Articulation tests have recently received similar criticism. Schissel and James (1979) found significant inconsistencies in scores obtained by young children on the Deep Test of Articulation vs. the Arizona Articulation Proficiency Scale. An investigation of responses on the Photo Articulation Test, the Templin Darley Test of Articulation and the Goldman Fristoe Test of Articulation also found significant differences between test scores (Madison, Kolbeck, & Walker, 1982). Descriptive analysis of the response differences indicated that test items were sometimes inappropriate for younger children and contributed significantly to inconsistent results. Others have criticized the stimulus items on articulation tests for being ambiguous (Whitehead & Mullen, 1975) and unreliable with younger children (Shanks, Sharpe & Jackson, 1970).

Although these studies of inconsistency between tests do not establish a situation of discrimination against any particular group, they do establish the presence of conditions which are conducive to discriminatory assessment. When several conflicting scores are available for one examinee or if scores are known to be unreliable, test scores cannot be interpreted in a standardized manner. Diagnosticians must rely on their own individual criteria which was

earlier shown to be associated with biased interpretation. Other issues in test interpretation and usage are presented in the following section.

Discrimination in Speech-Language Diagnosis: Non-measurement Bias

This section begins with an overview of the speech-language diagnostic process. Familiarity with the various types of communication disorders and procedures for assessment is essential to an understanding of sources of non-measurement bias. Discriminatory issues related to the definition of eligibility criteria and to the diagnostician's background are then reviewed.

The nature of the speech impaired diagnosis. Five major areas have been identified in the assessment of speech impairment (Longhurst, 1977; Prutting, 1979). These include:

(1) articulation - production of phonemes (sounds);

(2) language - production (expressive) and comprehension (receptive) of all units of communication, includes phonologic, morphologic, syntactic, vocabulary and auditory processing skills;

(3) rhythm - fluency of speech, primarily includes stuttering disorders;

(4) voice - quality, pitch, nasal emission and volume of speech; and

(5) pragmatics - socio-linguistic skills such as modifying speech as a function of the listener's age, initiating discourse and responding to subtle turn-taking cues.

The majority of speech impaired children are affected by disorders of language and/or articulation. Powers (1957) noted that articulation was by far the most common type of SI. New assessment procedures have focused on language diagnosis and incidence figures have begun to change; however,

articulation disorders still represent the major type of SI. In Head Start (U. S. Department of Health and Human Services, 1980), SI enrollment includes 46% with articulation as the primary problem and 44% with language disorders. In the public school system 50% of all children with SI have articulation disorders, while approximately 25% are diagnosed as having a language problem (Sabatino, 1979).

The following review of assessment features will focus on problems inherent in the evaluation of language and articulation. Other types of speech disorders will not be included in the discussion due to their low incidence in the preschool and school-age population.

Language. Assessment of receptive and expressive language skills includes a wide variety of procedures designed to evaluate several different skills. Lamberts (1979) noted the scope of abilities to be included in language testing:

> Language cannot be separated from children's ability to organize information and construct a cognitive foundation of knowledge about the environment. This knowledge, in turn, depends on ability to receive and process environments sensory stimulation, and derive percepts of objects/events reflecting past experiences. Language is seen as part of, and deriving from, a child's general cognitive processing ability (p. 290).

Longhurst (1977) reviewed language tests available for use with preschool and school-age children. She found a large number of tests which purported to measure one or more specific aspects of language. After attempting to compare several tests which were labeled as measures of syntax and morphological comprehension skills, Longhurst concluded that "they vary

greatly among themselves in terms of structures tested, presentation of items, scoring, and interpretation" (p. 122). Additionally, most of the tests have been developed since the relatively recent interest in language assessment in the early 1970's[1] and no one test or set of tests designed since then has yet become a "classic" which is widely used and accepted. A recent survey (Davis & Shephard, 1982) of speech therapists in Colorado indicated that the most frequently used language measure is the Peabody Picture Vocabulary Test which had been developed as a measure of verbal IQ before the interest in language testing began (Dunn, 1965). The Colorado survey listed over 20 language tests commonly used by diagnosticians in that state.

Articulation. Although several articulation measures were found to be in common use throughout the United States, Longhurst (1977) noted that "the procedures for testing in most of these tests are quite similar" (p. 114). The Oregon survey noted that a wide variety of language tests were commonly used by speech-language diagnosticians but found that only one articulation test (Goldman-Fristoe Test of Articulation) accounted for most of this type of measurement (Davis & Shephard, 1982). Longhurst concluded that the major problem with inconsistency in articulation testing is the norms available for judging the child's articulation mastery. This problem was of particular concern in the assessment of preschool children, as some articulation errors

[1]The American Speech, Language, and Hearing Association, for example, did not add "language" to its title until 1979.

are expected in their speech. Decisions concerning which errors or combination of errors constitute a speech impairment are confounded by inconsistencies in articulation norms. (See e.g., Fudala, 1963; Templin & Darley, 1969). According to Longhurst (1977) most experienced examiners would suggest that the Templin-Darley (1970) provides the best norms of any of the tests. Studies of norms and other criteria used by speech-language diagnosticians, however, have seriously challenged this statement. Before reviewing articulation criteria presently in use, however, it is useful to consider the interaction between language and articulation.

Language influences on measures of articulation. The interaction between language and articulation disorders further complicates the type of SI diagnosis. Children labeled as "severely language disordered" often have average language performance but severe articulation difficulties (Randal, Reynell, & Curven, 1974). The severity of the phonological problem is misinterpreted as a language deficit when the child's speech is unintelligible. Although it would be possible to err in the opposite direction (i.e., diagnosing an articulation problem as a language disorder) Menyuk (1971) observes that this is far less likely to occur. She notes that language-delayed children often have phonological problems, while the frequency of language delay in articulation disordered children is much lower. A number of other studies have confirmed that a clear relationship exists between articulation and language skills (Marquardt & Saxman, 1972; Schumauch et al., 1978; Schwartz, Leonard, Wilcox, & Folger, 1980; Shriner et al., 1969).

Articulation and language assessment is thus not an open and shut case. Language was shown to be a complex set of skills which are measured

differently by various instruments, all purporting to measure the same skills. It is difficult even to establish a clear definition of language or of the skills encompassed by the term. Articulation assessment, while more clearly defined, is complicated by an interaction of language skills and inconsistent test norms. Clearly, the articulation and language criteria used by speech-language diagnosticians is a potential source of non-measurement bias.

Criteria available to assist in the diagnosis of SI children. As early as 1966 a number of speech-language professionals from school districts all over the country expressed concern about the criteria used to select children as SI (Allen, Black, Burkland, Byrne, Farquhar, Herbert, & Robertson, 1966). Their discussion of different forms of articulation testing (isolation, deep testing) and consideration of other potential factors (parental concern, prognosis, other learning problems) underscores the wide differences in diagnostician judgment concerning SI criteria. Diagnosticians could not come to any philosophical agreement concerning what factors to consider in the diagnosis, how to evaluate or how to interpret test results. During the same year, several university directors of clinical training programs also expressed their points of view about the wide range of SI criteria (Webster et al., 1966). They commented on the types of information included in their training programs, further emphasizing the inconsistency and range of eligibility criteria. One observor summarized the results of their discussion: "Each of these professionals has indicated a wide range of variables that may influence the final decision on which children will receive a speech therapy program" (p. 358). Other speech-language professionals have indicated additional variables which individual diagnosticians may or may not consider, depending upon their

personal view of case selection responsibilites; these included case load priorities, prognosis (Flower, Leach, Stone, & Yoder, 1967) and the diagnostician's need for learning, innovation and experimentation (Henrickson, 1968).

Speech and language diagnosticians have continued to emphasize problems of inconsistent application of criteria, calling for greater reliability in the SI diagnosis (Adams, 1976; MacDonald & Martin, 1973; Young, 1975). Wingate (1977) looked specifically at the definition of stuttering. He noted that inter-observer reliability dropped significantly when stuttering was not defined in terms of specific behaviors, but was left to the speech diagnostician. He concluded that the diagnosticans each have very different ideas of what constitutes stuttering.

Michel (1978) conducted an extensive review of the history of the articulation diagnosis. She notes that test instruments have become increasingly sophisticated with multidimensional interpretation and more meaningful recommendations, but concludes that "we must strive for greater precision in the definition of terms" (p. 424). Longhurst (1977) also focused on the diagnosis of articulation. She concluded that one of the most difficult issues was the articulation assessment of preschool children. She noted that "it is perfectly normal for preschool children to have difficulty with articulation and there is a tremendous range of normal for articulation skills" (p. 101). Templin's classic 1957 study supports this contention. She found that the average three year old child correctly articulated 93 out of 176 items, but with a standard deviation of 34 items.

After reviewing the diagnosis of all types of SI, Freilinger (1978) concludes that case selection criteria has troubled the speech pathology profession for many years. He criticizes a table of criteria for case selection which he helped to develop (Shine & Freilinger, 1962). Freilinger now feels that the table is "not very helpful" (1978, p. 9). Rees (1978) reviews eight unresolved issues in the diagnosis of speech and language disorders coming to the "inescapable conclusion that the profession lacks a set of unified principles underlying current approaches to diagnosis" (p. 20).

Other professionals have noted the same variation in diagnostic criteria, but concluded that it represents a positive situation. Emerick and Hatten (1974) de-emphasize the importance of objective scores, focusing instead on the importance of other factors specific to the individual child: "We diagnose communicators, not communication" (p. 20). The American Speech, Language and Hearing Association (ASHA) (Dublinske, Karr, & Downey, 1981) endorses this position, calling for flexibility in the interpretation of test scores. ASHA states that, while careful testing and consultation of norms is important, the diagnosis "must culminate in clinical judgment" (p. 32).

The ASHA position may be defensible if clinical judgment results in relatively consistent decisions across diagnosticians. The following sections explore (1) the data available to assist speech-language professionals in making diagnostic decisions and (2) the decision-consistency which results from application of that data in conjunction with clinical judgment.

There are two basic types of information available to assist in the SI diagnosis; i.e., normative data and official criteria. Issues related to both of the sources are reviewed below.

Normative data. Assessment of language skills was earlier described as a very complex process due to the all-encompassing nature of the term "language". No one set of norms for language development could reflect all possible language skills in the many different forms in which they are tested. Instead, normative data for receptive and expressive language development are generally provided by developers of specific tests. This allows comparison with a norm group for a specific skill in the particular manner in which it is being measured by that test. Separate norms thus exist for a variety of language skills including receptive vocabulary, language comprehension, elicited language, environmental language, pre-language skills, language facility, spontaneous and structured language sampling, auditory discrimination, auditory memory, selective attention, auditory closure syntactic skills, visual perception and visual sequential memory (Darley, 1979). Unfortunately, as these instruments have often used different scoring systems, measured a slightly different aspect of one skill or used different age or other types of norming groups, test data are generally not comparable across tests of the same skill.

Assessment of articulation skills was shown earlier to be a more universal process. Although some articulation tests present normative data (Fudala, 1963), test performance is generally compared with norms not specifically developed for the test (Arlt & Goodban, 1976). The use of articulation norms is particularly crucial for preschool children. While older children may be expected to produce all sounds correctly, younger children exhibit a wide range of articulatory "errors" which are considered to be part of normal development (Longhurst, 1977).

The classical approach to the development of articulation norms has been to provide developmental ages at which specific sounds are produced correctly by a given percentage of children who purportedly are selected from a normal population. Until recently, three studies (Poole, 1934; Templin, 1957; Wellman et al., 1931) have served as the primary normative references for comparing articulation performance. Wellman et al. published the first data concerning articulation development in the U.S. Their sample included 240 children between the ages of three and six. Phoneme production was measured by 75% correct production of sounds in initial, medial and final positions in spontaneous responses to questions or stimulus pictures. The representativeness of this sample has been challenged as the subjects were all drawn from the same geographic region and high socioeconomic group (Winitz, 1969).

In 1934 Poole developed a similar set of norms from a study of 65 children between the ages of two years, six months and eight years, six months. These children also lived in the same area and came from upper socioeconomic backgrounds. Poole used a more stringent criterion of correct production by the entire sample. She tested production of 23 consonant phonemes in all three positions within a word. Most of the data were collected from spontaneous responses, although Poole has acknowledged that it was sometimes necessary to have the child repeat a response after the examiner.

Templin (1957) gathered developmental articulation data on 480 children ranging in age from three to nine years. The sample was carefully selected by sex, ages and father's occupation to form a group repesentative of the U.S. urban population. Poole included 176 sound elements in this study which also

used a 75% criterion as a measure of correct production. Spontaneous speech, elicited imitation and oral reading were used to obain responses.

Comparison of data from these three normative studies is hampered by the use of different phonemes, different response modes, different criterion levels, varying numbers of phonemes tested and differences in both the size and demographics of the sampled populations. Nevertheless, Templin (1957) found considerable agreement in the age level assignments of sounds among the three studies. Winitz (1969), however, noted that the largest discrepancies occurred for the sounds which are among the most frequently misarticulated by young children. Sander (1972) has also criticized these classical studies of phoneme mastery: as "relecting upper age limits rather than average performance" (p. 56). He has suggested that a less stringent and more meaningful way to describe phoneme acquisiton is in terms of ranges of customary production. Using this method, Sander defined the earliest age at which a phoneme is correctly produced in two positions by 51 percent of the children and extended the range to that age at which 90% produce it correctly. Darley (1979) has praised this approach becuase it allows for the variability of phoneme development in young children. When Sander's method was used to analyze the Wellman et al. and Templin data, far greater variability in the acquisition of consonant phonemes was found than was evident in Templin's comparison (Sander, 1972).

Prather, Hedrick, and Kern (1975) analyzed articulation acquisiton data from a normative study of 174 preschool children. Phoneme production from spontaneous speech and imitation was assessed in the initial and final positions using the 75% criterion. The sample included white children from all

socioeconomic classes. The researchers found that their sample produced phonemes at consistently earlier age levels than those noted in the earlier studies. In a similar study Arlt and Goodban (1976) found that 43% of the 79 sound elements tested were produced correctly from six months to four and one-half years earlier than would be expected from the classical normative data. The authors cite a variety of factors including television, earlier schooling and new teaching methods which may have contributed to earlier acquisition of articulatory skills and a subsequent need to update articulation acquisition norms. Darley (1979), after reviewing these recent developments, also concluded that "it is evident that we need more data on phoneme development of young children" (p. 229).

The validity, even of the most recently developed norms, has been challenged by several other researchers. Some have argued that separate male and female norms should be developed as girls often experience a far more rapid increase in speech and language skills begining at about four years of age (Koenigskneecht & Friedman, 1976; Nelson & Bonvillan, 1978). Others argue that, while some of the norms have included some ethnic minorities, no systematic sampling of phoneme acquisition in nonstandard English speakers has been conducted (Stockman & Vaughn-Cooke, 1982). Although the results are not yet available, a large scale longitudinal and cross-sectional investigation on the acquisiton of black English is being conducted at the Center for Applied Linguistics in Washington, D.C. This study will also examine sex as a possible variable in accounting for variation among black speakers.

The analysis of normative data thus appears to be in a developmental stage, at least for articulation skills. Language norms are difficult to compare across tests and generally do not represent all of the skills included under the umbrellas term "language". In the absence of a single, universally accepted source of criteria for evaluating SI, many federal, state and local agencies have developed their own standards. These official crtieria are described in the following section.

Official criteria. As outlined in Part I of this review, P.L. 94-142 defines each handicaping condition in general terms and then requires each state and its local agencies to develop further eligibility criteria and assessment procedures. Although there have been several national evaluations of LD, EMR and ED definitions (see Part I), there appears to be no similar review of SI criteria. Given the state of the art in SI diagnosis as reviewed above, the inconsistencies and lack of operationalized definitions found in other handicap definitions might also be expected for the SI category. Because no overall evaluations of SI criteria are yet available the following section includes an in-depth review of state and local definitions and criteria.

The federal government defines SI in P.L. 94-142 (1975) as "a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, which adversely affects a child's educational performance" (U. S. Congress, 1975). Project Head Start offers a slightly more specific definition which builds upon the first one:

> A child shall be reported as speech impaired with such identifiable disorders as receptive and/or expressive language impairment, stuttering, chronic voice disorders, and serious articulation problems affecting social, emotional, and/or educational

> achievement; and speech and language disorders accompanying conditions of hearing loss, cleft palate, cerebral palsy, mental retardation, emotional disturbance, multiple handicapping conditions and sensory and other health impairments. This category excludes conditions of a transitional nature consequent to the early developmental processes of the child (Administration for Children, Youth, and Families, 1975).

The American Speech, Language, and Hearing Association (ASHA) developed universal definitions for communications disorders, but then advised that they "are not intended to address issues of eligibility and compensation" (1982, p. 949). The ASHA definition includes an overall delineation of communicative disorders and then treats each type of speech and language disorder separately. It also includes the hearing impaired. ASHA defines a communication disorder as:

> an impairment in the ability to (1) receive and/or process a symbol system, (2) represent concepts or symbol systems, and/or (3) transmit and use symbol systems, The impairment is observed in disorders of hearing, language, and/or speech processes. A communicative disorder may range in severity from mild ro profound. It may be developmental or acquired, and individuals may demonstrate one or any combination of the three aspects of communicative disorders. The communicative disorder may result in a primary handicapping condition or it may be secondary to other handicapping conditions (p.949).

These are the three basic SI definitions which have been developed for national use. The federal and ASHA definitions do little more than describe some basic types of speech problems. The ACYF definition was developed for a specific population -preschool children. It suggests only two general criteria: (1) the SI must affect achievement (the federal definition contains this provision also) and (2) excludes developmental disorders which the child will "grow out of". While these are useful concepts, more specific guidelines are needed for those working with preschool children who, as explained earlier,

may exhibit a wide range of errors and still be considered "within normal limits". The ACYF admonition that articulation errors be "serious" is useless without some operational definition of this term. Clearly, the responsibility for establishing usable definitions and criteria rests with state and local education agencies.

A review of state SI criteria reveals a wide range of specificity and expansion of the federal regulations. Some states have opted merely to reiterate the federal criteria (Oklahoma State Department of Education, 1979), while others have made very minor changes. One SEA added the provision that a screening and an I.E.P. would have to indicate a speech problem (Pennsylvania Department of Education, 1980). There is no mention in these state guidelines of any need for a diagnostic evaluation.

A very common approach among SEA's has been to more thoroughly define a speech disorder by listing and describing each of the 4 speech and language areas, but still offer no criteria. Among the briefest descriptions are those offered by North Carolina (State Dept. of Public Instruction, 1979):

> Children may be identified as needing speech-language evaluations through mass screening efforts and/or teacher referral. Children determined through screening or referral to need evaluations shall be assessed in the areas of phonology (speech sounds), language (syntax, morphology, semantics), voice and fluency. It is on the basis of such an evaluation that the determination as to the type and itensity of services is to be made.

Other states such as Rhode Island use only slightly more specific guidelines (Bd. of Regents, 1978). Articulation disorders, for example, are described as defective production of phonemes (speech sounds). Types of misarticulations include: substitution of one phoneme for another, omission of

phonemes in words, phonemic distortions, and inappropriate additions of phonemes (p. 43)."

Some SEA's have added the stipulation that the articulation impairment must interfere with the inteliigibility of speech (Kansas State Department of Education, 1980) or be inappropriate for the child's age, sex or sociocultural background (Washington Department of Public Instruction, 1980; West Virgnia Department of Education, 1981). The Alaska Department of Education (Brown, 1978) stipulates only that the impairment be "significant," although this term is not operationalized.

Other state criteria are not based on the nature of the child's disorder, but on the capacity of the system to work with the child. Alabama (Department of Education, 1980) leaves all decisions concerning criteria to the discretion of the individual diagnostican, but stipulates the acceptable number of students who may be served at each SI severity level from mild to severe. The Georgia Department of Education (1980) uses the same caseload approach, but includes broad definitions of severity level. A severe SI, for example, is described as a student who "cannot be understood by and/or understand most persons without great difficulty; significantly prevents student from using verbal expression as a communication avenue (p. 33)."

Some states provide lists of recommended or required tests or types of test to be used in diagnosing SI. These guidelines may range from general comments about the usefulness of obtaining a developmental and social history (Florida Department of Education, 1980) to a required list of nine assessment measures (Alabama). Other states focus only on one aspect of the speech and language evaluation. Michigan (State Board of Education, 1980), for example,

requires the administration of a spontaneous language sample and at least two standardized language test or subtests. Requirements that tests be standardized or that published normative data be used are not included in the majority of state regulations. Some states require that such measures be used in general (e.g., Florida), that they be used for specific age groups (e.g., Ohio Department of Education, 1980 requires normative measures for children less than 8 years old) or for testing specific types of SI (e.g., Washington requires standardized measures for articulation and language only).

There have been some attempts at the state level to provide operationalized criteria for SI eligibility. In this area, too, there is a wide variation among states. Nebraska (State Board of Education, 1979) provides a example of very broadly defined criteria which include other handicaps in addition to SI:

> The developmental delay resulting from mental retardation, speech and communication disorders and specific learning disabilities as assessed on a standardized wide range developmental instrument are two or more standard deviations below the norm or if any handicapping conditions can be expected to produce such delay in later childhood (p. 54-10).

Other state-level criteria focus specifically on the SI diagnosis and on types of speech disorders. Washington State, for example, defines an articulation disorder as follows:

> Achieves a rating of moderate or severe on a standardized articulation test that yields a severity rating and/or misarticulates three or more unrelated phonemes for students up to age eight, one or more for students over age eight, with consideration given to the student's speech intelligibility, physical ability, and/or therapy history (Department of Public Instruction, 1980, p. 36).

Other states have focused on criteria not only for each type of SI, but also for severity levels within each type. Louisiana (Board of Education, 1980) guidelines for articulation include separate criteria for mild, moderate, severe and profound/multiple impairment. The criteria for a moderately severe disorder, for example, is defined as "three or more misarticulations of phonemes, still not interfering with intelligibility" (p. 73).

School districts are required to review their state's annual program plans and then submit their own assessment procedures based on the state regulations. District-level definitions and criteria for SI thus have an even greater potential for variation than those developed at the state level. Many districts such as Allegheny County in Pennsylvania (Allegheny Co. P.S., 1980), Guntersville District in Alabama (Guntersville Bd. of Ed., 1980), and Bibb County in Georgia (Bibb Co. Bd. of Ed., 1979) simply restate the SEA guidelines with no further clarfication. Oklahoma City (Ed. Dept., 1980) also copies the state SI definition, which is only a copy of the limited federal guidelines.

Another local district approach is to modify only part of the state procedures. The Georgia Department of Education (1978), for example, has defined each type of SI and required that standardized tests be used in the diagnosis. Decisions about severity rating criteria and caseload selection guidelines are left to the LEA's which have adopted a variety of criteria. Gwinnet County Schools define a severe language disorder in specific measurable terms: "The child's language is at least three years delayed in expressive and/or receptive language. The child's language is disordered; three standard deviations and/or lower five percentile in two or more areas of

language competencies" (Georgia Dept. of Ed., 1978, p. 6). The Muscogee County Schools interpreted the state guidelines differently as they chose to develop only one definition for a severe SI regardless of whether the problem was articulation, language, voice or fluency. This district defines severe SI as follows:

1. Student cannot be understood by and/or understand most persons without difficulty.
2. Communicative ability is inadequate when compared with the student's overall ability level.
3. Communication difference are other than dialectal differences.
4. Prognostic variables indicate student should be classified as "severe" (Georgia Dept. of Ed., 1978, p. 6).

The severity rating scales developed by the Metro Atlanta Speech Language Consortium (Georgia Dept. of Ed., 1979) specify only measures of articulation. Severe articulation, for example, is defined as consistently unintelligible speech which is distracting to all listeners. An additional measure may include unwillingness on the student's part to initiate communication. Some other districts have found it necessary to develop specific criteria for some handicaps while adopting state guidelines verbatim for SI and other handicap categories. The Half Hollow Hills district in New York state, for example, has modified only the state definition for LD (Half Hallow Hills Dept. of Sp. Ed., 1979).

It is generally only the largest and/or more wealthy districts which have developed their own complex and comprehensive special education plans. Dade County Public Schools (1979) requires a 300 page bound volume to

describe handicap evaluation and placement procedures. This district's SI eligibility criteria are outlined in detail for each type of speech or language disorder. This includes screening procedures, suggested diagnostic measures and re-evaluation guidelines. Criteria are specific and operationalized. An articulation disorder, for example, is defined as "one or more phonemes misarticulated which are not commensurate with expected developmental age norms" (p. 42). Recognizing the inconsistency of articulation norms, Dade County provides their own "phonological acquisition chart" based on the work of Sanders (1972) and Arlt and Goodban (1976).

The specific, thorough guidelines offered in Dade County appear to be the exception, not the rule. Fresno Unified School District (1981) in California, for example, provides an equally long, comprehensive plan but does not provide the same level of specificity or operationalized definitions. "Phonology criteria" include (1) "phonemic errors at least one year below develomental expectation and limits communication, and (2) reduced intelligibility" (p. 100). Normative data and measures of intelligibility are not offered.

Official criteria for SI vary widely, just as those for other types of special handicaps. The range and specificity of definitions is broad and includes states and districts with no elaboration beyond the very limited federal guidelines to those with fairly specific operationalized procedures for each type of communication disorder at each severity level. The more specific guidelines, however, appear to be based primarily on normative data which were shown earlier to be inconsistent. The overall impression of SI

criteria guidelines throughout the United States is one of extreme variability and inconsistency.

Criteria used in the diagnosis of SI children. After reviewing the plethora of inconsistent and contradictory normative data and other criteria available for diagnosing SI chidren, Stewart and Weybright (1980) attempted to find out what diagnosticians are actually using. They focused particularly on the diagnosis of articulation impairment. Their survey of 145 speech-language pathologists in Oregon revealed that the majority refer to the Poole (1934) or Templin (1957) norms rather than the more recent analyses furnished by Prather et al. (1975) or Arlt and Goodban (1976). Diagnosticians who did not rely on these classical studies reported using 20 other sources of normative information.

It appears that inconsistent sources of normative data and use of outdated information are not the only sources of inconsistency in SI diagnosis. There is some evidence that norms and official criteria are often not used at all. Over 15% of those in the Oregon sample did not use any norms but relied on their own internal criteria. One diagnostician, for example, reported using "common sense" rather than norms. Another respondent indicated that "motherhood" qualified her to make her own judgments. A similar study conducted by the Indiana Speech and Hearing Association revealed that 50% of the 276 speech-language diagnosticians sampled did not rely on any formal, written criteria for case selection or for special education enrollments (Anderson, 1982). Project Head Start conducted a national survey of criteria used by their diagnostic providers the majority of which were speech-language diagnosticians (Randolph, 1977). The survey revealed that almost all of the

respondents had never heard of or seen the officially legislated categories or diagnostic criteria provided by the federal or state agencies.

A number of studies have explored the effects of using internal criteria or a variety of guidelines in assessing children who may have speech impairments. In an early study Henderson (1938) found that diagnostician's judgments of articulation errors were fairly consistent if based on a correct/incorrect dichotomy, but were far less consistent when identifying specific types of errors. In a more recent study Shreiberg (1972) found significant differences in interjudge agreement concerning articulatory characteristics in each of four children. The diagnosticians in the study relied on their own judgment. Norris, Harden, and Bell (1980) found similar diagreements among four graduate students in speech pathology as to place and manner of articulation for 97 children. In three separate paired comparison studies Young (1981) asked 81 speech-language diagnosticians to rate and re-rate the articulation effort of several different speakers. Relying on their own criteria, the observors demonstrated very low rates of agreement with themselves or with each other.

Studies of speech disorders other than articulation generally result in the same internal inconsistencies. MacDonald and Martin (1973) and Young (1975) found that interobservor agreement for identifying stuttering was as low as 50% when diagnosticians used their own guidelines. Wertz and Mead (1975) observed speech clinician and teacher severity ratings of children with four types of speech disorders. They found that classroom teachers displayed significantly more agreement among themselves than did the speech professionals. Allen and associates (1981) examined the diagnosis of 3 year old

children with language disorders. They compared the clinical judgments of three experienced speech diagnosticians with the diagnostic decisions resulting from three language tests. The found "substantial disagreements between test decisions and clinical judgment: (p. 68).

> It appears that the SI diagnosis suffers as much from a lack of consistent, operationalized definitions as do other special education categories such as LD, EMR and ED which have received so much legislative and educational attention. There is much evidence that a reliance on clinical judgment, the approach advocated by ASHA, results in inconsistent diagnostic practices.

Diagnostician background and training. Another non-measurement issue which may be a potential source of inconsistency and bias in the SI diagnosis is the experience and education of the diagnostician. This was shown earlier to be a factor in the special education diagnosis for all types of diagnosticians, including speech-language professionals. There have been few studies, however, which have concentrated specifically on the speech-language diagnostician. Results of the research which has been conducted offer conflicting conclusions. Studies by Fristoe and Goldman (1968) and Irwin and Musselman (1962) indicate that there is little difference in the effectiveness of experienced vs. inexperienced listeners in judging articulation errors. Schissel and Flourny (1978), however, reached a different conclusion. They compared the consistency of articulation error judgments of inexperienced graduate students in speech pathology with those of certified, practicing speech-language clinicians. There were significant differences in variances between the two groups with higher consistency for the more experienced clinicians.

Other observers have criticized the content rather than the length of diagnostician training or experience. Wolfram (1979) advised that speech-language professionals should be expected to learn about the basic structures of nonstandard dialects encountered in their area. They should know the actual sounds or structures which compose the dialect and precisely how they are used. The content of training may be affected by the year(s) in which the diagnostician attended school. As discussed earlier, the study of language and of nonstandard dialects is a relatively recent discipline. Diagnosticians who received all of their training before the 1970's may be less likely to have been exposed to these new developments in speech and language assessment. Muma, Webb, and Muma (1979) note that "only a decade ago, little atention was given to language-in-training programs" (p. 467). Their recent survey of 84 speech-language pathology programs throughout the United States indicated that language is now included in all graduate programs. It is less likely, however to be required for undergraduate students.

A review of state certification requirements for speech/language professionals indicates that these regulations are determined at the state level, with little or no district-level intervention. The majority of SEA regulations for speech/language clinicians are based on general certification in special education with an endorsement in speech pathology (e.g., Alabama Dept. of Ed., 1980; Dept. of Public Instruction of Wash. State, 1980; Florida Dept. of Ed., 1980). Georgia (Dept. of Ed., 1978) has developed specific requirements for related disciplines of LD and audiology but has only very general guidelines for speech personnel. Only a few states require a master's degree or the equivalent (Michigan State Board of Ed., 1980; Rhode Island Bd.

of Regents, 1978). Nebraska (Campbell, 1979) requires some coursework and practicum beyond the bachelor's degree which is equivalent to approximately six months of graduate work. Some of the SEA's require that all teachers, including speech therapists, maintain their certification by taking some coursework at prescribed intervals (e.g., Rhode Is. Bd. of Regents, 1978; West Virginia Dept. of Ed., 1980). Yet none of the states require that these courses be taken in the teaching endorsement area.

Taylor (1980) reviewed state certification standards from 46 states. She found no uniformity in any aspects of the certification requirements. Nineteen different professional titles and fourteen different certificate designations were used by the various states.

The stringency of certification requirements also varied greatly. Only twelve states met ASHA practicum requirements in the speech-language pathology area. One state appeared to have no standards for speech-language staff. Academic requirements also showed great variability. Sone states required 1 or 2 courses, others required up to 36 semester hours and 2 states specified no particular academic standards. The majority of states simply required that the college or university's approved program be completed.

Speech-language personnel may have very different training and experience as evidenced by certification requirements. There is some evidence that differences in level of training, recency of training or years of experience may influence the diagnostician's decisions about the SI child. Information conerning this potential source of non-measurement bias, however, is insufficient to draw definite conclusions.

Summary

The fact of discrimination in special education diagnosis cannot be denied. It has been well documented that students with particular age, sex, ethnic, linguistic and behavioral characteristics are overrepresented in special education enrollments. Both measurement and non-measurement sources of these biases have been identified. Bias resulting from non-test issues is of particular concern as so much effort already has been concentrated on test reform with questionable results. The major non-measurement factors identified in the literature were related to (1) the use of diagnostic criteria and (2) the background of the diagnostician.

Discrimination in the diagnosis of SI has not been examined as closely in the literature as have other types of handicaps. The potential for bias in speech assessment appears to be as great as for other well-researched handicaps for which discrimination has been well documented. Non-measurement factors were shown to be an important consideration in the SI diagnosis. The use of inconsistent or vague criteria, for example, is thought to be a particular problem in the diagnosis of preschool children. The diagnostician's educational training and recency of training may also serve as non-measurement factors in diagnostic bias. Although these non-measurement variables have been linked wth discrimination in special education assessment, there have been no investigations of their relationship to speech assessment. In fact, there is little information concerning criteria used in the SI diagnosis or factors related to the speech diagnostician's background.

CHAPTER III

METHODOLOGY

The investigation utilized a secondary data analysis from a non-experimental survey design. Data collected from a subset of the original sample were used to investigate variables not considered in the original research. Data collection procedures from the initial study as well as modifications for the present investigation are described in this section.

Population and Sample

The population consisted of all Head Start children in the 1980-1981 school year diagnosed as having a primary handicap of speech impairment with an articulation problem. The articulation problems ranged from mild to severe. A multi-stage procedure was used to select an appropriate sample from this population. It consisted of the following steps:

1. The desired sample size for the survey was set by the Administration for Children, Youth, and Families at approximately 500 Head Start SI children drawn from an estimated universe of 21,988 such children enrolled during the 1980-1981 schol year (U.S. Dept. of Health and Human Services, 1980).
2. The average number of SI children enrolled in individual Head Start programs throughout the United States was determined through an analysis of the 1980 Annual Head Start Handicapped Survey. It was

determined that, if 8 SI children were to be selected from a program, then a total of approximately 69 programs would need to be selected to obtain a sample of 500 children, as many of the smaller programs have less than 8 SI children.

3. A total of 69 Head Start programs were selected from the 1,750 programs throughout the United States and Puerto Rico, using Probabilities Proportional to Size (PPS) sampling techniques. This methodology was employed to ensure representation in the sample of children from very large and very small programs. This list was ordered by the MOS (Measure of Size; i.e., the number of Speech Impaired children in the program). The initial sample element was selected using a random start. Sequential selections were made by summing the MOS for subsequent programs on the list until the interval zone size (total number of SI Head Start children/500) was reached. Each time a series of summations reached the interval zone size, a program was selected and the summation was restarted.

4. Each Head Start program administrator provided a list of all children in their program who had been diagnosed as having a primary handicap of speech impairment. Systematic random sampling was used to select 8 children from each program. For programs with less than 8 such children, a saturation sample of all children fitting this description was included. The final sample consisted of 497 Head Start children.

5. A subsample of all children with articulation disorders was selected from the 497 subjects. This articulation subsample included 393 subjects. A total of 48 children were later dropped from the subsample because (1) they were untestable (N=20) or (2) their records had been lost or destroyed (N=25), or (3) they spoke only Spanish and could not be evaluated for articulation error scores on the same basis as English speakers (N=3). The final sample selected for the present investigation included 345 children. Due to missing data some analyses were conducted on only 311 of the subjects as explained in a later discussion of the statistical analysis procedures.

Instrumentation

Interviews

Each Head Start program provided the names of the speech-language diagnostician who had evaluated each child in the sample. (Some of the diagnosticians had evaluated more than one child in the sample.) A total of 130 diagnosticians were thus identified. Structured Computer Assisted Telephone Interviews (CATI) were used to collect specific information from these diagnosticians about the children in the sample and about the background and experience of the diagnosticians.

The CATI consisted of a series of interview questons and possible response choices entered into a microcomputer. Prior to each interview the CATI program was loaded into the computer and displayed on a screen, one

question at a time. The interviewer then contacted the diagnostician at a pre-arranged time and initiated the interview. When the respondent answered a question the interviewer entered the code for that response. The response was immediately evaluated by the computer program. If it contained an out of range value (e.g., indicating that the respondent had more than 50 years of diagnostic experience) a message advising the interviewer to repeat the question appeared on the screen. If the response fell within the accepted range the computer evaluated the answer to determine the next appropriate question. If a diagnostician indicated, for example, that she had pertinent experience prior to her present position then three questions related to this experience were included. For diagnosticians who had no previous experience, these questions were omitted. Thus CATI allowed for immediate data entry and editing and for more efficient conduct of interviews.

Two types of CATI's were used in this study; i.e., the Diagnostician General Interview (DGI) and the Child Specific Interview (CSI). Each diagnostician responded to one DGI. This interview requested specific information about the diagnostician's speech-language degree(s), date(s) of graduation, years of experience, demographic characteristics, etc. A copy of the DGI questions is presented in Appendix I. One CSI was administered for every child in the sample. Diagnosticians who had diagnosed four children in the sample, for example, provided data for four separate CSI's. The CSI requested specific information about the severity of the child's speech problem, the criteria used to diagnose the child, child demographics, etc. A copy of the CSI questions is also included in Appendix I.

Case File Review

Diagnostic case files were provided by individual Head Start programs for each of the sampled children. Case files supplied demographic data, articulation and language severity ratings, names of articulation tests used and specific articulation errors. The percentage of types of sounds in error (AES1) and the percentage of total articulation errors (AES2) were calculated from the results of tests in the case files. The following formulae were used:

AES1 = (# sound errors/# sounds tested) x 100

AES2 = (# errors/total # observations) x 100.

The sample case file review form, which is included in Appendix I, illustrates how articulation errors were recorded.

Statistical Analysis

The statistical analysis consisted of three basic procedures, which were conducted through the use of the Statistical Package for the Social Sciences (SPSS) programs (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). First, frequency counts and crosstabulations were calculated for all of the independent variables in order to (1) provide a description of diagnostic criteria (Hypothesis 1) and diagnostician background (Hypothesis 2), (2) provide a description of the SI sample, and (3) assist in selection of predictor variables for additional analysis. Secondly, scatterplots and correlations among variables were computed to test for possible violations of assumptions associated with significance testing in multiple regression. Finally, multiple regression equations were computed to test Hypotheses 3 and 4. The regression analyses were conducted twice for the same sets of predictor

variables in order to analyze their prediction first of AES1 and then of AES2. Each of the three statistical procedures are described in this section.

Selection of independent variables

Data were collected for a total of 18 independent variables, classified in the following general categories: diagnostic criteria, diagnostician's background, and child characteristics. Exhibit 2 presents a listing of variables within each category and indicates how categories for some of those variables were modified. Appendix II presents frequency counts for the original categories of each variable. Recoding was used for some variables to eliminate low-frequency categories. Some other variables were dropped from further analysis due to very small sample sizes within secondary categories. These variables, which are indicated in Exhibit 2, include the sex and ethnicity of the diagnostician and the child's additional handicapping conditions.

Missing data were found only for the language severity variable. Diagnosticians did not appraise the severity of the language impairment for 34 of the 345 children. All analyses related to language impairment were thus conducted with a sample size of 311 SI children.

Analysis of correlations among variables resulted in the elimination of two additional variables to avoid problems associated with extreme multicollinearity, as explained in the following discussion of multiple regression assumptions. A total of 12 predictor variables was thus selected from the original 17 for the regression analysis. These variables are indicated by an asterisk in Exhibit 2.

Exhibit 2

Recoded Independent Variables

Variables	Categories/Ranges of Values	
CHILD CHARACTERISTICS		
* Age	Range: 34-76 months	
* Sex	1=Male	2=Female
* Ethnicity	1=Black	2=Non-black
Primary language or dialect	1=Black dialect	2=Standard Eng or other dialct.
Other handicaps	1=Hearing	2=Physical
	3=Health	4=EMR
	5=ED	6=None
* Appraisal of language severity	Range: 1-6	
	1=Normal	6=Severe
DIAGNOSTICIAN BACKGROUND		
Yrs. of experience	Range: 1-23 years	
* Highest earned speech degree	1=B.A. or less	2=M.A. or higher
* Certification type	1=CCC or priv. prac	2=Other
Sex	1=Male	2=Female
Ethnicity	1=Black	2=White
* Recency of training	Range: 1954-1982	
* Awareness of Head Start diagnostic criteria	1=No	2=Yes
DIAGNOSTIC CRITERIA		
* Sound errors (AES1)	Range: 4-95%	
* Total errors (AES2)	Range: 3-96%	
* Articulation test used	1=Goldman Fristoe	2=Arizona APS
	3=Photo Artic	4=Other
* Behavior problem	1=No	2=Yes
Other criteria	1=Family factors	2=Other factors
	3=No other criteria considered	

Assumptions of multiple regression significance tests

The significance tests associated with multiple regression are based on several assumptions, including the following:

1. The sample is drawn at random;
2. X is measured without error;
3. The distribution of Y values is normal;
4. The regression of Y on X's is linear;
5. The variance of errors at all levels of X is constant (homoscedasticity); and
6. None of the independent variables are highly correlated (Cohen & Cohen, 1976; Nie et al., 1975; Pedhazur, 1982).

The assumption of normal distribution of the Y values was relaxed due to the large sample size. Pedhazur (1982) notes that several sources have demonstrated that regression analysis generally remains robust when normality assumptions are violated. Cohen and Cohen (1976) note that this is particularly true for large sample sizes.

The assumptions of linearity of relationships and homogeneity of variance were evaluated through a direct examination of a scatterplot of the residuals. The residuals or deviation of observed scores on the criterion variable (articulation severity rating) with estimated scores were derived from the standard regression procedure, which is described in a later section. The scatterplot of residuals resulted in a generally straight band pattern with no curvilinear trends. This indicates relative freedom from abnormalities in linearity or homoscedasticity. There were no outliers, indicating that there were no deviant articulation severity scores.

Correlation coefficients were examined for all possible pairs of predictor variables as a test of multicollinearity. (See chapter 4 for a table of these values for all variables which were selected for regression analysis.) Three pairs of predictor variables were found to be moderately to highly correlated; i.e.,

(1) Diagnostician's years of experience and year of receiving highest degree ($r = -.70$; $p < .001$)

(2) Child's primary language/dialect with child's ethnicity ($r = -.71$; $p < .001$)

(3) Total percentage of articulation errors (AES2) with percentage of sounds in error (AES1) ($r = .86$; $p < .001$.)

Some statisticians (Darlington, 1968; Gordon, 1968; Nie et al., 1975) have suggested that, in such cases of extreme multicollinearity, either a composite variable be created or that only one of the variables in the correlated set be used to represent the common underlying dimension. Composite variables were not used as they would not have assisted in testing the hypothesis. Instead a single variable from each pair was selected. The year of receiving highest degree was chosen over years of experience because the first variable received much greater support in the literature as a possible predictor (Fristoe & Goldman, 1968; Irwin & Musselman, 1962; Muma, Webb, & Muma, 1979). The child's ethnicity was selected, rather than the child's primary language/dialect because of the relatively larger sample sizes in ethnic categories than in language categories.

The two measures of articulation errors were both retained but were used in separate analyses. Regression equations were developed using the same sets of predictor variables, first for AES1 and then for AES2.

Multiple Regression Procedures

The ordering of predictor variables within the equation was carefully considered. Unrestricted stepwise procedures were rejected for several reasons. First, tests of significance based on regression weights from stepwise analysis have been criticized for having little statistical validity (Finn & Mattsson, 1978). Secondly, 6 of the 18 predictor variables were selected in the third hypothesis as being the best potential predictors of the articulation severity rating because of strong support for these variables in the literature. By ignoring this information and basing the ordering of predictor variables on correlation alone, stepwise regression may decrease the power of the analysis. Finally, a variable may appear unimportant in stepwise regression not because it has a low correlation with the criterion, but because it is highly intercorrelated with other predictors already in the model.

Hierarchical blockwise ordering of variables is the recommended procedure if the correlation among the independent variables is considered to be causal or if the contribution of a subset of variables is of interest (Cohen & Cohen, 1976; Nie et al., 1975). This procedure was selected as a test of the third hypothesis because it could provide relevant information about the three subsets of predictor variables; that is, diagnostic criteria, diagnostician background, and child characteristics. No causal ordering of the predictor variables could be justified through the literature.

Simultaneous regression analysis enters all the predictor variables together. The independent variables are simultaneously correlated with the dependent variable. This procedure considers the observed correlations among variables to be noncausal and was considered most appropriate as the primary

test of the fourth hypothesis. Simultaneous regression procedures were planned to be used with (1) the interaction terms specified in Hypothesis 4, (2) the predictor variables found to explain the most variance in the investigation of the third hypothesis, and (3) interactions found to be of possible significance during the statistical analysis.

A set of dummy variables was created for each category of the nominal variables. The articulation test variable was recoded as (1) Goldman Fristoe and (2) all other articulation instruments. All other dummy variables were based on the codes presented earlier in Exhibit 2. The dummy variables were used in both the hierarchical and standard regression analyses.

Summary

Computer Assisted Telephone Interviews (CATI) and case file review were used to collect information about 345 SI children[1] enrolled in Head Start and the 130 speech-language professionals who diagnosed these children as speech impaired. Data were collected for a total of 18 variables which were summarized in Exhibit 2. Some of the variables were used to describe the sampled children and diagnosticians and the diagnostic criteria. Twelve of the variables were used in multiple regression analyses as potential predictors of the articulation severity rating, as discussed in Chapter 4.

[1]The mulitple regression analyses included only 311 SI children as language severity ratings were not available for 34 of the original 345 children sampled.

CHAPTER IV

RESULTS

This chapter summarizes the results of each analytic procedure. It begins with a description of the children and diagnosticians included in the sample. The results of the investigation for each of the four hypotheses are then presented. Finally, some observations of language diagnosis, which were not included in the original study plan, are discussed.

Description of the sample

The data were collected from a sample of 130 diagnosticians who provided information about a total of 345 articulation impaired Head Start children. Language severity ratings were not available for 34 of the children sampled. Thus, the total sample size for analyses involving language severity was limited to 311 children. A summary description of variables related to the children and diagnosticians included in the sample is presented below. Appendix II provides specific frequency counts for each variable.

Child sample

There were twice as many male (n=230) as female (n=115) SI children. This ratio is similar to data reported previously for Head Start (U.S. Dept. of Health & Human Svcs., 1980) and for SI children thrughout the United States (U.S. Dept. of Ed., 1930). The children ranged in age from 34-76 months. While Head Start typically serves children from 34-69 months, the diagnosticians explained that some handicapped children are retained in the

Head Start speech-language program when other appropriate placements are not available. The majority (97%) of the children were in the 3 to 5 year age bracket, which is typical of the Head Start population.

The children represented four of the five federally defined ethnic groups. No Asian/Pacific Islanders appeared in the sample. Three of the Hispanic children originally included in the sample were omitted, as explained earlier, because they spoke no English and their articulation error scores could not be compared with English speakers. The final sample included 66.4% white, 27.2% black, 4.1% American Indian, 2.3% Hispanic, and 0% Asian children. These figures are approximately representative of the Head Start populaton in general for the Asian and American Indian categories (U.S. Dept. of Health & Human Svcs., 1980). The Hispanic category was appropriately reflected in the original sample, but is significantly smaller here due to the lack of adequate case file records supplied for these children. The proportion of white and black children accurately reflects the ethnicity of the original sample. It differs, however, from the general Head Start population which includes approximately 42% black and 35% white children. The PPS sampling strategy used in the study resulted in the selection of proportionally smaller samples of students from larger programs. Smaller programs typically serve more white children, while the large urban programs serve a higher proportion of blacks. Since a total of eight children were selected from each program, regardless of size, a relatively higher proportion of whites is included in the sample.

Approximately half (60.3%) of the black children were reported to be black dialect speakers. While there are no national figures which might be used to compare the sample data, this proportion appears reasonable given the

representation of western and mid-western rural programs (typically non-black dialect speaking) as well as southern and urban (typically black dialect speaking).

Diagnostician sample

The sampled diagnosticians supplied information only for the children they had personally diagnosed. Most (64.5%) had evaluated only one or two children, although some (5.4%) had tested as many as eight children. Table 1 presents a breakdown of the number of sampled children assesed by each of the 130 diagnosticians. All of these respondents were speech-language professionals who had served in this capacity from 1 to 23 years, with an average of 8.3 years of experience.

Hypothesis 1: A wide range of objective and subjective criteria are used to rate the severity of articulation impairment in preschool children

Support for the first hypothesis of a wide range of diagnostic criteria was sought through descriptive statistics. As all of the diagnosticians in the sample indicated that they considered articulation scores as an important criteria in the SI diagnosis, the type of test used to measure articulation proficiency was also investigated. As indicated in Table 2, the Goldman Fristoe Test of Articulation (GFTA) was used to diagnose the majority of the SI children (55%). The Photo Articulation Test and the Arizona Articulation Proficiency Scale were used with 15% and 12% of the sample, respectively. A total of 12 other articulation tests accounted for the instruments used with the remaining 18% of the sample. These additional tests included informal,

Table 1
Number of Children in the Sample
Tested by Individual Diagnosticians*

Total No. of Children Per Diagnostician	No. of Diagnosticians	Percent of Total**
1	54	41.5
2	30	23.0
3	10	7.7
4	11	8.5
5	8	6.1
6	9	6.9
7	1	0.8
8	7	5.4

*No child was tested by more than one diangostician

**Total does not equal 100% due to rounding error

Table 2

Summary of Descriptive Statistics for

Diagnostic Criteria

Categorical Data	Categories	Frequency	Rel. Freq.*
Articulation test	Goldman Fristoe	189	54.8
	Photo Artic Test	52	15.1
	Arizona APS	42	12.2
	Other	62	18.0
Behavior problems	Hyperactive or agressive	50	14.5
	Not a problem	295	85.5
Other criteria	Family factors	55	15.9
	Other factors	49	14.2
	No other criteria	241	69.9

*Does not always add to 100% due to rounding error.

"teacher-made" devices used with 7% of the sampled children. Some of the articulation instruments listed were actually screening, not diagnostic devices, such as the Denver Articulation Screening Exam and the Fluharty Speech and Language Screening Test for Preschool Children. Approximately 3% of the sample was diagnosed as SI having received only a screening of articulation skills.

Diagnosticians were also asked to indicate factors other than test scores which influenced their diagnosis. The most commonly mentioned criteria were related to behavioral problems of hyperactivity or aggression. Some respondents used additional descriptors such as "acting out" or "disruptive". A total of 50 children (14.5%) were described as having behavioral problems of this nature which were considered in the SI diagnosis.

While the majority (70%) of diagnosticians indicated that they considered only the child's test scores, those who were influenced by other factors listed a wide variety of considerations. Some commonly considered factors centered around the child's family environment. These included concerns about child abuse, lack of verbal stimulation in the home, and speech impairments of other family members. Diagnosticians indicated that, for approximately 16% of the sample, they were influenced by one or more of these family issues in making their final diagnosis. An additional 14% of the diagnostic decisions were partially based on a wide variety of other factors which included the following:

- child's maturity or age
- withdrawn behavior
- no other services available for child's problems/handicaps
- child bothered by his/her speech problems

- parent or teacher concerned about child's speech
- physical coordination or appearance
- child's cultural background
- speech-language clinician's caseload
- stigma of handicap labels other than SI

No one of these concerns was listed as a consideration in more than 5% of the diagnoses.

These findings offer some support for the first hypothesis. While three formal, comprehensive, and standardized instruments accounted for the majority of articulation measures, approximately 18% of the children were evaluated very differently. They were administered other tests which varied widely in content, format, comprehensiveness, standardization, and difficulty. Additionally, the normative data provided by the three major instruments varies widely. The information from these measures thus provided different criteria on which to base the diagnostic decision. A variety of additional factors, unrelated to the child's articulation, influenced more than 30% of the diagnostic decisions. These factors varied widely and were sometimes totally unrelated to the child. In summary, it appears that the articulation diagnosis for the majority (70-80%) of the children was based on a similar set of criteria related to articulation proficiency. For the remaining group, however, test measures and other factors influencing diagnostic criteria varied widely.

Hypothesis 2: Speech-language diagnosticians have a wide range of training and experience.

Support for the second hypothesis of a wide range in diagnostician's backgrounds was sought through descriptive statistics of the staff's education,

certification, experience, and demographic characteristics. The 130 diagnosticians in the sample indicated that they had an average of 8.3 years of experience in the field of speech-language pathology. It was most common for children to be diagnosed by a professional with three years of experience (13%). Although diagnostician backgrounds ranged from 1 to 23 years, the vast majority of SI children (82%) were served by professionals with less than 13 years of experience. (See Table 3 for a summary of descriptive statistics for diagnostician background variables.)

The diagnosticians indicated that they had received their highest speech-language degree as early as 1954 and as late as 1982, although the majority (83%) had earned this degree after 1970. The average graduation date was approximately 1974. This measure of recency of training was found to be highly correlated with the diagnostician's years of experience; that is, respondents with earlier degrees tended to have more experience ($r=-.70$, $p < .001$). This indicates that speech-language professionals tend to seek graduate degree training early in their careers. Less than 3% of the children were diagnosed by staff who had earned no speech-language degree. The large majority of SI children (74%) were diagnosed by staff who held a master's degree in speech-language pathology.

Two thirds of the children were diagnosed by speech-language staff who had the professional certification necessary to enter into private practice; that is, they had a state license and/or the Certificate of Clinical Competence awarded by ASHA. The majority (92%) of those with state or ASHA certification held at least a master's degree. Less than one third of the staff were aware of the official Head Start criteria for the diagnosis of speech-language disorders.

Table 3
Summary of Descriptive Statistics for
Diagnostician Background Variables

Continuous Variables	Mean	S.D.	Range
Years of experience	8.3	5.1	1-23
Recency of training	1974	5.5	1954-82

Categorical Variables	Categories	Frequency	Rel. Freq.*
Highest speech language degree	No degree	9	2.6
	B.A.	65	18.8
	M.A.	255	73.9
	Ph.D	16	4.6
Certification	CCC +/or private practice	233	67.5
	Other	112	32.5
Sex	Male	36	10.4
	Female	309	89.6
Ethnicity	Black	21	6.1
	White	324	93.9
Aware of Head Start diagnostic criteria	No	225	65.2
	Yes	120	34.8

*Does not always add to 100% due to rounding error.

Almost all of the diagnosticians were white females. There were no Hispanic, Asian or Native American speech-language professionals in the sample. Only 6% of the SI children were served by black diagnosticians, and only 10% were diagnosed by males.

These findings offer some support for the second hypothesis. While their demographic characteristics were similar, the diagnosticians had diversified levels of experience in the field of speech-language pathology. While certification and training credentials fell within the same general categories, there were some important differences. There were 24 different combinations of certification types among the 130 staff, indicating a wide range of certification practices throughout the United States. The most recent degree training for these professionals varied by as much as 28 years. As discussed earlier, recent changes in the field of speech-language pathology indicate that there were significant differences in the content of training for those who earned degrees before the last decade.

Hypothesis 3a: Diagnostic criteria will be a better predictor of articulation severity ratings than will child characteristics or the diagnostician's background.

The third hypothesis of significant prediction of variance in the articulation severity rating was tested through the development of multiple regression equations. Regression procedures were conducted twice so that the two measures of articulation proficiency (AES1 and AES2) could be tested in separate equations.

Correlation Coefficients

The first step in the analysis was the calculation of correlation coefficients for all possible pairs of regression variables. These coefficients are presented in Table 4. (The coefficients for interaction terms are not included in this table as they will be treated later in the discussion of the fourth hypothesis.) AES1 and AES2 were highly correlated (r=.87) and were entered in separate analyses. Child age was moderately and negatively correlated with both AES1 (r=-.32) and AES2 (r=-.36); i.e., higher articulation error scores tended to be associated with younger children. The language severity rating was also moderately correlated with AES1 (.24) and with AES2 (.23). The diagnostician's degree was moderately correlated (r=.43) with type of certification, as diagnosticians with more advanced degrees tended to have private practice certification. The criterion variable, articulation severity rating, was most highly correlated with AES1 (r=.51) and AES2 (r=.55).

Identification of these correlations was potentially important for the regression procedure, since at each stage of heirarchical regression the effects of previously entered variables are removed from the effects of new predictor variables being entered at this next stage. Thus, if the child's age were entered first, the contribution attributed to AES would be reduced by the correlation of the two variables.

Variable Subsets

Hierarchical blockwise regression was used to test the predictive power of the three subsets of independent variables; i.e., diagnostic criteria, diagnostician background, and child characteristics. Diagnostic criteria

Table 4

Correlation Coefficients of Main Effects and Criterion

(N=311)

	AES1	AES2	TEST	BEHAV	AGE	SEX	ETHNIC	LSR	DEGREE	CERTIF	YEAR	CRIT
ASR	.5055	.5473	-.0271	-.0258	-.1646	.0601	-.0339	.0738	.0534	.0946	.0164	-.0432
AES1		.8660	.1296	-.0896	-.3211	.0479	.0410	.2443	.0234	.0290	.0727	.0135
AES2			.0112	-.1323	-.3590	.0304	-.0128	.2288	.0151	.0190	.0529	-.0366
TEST				.0057	.0741	.0466	.1750	.0713	-.0022	.0478	.2474	.0599
BEHAV					.0541	-.0627	.0054	-.1033	.0083	.0351	-.0294	.0415
AGE						-.0705	.0183	-.1342	.0822	.0399	.0425	.0312
SEX							-.0096	.0196	.0328	-.0497	.1603	.0715
ETHNIC								.1198	.1775	.0542	.1640	.2075
LSR									.1254	.2075	.0985	.0827
DEGREE										.4297	.2074	.0224
CERTIF											-.1728	.1894
YEAR												.0620

Abbreviations used

ASR = Articulation severity rating
BEHAV = Behavior problem considered as a factor in the diagnosis
LSR = Language severity rating
CERTIF = Diagnostician's certification
YEAR = Year diagnostician received degree
CRIT = Diagnostician's awareness of Head Start's criteria

$p < .05$ for all $r \geq .0896$
$p < .01$ for all $r \geq .1296$

included AES measures, the articulation test used, and whether the child's behavior was listed as important criteria. The highest speech degree, certification type, recency of training, and awareness of Head Start diagnostic criteria constituted the diagnostician background variables. The child characteristics included age, sex, ethnicity, and language severity. The entire regression analysis was conducted twice, first using AES1 in the diagnostic criteria block and second, replacing it with AES2. Table 5 lists the variables included in each subset and presents the results of the hierarchical analysis with AES1. Table 6 displays the same information for the second analysis; i.e., using AES2.

The results of the regression analysis of the three subsets are as might have been predicted from an examination of the bivariate correlations. The only block which explained a significant proportion of the variance (30%) was the group of diagnostic criteria variables; that is, the block containing AES1 or AES2. The regression of diagnostician background and child charachteristics were not significant at the .05 level. The null hypothesis, that R=0, was rejected for diagnostician criteria, but was accepted for the other two variable subsets .

Hypothesis 3b: The following single variables will be significant predictors of variance in the articulation severity rating: AES, diagnostician's highest educational degree, child's sex, child's ethnicity, behavioral problems, and language severity rating

Due to the large discrepancy in the predictive power of the three blocks of variables, further analysis concentrated on specific variables within the

Table 5
Hierarchical Regression with Three Variable Groups: Using AES1

Entry Order	Variable Subset	Cumulative Multiple R	Increase in R^2	F	Degrees of Freedom
1.	Diagnostic Criteria AES1 Behavior Artic. Test	.515	.265	36.84*	3,307
2.	Child Characteristics Age Sex Ethnicity Language Severity	.519	.005	15.98*	7,303
3.	Diagnostician Background Degree Certification Awareness of criteria Yr rcvd highest degree	.529	.001	10.57*	11,299

*p < .01

Table 6
Hierarchical Regression with the Three Variable Groups: Using AES2

Entry Order	Variable Subset	Cumulative Multiple R	Increase in R^2	F	Degrees of Freedom
1.	Diagnostic Criteria AES2 Behavior Artic. Test	.550	.303	44.46*	3,307
2.	Child Characteristics Age Sex Ethnicity Language Severity	.557	.006	19.33*	7,303
3.	Diagnostician Background Degree Certification Awareness of criteria Yr rcvd highest degree	.565	.011	12.74*	11,299

*p < .01

blocks. Regression coefficients, tests of their significance, and R^2 values for individual predictors are reported in Table 7 with AES1 in the equation and in Table 8 with AES2 in the equation. The only variables which explained more than one percent of the variance were AES1 and AES2.

The total variance explained by the combination of all the predictor variables was 31.9% using AES2 and 27.99% using AES1. The percentage of articulation errors (AES1 and AES2) was by far the best predictor, with an R^2 value of .255 for AES1 and .300 for AES2. No other individual predictors accounted for more than 1% of the variance.

The calculation of the standardized regression coefficients supports the explanation of variance data. The largest beta weights were for AES1 (Beta=.531; F=98.0; $p < .01$) and for AES2 (Beta=.577; F=120.8; $p < .01$). None of the other regression coefficients were statistically significant. The overall test for goodness of fit of the regression equations was also shown in Tables 5 and 6 for each regression. All of the regressions, except those involving child characteristics, were significant ($p < .01$), indicating that $R \neq 0$. Thus the null hypothesis that R=0 was rejected. The standard error of estimate for all the hierarchical regressions ranged from 1.12 to 1.32. This indicates that, on the average, the articulation severity ratings predicted by the three sets of variables will deviate from the actual scores by a little more than one point on the five point scale.

Table 7
Regression Analysis of Individual Predictors
Using AES1

Variables	B	BETA	F	Signif	R^2 Increase
AES1	.031	.531	98.0	p <.01	.255
Behavior	.097	.026	0.3	NS	.000
Test	-.227	-.084	2.6	NS	.009
Child Age	-.002	.012	0.1	NS	.000
Child Sex	.109	.038	0.6	NS	.002
Child Ethnicity	-.081	-.027	0.3	NS	.002
Lang Severity	-.014	-.018	0.1	NS	.002
Degree Year	-.003	.011	0.1	NS	.002
Certification	-.263	-.090	2.2	NS	.000
Degree	-.060	-.017	0.8	NS	.003
Awareness of Head Start criteria	-.704	-.025	0.3	NS	.000

Table 8
Regression Analysis of Individual Predictors
Using AES2

Variables	B	BETA	F	Signif	R^2 Increase
AES2	.035	.577	120.8	p <.01	.300
Behavior	.193	.052	1.2	NS	.002
Test	-.067	-.025	0.2	NS	.001
Child Age	.007	.045	0.7	NS	.001
Child Sex	.142	.049	1.0	NS	.003
Child Ethnicity	-.025	-.008	1.0	NS	.000
Lang Severity	-.014	-.020	0.1	NS	.002
Degree Year	-.007	-.028	0.3	NS	.000
Certification	-.294	-.100	3.0	NS	.007
Degree	-.041	-.012	0.0	NS	.003
Awareness of Head Start criteria	-.003	-.001	0.0	NS	.000

Hypothesis 4: The following interactions will be significant predictors of variance in the articulation severity rating: sex and behavior, ethnicity and behavior, AES and age, and AES and language severity rating

Correlation coefficients for these four interactions with the criterion and with the main effects are shown in Table 9. None of the interactions were correlated as highly with the criterion as the single main effect, AES2. The interaction of AES2 with age had the highest correlations (r=.5221) of all the interactions. Most of the correlation, however, was probably due to AES2 which was highly related to the AES2-age interaction (r=.9086). The interactions were not added to the main effects in the regression equation due to their relatively low correlations with the articulation severity rating. The null hpothesis, that there was no difference between any of the predictors, was accepted.

Additional Analysis

During the initial stages of the statistical analysis, the SPSS "Breakdown" program was used to provide descriptive information. The descriptive data generated by this program have already been reported. Although not called for in the original analysis plan, this procedure also provided a series of one-way ANOVA's. The variance of the independent variables was analyzed, first using articulation severity rating as the criterion, and then using the language severity rating. None of the articulation-based statistics were significant, while most of the language-based analyses resulted in significant findings. These data are reported in Table 10. The following significant results were provided by the series of one-way ANOVA's:

Table 9

Correlation Coefficients of Interactions with Criteria and Main Effects

(N=311)

	Beh-Sex	Beh-Eth	AES2-LSR	AES2-Age
Artic Severity	.0545	-.0226	.2921	.5221
Diagnostic Criteria				
AES2	-.0309	-.0017	.9086	.7286
Behavior	.4878	.2366	-.1163	-.1548
Artic Test	.0764	.1755	.0532	.0168
Diagnostician Bkgrd				
Degree	-.0221	.1806	.0398	.0875
Certification	-.0275	.1096	.0258	.1486
Degree Year	.1134	.1114	.0750	.1030
Know Criteria	.0011	.0628	-.0334	.0639
Child Characteristics				
Age	-.0531	.0432	.0351	-.2986
Sex	-.7968	-.0124	.0111	.0353
Ethnicity	.0048	.8958	.0112	.0830
Language Severity	-.0398	.1089	.1729	.7827

$p < .05$ for all $r \geq .0830$

$p < .01$ for all $r \geq .1163$

Abbreviations:

Beh-Sex=Interaction of child's behavior with child's sex

Beh-Eth=Interaction of child's behavior with child's ethnicity

LSR=Language Severity Rating

Table 10

One-way ANOVA Results with Language Severity Rating as the Criterion (N=311)

Source	DF	MS	F	P Level
Child Sex	1,309	0.4	0.1	.73
Child Language	1,309	8.8	2.6	.10
Diagnostician's Degree	1,309	16.5	4.9	.02
Degree Year	22,288	18.7	2.9	.00
Certification	1,309	45.1	13.9	.00

1. Children with behavior problems had higher language severity ratings than those with no such problems ($p < .06$).
2. Black dialect speakers had higher language severity ratings than Standard English speakers ($p < .10$).
3. Children diagnosed by staff with a bachelor's degree had lower language severity ratings than those diagnosed by staff with higher degrees ($p < .02$).
4. Children diagnosed by staff with CCC and/or private practice certification had higher severity ratings than those diagnosed by staff with no such certification ($p < .0001$).

There were also significant main effects associated with the recency of the diagnostician's training. As there were many categories for this variable, significant effects were not readily determined. For further analysis this variable was divided into the following discrete categories, based upon events related to the diagnosis of language disorders (and reviewed in Chapter 2):

(1) pre-1965: before the development of language diagnostic procedures

(2) 1965-1977: a period of awakening interest in the diagnosis of language disorders, but before federal intervention in the process

(3) after 1977: following the passage of the complete text of P.L. 94-142.

Although not included in the original ANOVA analyses, the child's age was considered as a potentially significant variable. There was no similar measure of language errors to substitute for the AES which had accounted for the majority of the explained variance in the articulation analysis.

At this point, a new hypothesis was formulated; i.e., that child's age, behavior, language, and diagnostician's recency of training and certification are significant predictors of language severity rating. As no subset of these variables was of interest and no causal ordering could be established through the literature, simultaneous regression analysis was used. The results of this multiple regression analysis, using language severity rating as the criterion, are presented in Table 11. The child's age and behavior and the diangostician's certification and degree year were significant predictors ($p < .01$) of the language severity rating. This combination of six predictor variables, however, explained slightly less than 10% of the variance.

The results obtained from this final analysis of language severity rating must be treated only as preliminary data, as there were no controls for level of language performance. These findings, however, provide some insight into the speech-language diagnostic process and offer implications for further research. Both of these topics will be discussed in the final chapter.

Table 11
Regression Analysis with Language Severity Rating as the Criterion

Variables	B	BETA	F	Signif	R^2 Increase
Child Age	-.328	-.147	7.23	.01	.017
Behavior	-.499	-.099	3.25	.01	.011
Degree Year	.469	.135	5.32	.01	.011
Diagnostician Degree	-.161	-.003	0.01	NS	.010
Certification	.978	.243	14.81	.01	.042
Child Language	-.361	-.074	1.79	NS	.005
				TOTAL	.099

CHAPTER V

DISCUSSION, SUMMARY, AND CONCLUSIONS

Hypothesis Testing

Four hypotheses were orignally proposed for this investigation. One additional hypothesis concerning language severity ratings was introduced later, based on preliminary findings. The following section summarizes and draws conclusions from the results of hypothesis testing.

Diagnostic Criteria

The first hypothesis of the use of a wide range of objective and subjective criteria in rating preschool articulation problems was supported through several findings. A total of 15 different instruments, ranging from formal, standardized measures to informal teacher-made devices were used. While these tests are intended as an objective measure of the child's articulation, most leave the interpretation to the diagnostician. The most commonly used measure, the Goldman Fristoe Test of Articulaton, provides no normative data. Tests such as the Templin Darley or Arizona measures, which provide normative and/or interpretive assistance, were much less frequently used. This finding corresponds with Davis and Shephard's observations of speech-language testing in Oregon (1982).

Even greater variability was found in the scope of factors cited as influencing the diagnostic decision. While the majority of diagnosticians indicated that they considered no factors other than speech-language test

results, some evaluated non-speech characteristics such as the child's aggressive or withdrawn behavior or maturity. Many diagnosticians appeared to consider the child's total social environment, including the home situation or the speech of family members. Others were influenced by factors unrelated to the child. Two professionals, for example, were concerned about the stigma which the community placed on mental retardation and other handicaps. They preferred to diagnose children with other severe problems as primarily SI to avoid potentially stigmatizing labels. Some were concerned about the lack of other services for children with multiple handicaps. Speech was diagnosed as the primary handicap in some cases because it was the only impairment for which remediation programs were available. Other staff were influenced by parent or teacher pressure in making their diagnostic decisions.

Diagnostician training and experience

The second hypothesis of a wide range of diagnostician training and experience was supported by several variables. Although the vast majority of children were diagnosed by white females, there were important differences among these staff. While the differences in experience (1 to 23 years) may not be related to diagnostic decision making (Fristoe and Goldman, 1968) they were related to the recency of training. Less experienced diagnosticians were much more likely to have benefitted from later schooling which provided training in language evaluation and possibly in the provisions of recent handicapped legislation. The sampled children were diagnosed by staff who had last earned a degree as early as 1954 or as late as 1982.

The fact that only a few states have graduate training requirements for speech-language diagnosticians, does not appear to have influenced the educational decisions of these staff. The majority held master's degrees in speech-language pathology, although a substantial number had no graduate degrees. The greatest variation among staff was found in the certification area. The diagnosticians were certified through a number of different authorities, including the public schools, state governments, and ASHA. At least 24 different combinations of certifications involving general education, special education, endorsement in speech, state license for private practic, state license for public school and the Certificate of Clinical Competence (CCC) were identified for the 130 diagnosticians. Less than five percent of the children were diagnosed by staff with none of these certifications.

Prediction

The differences in diagnostic criteria and in diagnostician background appeared to offer very little explanation for differences in diagnostic decision-making. The only variable which was a significant predictor of the diagnostician's rating of articulation severity was the child's performance on the articulation test. Although a wide variety of measures were used, the type of test did not exercise a signifcant influence on the child's performance or the diagnostician's severity rating.

None of the child characteristics offered a substantial explanation of differences in the severity ratings, either. There were twice as many boys as girls in the sample, exactly the ratio reported by Tomlinson and associates (1977). There was, however, no evidence of the gender-bias observed in an

earlier study (Owen, 1978). Children from minority ethnic groups were not overrepresented in the sample. There was some evidence of the "squeaky wheel concept" which Weatherly and Lipsky (1977) have noted; that is, diagnosticians considered "acting out" or aggressive behavior as a factor in about 15% of the diagnoses. Children with behavor problems, however, were not rated any differently than those with no reported behavor problems. The only dialect or language other than Standard English which was substantially represented in the sample was black dialect. Despite the concern over discriminatory speech-language testing for this linguistic minority (Vaughn-Cooke, 1982) there was no evidence of discrimination in the articulation diagnosis.

While language skills have been shown to influence the development of articulation skills and the diagnostician's perception of a child's articulation skills (Randall et al., 1974), this relationship was not evident in this investigation. Differences in language severity ratings explained little of the variance in articulation severity ratings. Language ratings were, however, related to some variables other than articulation, as explained in a later section.

None of the interactions between variables were significant predictors of how diagnosticians would rate articulation severity. Despite the findings of Phipps (1982) that boys are more often judged to have behavioral problems, there was no evidence that males or minority students with behavioral problems were diagnosed differently than other groups.

There was also no evidence that age was considered as a diagnostic factor when evaluating the child's test performance. Younger children with a

high percentage of errors were rated no differently than older children with the same percentage of errors. Existing articulation norms, while differing on specific sounds learned at each stage, all identify a significantly higher percentage of correct sound production for five-year old chilren than for three-year olds. The fact that diagnosticians appeared to ignore the age variables supports earlier findings that speech-language norms are too inconsistent to serve as a major tool in the diagnostic process (Stewart & Weybright, 1980).

Language Severity Rating

As noted earlier, language assessment is an even more complex, less clearly defined process than articulation. The present investigation focused on the diagnosis or articulation with the reasoning that, if variations in articulation criteria were found to be associated with the three groups of variables, then a similar relationship could be assumed in more complex, less clearly defined areas of testing, such as language. It is not surprising then that some relationships with predictor variables were tentatively found to be stronger for language severity ratings than for articulation ratings. Even in this analysis, however, no evidence of discrimination on the basis of sex or ethnic status was found.

The variables which were significant predictors of the language severity rating were related to the diagnostician's background and to the chld's behavor. It is not possible to draw conclusions about diagnostic criteria because of the limited information available specifically for the language diagnosis. The strongest factor was the effect of the diagnostician's

certification. Children diagnosed by staff with the CCC and/or private practice certification had higher language severity ratings than those diagnosed by staff with no such certification. The latter represents a higher, less easily obtained level of certification. While private practice certification varies across states, the CCC is a national certificate awarded by ASHA. It requires graduate-level coursework, extensive field work, and successful completion of a written examination. The certification variable is thus somewhat related to both the educational degree and years of experience variables. The finding that staff with higher certification levels rated language severity differently may mean that these diagnosticians have greater awareness or knowledge of language assessment. Graduate-level training programs are much more likely to provide language assessment training than are undergraduate programs (Muma et al., 1979). Additionally, the complexity of language assessment dictates greater diagnostician experience with evaluation instruments and interpretation.

As discussed earlier, the period in which training was conducted may exercise an influence on the diagnostician's knowledge of language assessment. Staff trained after the early 1970's are much more likely to have received instruction in language evaluation and in the provisions of recent federal legislation for the handicapped. The finding that diagnosticians trained after 1977 were likely to assign the highest language severity ratings may closely parallel the certification observations.

Greater awareness of the language component in speech-language assessment may influence the diagnostician's language severity ratings in at least two ways. First, he/she may place more emphasis on language testing

and remediation than on articulation, thereby recommending more children with language impairments for Head Start's SI programs. Secondly, the depth of the language assessment may be a factor in the variance of severity ratings. Two diagnosticians with varying levels of language assessment skills may both agree that a child has a language problem. The diagnostician with less training and experience may evaluate only the child's vocabulary, for example, concluding that the deficit was moderate. The more experienced staff may include an assessment of morphology, syntax, and auditory processing and find additional deficits. This person would be more likely to conclude that the language problem was more severe.

The fact that the child's age had a significant effect on the language severity rating is confounded by the lack of data on language testing results for the sampled children. It may be that "language error score" would have explained most of the variance attributed to the age variable.

Conclusions

This study investigated non-measurement factors which influence the interpretation of test information. While non-measurement bias may be a factor in the overall preschool speech-language diagnosis, it did not appear to affect the articulation component. Diagnosticians in this investigation appeared to be more heavily influenced by the actual test score than by any non-measurement factors. This leaves little room for the operation of individual biases, based on child characteristics or diagnostic training and experience.

It may be that diagnosticians are unaware of what factors influence their decisions. While staff indicated that they considered factors related to the child's social envoronment or administrative policies, there was no evidence that these factors were related to the final diagnostic outcome.

These findings appear to contradict charges of race, sex, and linguistic bias in speech testing. They also indicate that existing articulation measures do not result in discriminatrory testing, despite that the fact that they have not been modified for black dialect speakers or that their articulation norms may be inconsistent.

None of the issues associated with diagnostician's education, certification or demographic background were related to the articulation diagnosis. The type of degree or certification required by state and local jurisdictions for speech diagnosticians does not appear to be a necessary control in reducing non-measurement bias in the articulation diagnosis.

The study results lend no support to the notion that changes in training programs, certification requirements, tests, or diangostic criteria are needed to prevent bias in articulation assessment. They do, however, raise some additional questions. As discussed earlier, the area of language assessment is open to greater diffferences in test use and interpretation (i.e., non-measurement bias) than is articulation assessment. Results of exploratory research from this investigation indicate that the biases hypothesized for articulation diagnosis may affect the language diagnosis. While no race, sex, or linguistic biases were found, children with behavioral problems tended to be categorized as having more severe language problems. As no language error scores were available, it is not possible to say whether discrimination was

practiced. It may be that these hyperactive or aggressive children were less attentive and had less opportunity to develop linguistic skills.

The role of the diagnostician's knowledge and biases may have an important influence on test use and interpretaion. This was suggested by the significant effect that certification and recency of training had on the language severity rating. The present study, however, focused on the articulation diagnosis and was not rigorous enough to allow conclusions in the language area.

A major factor in the investigations of both articulation and language diagnosis was the amount of unexplained variance. In the observation of language severity ratings, however, the language error score was not included so it is possible that this variable exercised a heavy influence. Less than one third of the differences in articulation ratings were attributable to all of the variables combined. This leaves open the possibility that other measurement and/or non-measurement biases may influence the diagnosis. Further research in this area could focus on a number of other variables which may influence the speech-language diagnosis.

Suggestions for Further Research

The results of this research suggest two separate directions for further investigation of measurement and non-measurement bias in the speech-language diagnostic process. First, although articulation assessment may not be open to the diagnostic bias investigated here, there is some evidence that language assessment may be influenced by the variables included in this research. Future investigations may be directed toward the development and

inclusion of a "language error score" which universally measures language proficiency. The types of language instruments used in the diagnosis may also be an important variable.

A second area of investigation might focus on the amount of unexplained variance in both the articulation and language ratings. Some potential predictors which were not included in this research could include fiscal or political factors such as policies concerning diagnostic criteria, caseload restrictions, availability of personnel for diagnosis and therapy, and the availability of other local facilities for handicapped children. Future research might also focus more intensively on diagnostic criteria, analyzing variables such as the specific phonemes in error, types of phoneme errors (substitutions, omissions, distortions), or the norms used by the diagnostician.

Another procedure for investigating diagnostic criteria would be to observe the characteristics of those who were found not to be primarily SI. Children with other handicaps or in other facilities might be included in future research. Discrimination against boys or minority students, for example, may take the form of more restrictive placement in EMR programs or as SI in institutions outside of Head Start. Another strategy would be to include non-handicapped children. One objective of this procedure would be to investigate the speech-language skills of girls, minority ethnic groups, etc. who have never been referred for SI diagnostic testing. One potential finding of such a study might be that girls with speech-language problems are never referred due to factors such as withdrawn behavior or diagnostician or teacher bias.

Summary

Although variations in diagnostic criteria and diagnostician background were observed, no evidence of non-measurement bias in the diagnosis of preschool articulation impairment was uncovered in the present investigation. Diagnosticians appeared to be influenced primarily by articulation test scores in rating the severity of the child's impairment.

The combination of variables included in the present investigation explained slightly less than one third of the variance in severity ratings. This indicates a need for further study of other variables such as the fiscal and political arena in which the diagnostician operates which might also be considered as a potential source of non-measurement bias. Preliminary evidence from this study indicated, for example, that some diangosticians were influenced by caseload restrictions, availability of diagnostic and treatment staff, parent or teacher pressure or local diagnostic policies.

There is also a need for further investigation of non-SI children before any definite conclusions may be drawn concerning the influence of diagnostic criteria, child characteristics or diagnostician background on the assessment process. A study of children who have gone through the diagnostic process, but have not been diagnosed as SI and/or have been diagnosed as having a more serious handicap would provide more information about the influence of non-measurement factors. Discrimination against boys or minority students, for example, may take the form of more restrictive placement in EMR programs or as SI in institutions outside of Head Start. Similarly, girls with speech-language problems may never be referred due to factors such as withdrawn behavior or diagnostician or teacher bias.

REFERENCES

Abeson, A. & Ballard, J. State and federal policy for exceptional children. In Weintraub et al. (Eds), Public policy and the education of exceptional children. Reston, VA: The Council for Exceptional Children, 1976.

Abseon, A. & Zettel, J. The end of the quiet revolution. The Education for All Handicapped Children Act of 1975. Exceptional Children, 1977, 44, 114-128.

Adams, M.R. Some common problems in the design and conduct of experiments in stuttering. Journal of Speech and Hearing Disorders, 1976, 41, 3-9.

Administration for Children, Youth and Families. Announcement of diagnostic criteria for reporting handicapped children in Head Start: Transmittal notice 75.11. Washington, D.C., 1975.

Alabama Department of Education. Rules and regulations for special education programs. Mobile, AL: Author, 1980.

Alaska Department of Education. State guidelines for special education. Juneau: Author, 1978.

Algozzine, B. & Ysseldyke, J.E. & Shinn, M. Identifying children with hearing disabilities: When is a discrepancy severe? Journal of School Psychology, 1982, 20, 299-305.

Allegheny County Public Schools, Handicap definitions, Allegheny, PA: Author, 1980.

Allen, E.Y., Black, M.E., Burkland, M., Byrne, M.C., Farquhar, M.S., Herbert, E.L., & Robertson, M.L. Case selection in the public schools. Journal of Speech and Hearing Disorders, 1966, 31, 157-161.

American Speech, Language and Hearing Association. Definitions: Communicative disorders and variations. Asha, 1982, 24, 949-950.

Amira, S., Abramowitz, S. I. & Gomes-Schwartz, B. Socially-charged pupil and psychologist effects on psychoeducational decisions. The Journal of Special Education, 1977, 11, 433-440.

Anderson, J.L. Report on survey of speech-language pathologists and audiologists in public schools of Indiana. Unpublished manuscript, 1982. (Available from Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN.)

Arlington County Board of Education. Special Education Handbook. Arlington, VA: Author, 1981.

Arlt, P.B. & Goodban, M.T. A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six. Language, Speech and Hearing Services in the Schools, 1976, 7, 170-183.

Arnold, K.S. & Reed, L. The grammmatic subciosure subtest of the ITPA: A comparative study of black and white children. Journal of Speech and Hearing Disorders, 1976, 41,477-485.

Arvisu et al. v. Waco Independent School District. Federal Supplement, 1973, 373, 1264-1272.

Baratz, J.C. Language in the economically disadvantaged child, a perspective. Asha, 1968, 10, 143-145.

Barresi, J. & Bunte, J. Special education certification procedures: a summary of a national survey. Reston, VA: Council for Exceptional Children, 1979.

Beatty, J.R. Identifying decision-making policies in the diagnosis of learning disabilities. Journal of Learning Disabilities, 1977, 10, 201-209.

Bennet, R.E. Methods for evaluating the performance of school psychologists. School Psychology Monograph, 1980, 4, 45-59.

Bennet, R.E. Professional competence and the assessment of exceptional children. The Journal Special Education. 1981, 15, 437-446.

Berenthal, J.E. & Bankson, N.W. Articulation disorders. Englewood Cliffs, NJ: Prentice Hall, 1981.

Bernard, R. & Clarizio, T. Socioeconomic bias in special education placement decisions Pyschology in the Schools, 1981, 18, 178-183.

Bersoff, D.N. Testing and the law. American Psychologist, 1981, 36, 1047-1056.

Bibb County Board of Education. Special education policies and procedures. Bibb County, GA: Author, 1979.

Bischoff, H.G.W. & Shakel, J.A. Definitions and interventions for the emotionally disturbed in Alaska's rural schools. Anchorage, AK: Anchorage Public Schools, 1980.

Blom, G.E. Sex differences in reading disability. In E. Calkins (Ed.) Reading forum. Bethesda, MD: National Institute of Neurological Disease and Stroke, 1971.

Bracken, B.A. & Prasse, D.P. Comparison of the PPVT-R and WISC-R with urban mentally retarded students. Psychology in the Schools, 1981, 18, 174-177.

Brady, P.M., Manni, J.L., & Winnikur, D.W. A three tiered model for the assessment of culturally and linguistically different children. Psychology in the Schools, 1983, 20, 52-58.

Bransford, L. Social issues in special education: Stop gap measures may be a form of cultural genocide. Phi Delta Kappan, 1974, 55, 530-532.

Brown, T. Special Education Handbook: State of Alaska. Juneau, Alaska: Alaska State Department of Education, 1978.

Bureau of Education for the Handicapped, Unite States Office of Education. Annual report to Congress, 1978-1979 school year. Washington, D.C.: U. S. Government Printing Office, 1980.

Bureau of Education for the Handicapped, U. S. Department of Health, Education, and Welfare. State plan amendment for fiscal year 1975 under part B, Education for All Handicapped Act. Washington, D.C.: Author, 1974.

Campbell, M.A. Operation of special education programs for handicapped children below age five. Lincoln, NE: Nebraska Department of Education, 1979.

Carrell, J.A. Disorders of articulation. Englewood Cliffs, NJ: Prentice Hall, 1968.

Carrow, E.C. Test for auditory comprehension of language. Austin, TX: Learning Concepts, 1973.

Clarizio, H.F. In defense of the IQ test. School Psychology Digest, 1979, 8, 79-88.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A., & Wesman, A. Educational uses of tests with disadvantaged students. American Psychologist, 1975, 30, 15-41.

Cohen, J. & Cohen, P. General multiple regression and correlation analysis in Bentler, P.M., Lettieri, D.J., & Austin, G.A. Data analysis strategies and designs for substance abuse research. Rockville, MD:National Institute on Drug Abuse, 1976.

Cole, N.S. Bias in testing. American Psychologist, 1981, 36, 1067-1077.

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

Craig,P.A. & McEachron,N.B. Studies of handicapped students: Volume I. Whom do teachers identify as handicapped? Menlo Park, California Stanford Research Institute, 1978.

Dade County Public Schools, Dade County procedures. Miami, FL: Author, 1979.

Darley, F.L. Diagnosis and appraisal of communication disorders. Englewood Cliffs, NJ: Prentice Hall, 1964.

Darley, F.L. Evaluation of appraisal techniques in speech and language pathology. Reading, MA: Addison-Wesley, 1979.

Darlington,R.B. Multiple Regression. Psychological Bulletin, 1968, 34, 161-182.

Davis, W.A. & Shepard, LA. The use of tests by LD teachers, school psychologists and speech/language specialists in the diagnosis of learning disabilities. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, March 19-23, 1982.

Department of Public Instruction. Rules and regulations for programs providing services to children with handicapping conditions. Olympia, WA: Author, 1980.

Diana v. State Board of Education. Civil Action No. C-7037 RFP. Northern District California, January 7, 1970 and June 18, 1973.

Draper,N. & Smith,H. Applied Regression Analysis. New York:John Wiley & Sons, 1966.

Drumwright,A., VanNetta,P., Camp,B., Frankenburg,W., & Drexler,H. The Denver articulation screening exam. Journal of Speech and Hearing Disorders, 1973, 38,3-14.

Dublinske, S., Karr, S.T. & Downey, M. Serving speech impaired children under P.L. 94-142: response of the American Speech-Language-Hearing Association to the United States General Accounting Office report. Unpublished manuscript, 1981. (Available from ASHA, 10801 Rockville Pike, Rockville, MD 20852).

Duffey,J.B.,Salvia,J, Tucker,J. & Ysseldyke,J. Nonbiased assessment:a need for operationalism. Exceptional Children, 1981, 47, 427-434.

Dunn, L.M. Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service, 1965.

Dunn, L.M. Special education for the mildly retarded - is much of it justifiable? Exceptional Children, 1968, 35, 5-22.

Education Advocates Coalition. Misclassification of minorities in "Educable Mentally Retarded" classes. Integrated Education, 1980, 18, 113-115.

Emerick, L. & Hatten, J. Prognosis and evaluation in speech pathology. Englewood Cliffs, NJ: Prentice Hall, 1974.

Epps,S., McGue,M. & Ysseldyke,J. Interjudge agreement in classifying students as learning disabled. Psychology in the Schools, 1982, 19, 209-220.

Evans, P.L. Classification and labeling issues: toward operationalizing legal definitions of handicapping conditions. Paper presented at the Annual Convention of the American Psychological Association, New York City, September 1, 1979.

Evard, B.L. & Sabers, D.L. Speech and language testing with distinct ethnic-racial groups: a survey of procedures for improving validity. Journal of Speech and Hearing Disorders, 1979, 44, 271-281.

Fenton,K.S., Yoshida, R.K., Maxwell, J.P., & Kaufmann, M.J. Recognition of team goals: An essential step toward rational decision making. Exceptional Children, 1979, 45, 638-644.

Finn, J. & Mattsson, I. Multivariate analysis in educational research. Applications of the MULTIVARIANCE program. Chicago,IL: National Educational Resources, 1978. Flaugher, R. The many definitions of test bias. American Psychologist, 1978, 33,1671-1679.

Flaugher, R. The many definitions of test bias. American Psychologist, 1978, 33, 1671-1679.

Florida Department of Education. A resource manual for the development and evaluation of special programs for exceptional students. Tallahassee, FL: Author, 1980.

Flower, R.M., Leach, E., Stone, C.R., & Yoder, D.E. Case selection. Journal of Speech and Hearing Disorders, 1967, 32, 65-70.

Foster, C.R., Giddan, J.J. & Stark, J. Assessment of children's language comprehension. Palo Alto, CA: Consulting Psychologists, 1973.

Frame, R.E., Clarizio, H., Porter, J., & Vinsonhaler, P. Interclinician agreement and bias in school psychologists' diagnostic and treatment recommendations for a learning disabled child. Psychology in the Schools, 1982, 19, 319-327.

Freilinger, J.J. Iowa's severity rating scale for communication disabilities. Des Moines, Iowa: Iowa State Department of Public Instruction, 1978.

Fresno Unified School District. Comprehensive plan for special education. Fresno, CA: Author, 1981.

Fristoe, M. & Goldman, R. Comparison of traditional and condensed articulation tests examining the same number of sounds. Journal of Speech and Hearing Research, 1968, 11, 583-589.

Fudala, J.B. Arizona articulation proficiency scale. Los Angeles: Western Psychological Service, 1963.

Garrison, M. & Hammil, D. Who are the retarded? Exceptional Children, 1977, 43, 464-465.

Georgia Department of Education, Program for exceptional children: regulations and procedures. Atlanta, GA: Author, 1978.

Georgia Department of Education. Speech and language impaired: Resource manuals for programs for exceptional children. Volume VII: Speech and language impaired. Atlanta, GA: Author, 1980.

Giesbrecht, M.L. & Routh, D.F. The influence of categories of cumulative folder information on teacher referrals of low-achieving children for special educational services. American Educational Research Journal, 1979, 16, 181-187.

Gordon,R.A. Issues in multiple regression. American Journal of Sociology, 1968 73, 592-616.

Granger, R.C., Matthews, M., Quay, L.C., & Verner, R. Teacher judgments of the communication effectiveness of children using different speech patterns. Journal of Educational Psychology, 1977, 69, 793-796.

Gresham, F.M. Assessment of children's social skills. Journal of School Psychology, 1981, 19, 120-133.

Grill, J.J. & Bartel, N.R. Language bias in tests: ITPA Grammatic Closure. Journal of Learning Disabilities, 1977, 10, 229-235.

Guadalupe Organization, Inc. v. Tempe Elementary School Dist. No. 3, Civil No. 71-435 Phoenix (D. Arizona, Jan. 25, 1972).

Guntersville Board of Education, Definitions and characteristics of exceptional children. Guntersville, AL: 1980.

Gutkin, T.B. & Reynolds, C.R. Factorial similarity of the WISC-R for white and black children from the standardization sample. Journal of Educational Psychology, 1981, 73, 227-231.

Half Hollow Hills District Board of Education, Special education delivery system. Half Hollow Hills, NY: author, 1979.

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

Harber, J.R. Assessing the quality of decision making in special education. The Journal of Special Education, 1981a, 15, 77-90.

Harber, J.R. Evaluting utlitiy in diagnostic decision making. The Journal of Special Education, 1981b, 15, 413-428.

Hejna, R. Developmental articulation test. Madison, WI: College Printing and Typing, 1975.

Henderson, F.M. Accuracy in testing the articulation of speech sounds. Journal of Educational Research, 1938, 31, 348-356.

Henrickson, E.H. Case selection in the schools. Journal of Speech and Hearing Disorders, 1968, 33, 231-235.

Hobson v. Hansen. Federal supplement, 1967, 269, 401-408. Affd. en banc subnom. Smuck v. Hobson. The Federal Reporter 2nd Series, 1969, 408, 175.

Jackson, P. & Lahaderne, H. Inequalities of teacher-pupil contacts. Psychology in the Schools, 1971, 8, 139-147.

Jirsa, J.E. The SOMPA: A brief examination of technical considerations, philosophical rationale, and implications for practice. Journal of School Psychology, 1983, 21, 13-22.

Johnson, V.M. Analysis of factors influencing special educational placement decisions. Journal of School Psychology, 1980, 18, 191-202.

Jones, P.R. A practical guide to federal special education law. New York: Holt, Rinehart & Winston, 1981.

Kakalik, J.S., Brewer, G.D., Dougharty, L.A., & Fleischauer, P.D. Services for handicapped youth: A program overview. Santa Monica, CA: Rand, 1973.

Kansas State Department of education. State plan for fiscal year 1981: Special education. Topeka, KS: Author, 1980.

Kirp, D. & Kirp, L. The legislation of the school psychologist's world. Journal of School Psychology, 1976, 14, 83-89.

Koenigskneecht, R.A. & Friedman, P. Syntax development in boys and girls. Child Development, 1976, 47, 1109-1115.

Labov, W. The study of nonstandard English. Champaign, IL: National Council of Teachers of English, 1970.

Lamberts, F. Describing children's language behavior. In Sabatino, D.A. & Miller, T.L. (Eds.) Describing learner characteristics of handicapped children and youth. New York: Grune & Stratton, 1979.

Larry P. v. Riles. Federal Supplement, N.D. Col., 1972, 204, 1306; The Federal Reporter 2nd Series, 1974, 502, 963.

Lessen, E.I. & Rose, T.L. State definitions of preschool handicapped populations. Exceptional Children, 1980, 46, 467-469.

Li, A.K.F. Learning disabilities: a review essay. Journal of Educational Thought. 1977, 11, 253-263.

Lietz, J.J. & Gregory, M.K. Pupil race and sex determinatins of office and exceptional educational referrals. Educational Research Quarterly, 1978, 3, 61-66.

Longhurst, T.M. & File, T.J. A comparison of developmental sentence scores from Head Start children collected in four conditions. Language, Speech and Hearing Services in Schools, 1977, 8, 54-64.

Longhurst, T.M. Language acquisition and assessment in normal and handicapped preschool children: a review of the literature. Final Report. Topeka, KA: Kansas Neurological Institute, 1977.

Louisiana Board of Education. Special education regulations. Baton Rouge, LA: Author, 1980.

Low, B.P. & Clement, P.W. Relationships of race and socioeconomic status to classroom behavior, academic achievement, and referral for special education. Journal of School Psychology, 1982, 20, 103-112.

MacDonald, J.D. & Martin, R.R. Stuttering and disfluency as two reliable and unambiguous response classes. Journal of Speech and Hearing Research, 1973, 16, 691-699.

Mack, J. H. An analsyis of state definitions of severely emtionally disturbed. Washington, D.C.: Office of Special Education, 1980.

MacMillan, D. & Meyers, C. The nondiscriminatory testing provision of P.L. 94-142. Viewpoints, 1979, 53, 39-56.

Madison, C.L., Kolbeck, C.P., & Walker, J.L. An evaluation of stimuli identification on three articulation tests. Language, Speech and Hearing Services in the Schools, 1982, 13, 110-115.

Malgoire, M.A., Craig, P.A. & Kaskowitz, D.H. Identification of handicapped students (ages 12-17) using data from teachers, parents and tests. Washington, D.C.: Department of Health, Education, and Welfare, 1979.

Manni, J.L., Winikur, D.W., & Keller, M. The status of minority group representation in special education programs in the State of New Jersey. New Jersey Department of Education, 1980.

Marquardt, P.T. & Saxman, J.H. Language comprehension and auditory discrimination in articulation deficient kindergarten children. Journal of Speech and Hearing Research, 1972, 15, 382-389.

Martin Luther King Junior Elementary School Children et al. v. Ann Arbor School District Board. Federal Supplement, 1979, 473, 1371-1378.

Martin, R. Educating handicapped children. Champaign, Illinois: Research Press, 1979.

Matuszek, P. & Oakland, T. Factors influencing teachers' and psychologists' recommendations regarding special class placement. Journal of School Psychology,1979, 17, 116-125.

McDaniels, G. Assessing handicapped students: Beyond identification. In W. Schrader (Ed.), Measurement and educational policy: Proceeedings of the 1978 ETS Invitational conference. San Francisco: Jossey-Bass, 1979.

McLaughlin, J.A., Miller, C.K., & Chansky, N.M. Prognoses made by special educators. Mental Retardation, 1970, 8, 20-24.

Menyuk, P. The acquisition and development of language. Englewood Cliffs, NJ: Prentice Hall, 1971.

Mercer, C.D., Forgnone, C. & Wolking, W. Definitions of learning disabilities used in the United States. Journal of Learning Disabilities, 1976, 9, 376-386.

Mercer, J. Current retardation procedures and the psychological and social implications on the Mexican American: Position Paper #2., 1970. (Available on Education Resources Information Center Microfiche ED058848).

Mercer, J. Psychological assessment and the rights of children. In N. Hobbs (Ed.), Issues in the classification of children. San Francisco: Jossey Bass, 1975.

Mercer, J. Technical manual. SOMPA: System of Multicultural Pluralistic Assessment. New York: Psychological Corporation, 1979.

Meyers, C. Sundstrom, P. & Yoshida, R. the school psychologist and assessment in special education. School Psychology Monograph, 1974, 2, 1-57.

Michel, L.I. Evaluation of articulation disorders: traditional approach. In S. Singh & J. Lynch (Eds.) Diagnostic procedures in Hearing, Language and Speech. Baltimore: University Park Press, 1978.

Michigan State Board of Education, Michigan special education rules. Lansing, MI: Author, 1980.

Miller, K., Prutting C. Consistencies across three language comprehensive tests for specific grammatical features. Language, Speech and Hearing Services in Schools, 1979, 10, 162-170.

Montgomery County Public Schools. Developmental evaluation services for children: a replication manual. Rockville, MD: Author, 1978.

Mullen, P.A. & Whithead, R.L. Stimulus picture identification in articulation testing. Journal of Speech and Hearing Disorders, 1977, 42, 113-116.

Muma, J.R., Webb, P.H., & Muma, D.B. Language training in speech-language pathology and audiology: A survey. Asha, 1979, 21, 467-473.

Murai, H.M. Eligibility criteria for learning disabilities programs. Journal of Learning Disabilities, 1982, 15, 267.

Nazarro, J. Comprehensive assessment for educational planning. In F. Weintraub et al. (Eds), Public policy and the education of exceptional children. Reston, VA:Council for Exceptional Children, 1976.

Nelson, K.E. & Bonvillan, J.D. Early language development: conceptual growth and related processes between 2 and 4½ years of age. In K.E. Nelson (Ed.), Children's Language: Volume 1, New York: Gardner Press, 1978.

Nie, N.H., Hull, H.C., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. SPSS: Statistical Package for the Social Sciences. New York: McGraw Hill, 1975.

Norris, M., Harden, J.R., & Bell, D.M. Listener agreement on articulation errors of four- and five-year old children. Journal of Speech and Hearing Disorders. 1980, 45, 378-389.

Oakland, T., DeLuna, C., & Morgan, C. Annotated bibliography of language dominance measures. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Oakland, T. & Feigenbaum, D. Multiple sources of test bias on the WISC-R and Bender Gestalt test. Journal of Consulting and Clinical Psychology, 1978, 46, 417-422.

Oakland, T., & Matuszek, P. Using tests in nondiscriminatory assessment. In T. Oakland (Ed.), Psychological and educational assessment of minority children. New York: Brunner/Mazel, 1977.

Ohio Department of Education. State administrative procedures for the education of the handicapped. Columbus, OH: Author, 1980.

Oklahoma City Education Department. Special education policies and procedures. Oklahoma City, OK: Author, 1980.

Oklahoma State Department of Education. Annual program plan for special education. Oklahoma City, OK: Author, 1979.

Olmedo, E.L. Testing linguistic minorities. American Psychologist, 1981, 36, 1078-1085.

Owen, F.W. Dyslexia - Genetic aspects. In A.L. Benton & D. Pearl (Eds.) Dyslexia: An appraisal of current knowledge. New York: Oxford University Press, 1978.

Palardy, J. What teachers believe - what children achieve. Elementary School Journal, 1969, 69, 370-374.

Pedhazur,E.J. Multiple regression in behavioral research. New York:Holt, Rinehart, & Winston, 1982.

Pennsylvania Department of Education. Special education policies and procedures manual. Harrisburg, PA: Author, 1980.

Phipps, P.M. The LD learner is often a boy - why? Academic Therapy, 1982, 17, 425-430.

Podilla, A.M. & Garza, B.M. IQ tests: A case of cultural myopia. The National Elementary Principal, 1975, 54, 53-58.

Poland, S., Ysseldyke, J., Thurlow, M., & Mirkin, P. Current assessment and decision-making practices in school settings as reported by directors of special education, (Research Report No. 14). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Poole, I. Genetic development of articulation of consonant sounds in speech. Elementary English, 1934, 11, 159-161.

Prather, E.M., Hedrick, D.L. & Kern, C.A. Articulation development in children aged two to four years old. Journal of Speech and Hearing Disorders, 1975, 40, 179-191.

Powers, M. functional disorders of articulation: symptomatology and etiology. In L.E. Travis (Ed.), Handbook of speech pathology. New York: Appleton-Century- Crofts, 1957.

Prutting, C.A. Process \'prä ˊ /, ses\ n: The action of moving forward progressively from one point to another on the way to completion. Journal of Speech and Hearing Disorders, 1979, 44, 3-30.

Purvin, G. The hidden agendas of IQ. The National Elementary Principal, 1975, 54, 44-48.

Randall, D., Reynell, J. & Curven, M. A study of language development in a sample of 3 year old children. British Journal of Disordered Communication, 1974, 9, 3-16.

Randolph, L.A. Diagnosis and assessment of handicapped children in Head Start: Executive summary. Washington, D.C.: Department of Health, Education, and Welfare, 1977.

Rees, N.S. Art and science of diagnosis in hearing, language, and speech. In Singh, S. & Lynch, J. (Eds.), Diagnostic procedures in hearing, language, and speech. Baltimore, MD: University Park Press, 1978.

Reschly, D.J. Psychology testing in educational classification and placement. American Psychologist, 1981, 36, 1094-1102.

Reschly, D. Nonbiased assessment. In G. Phye & D. Reschly (Eds.), School psychology: Perspectives and issues. New York: Academic Press, 1979.

Reschly, D. WISC-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. Journal of Consulting and Clinical Psychology, 1978, 46, 417-422.

Reynolds, M.C. Education of handicapped students: some areas of confusion. Phi Delta Kappan, 1980, 61, 603-604.

Rhode Island Board of Regents for Education, Education for handicapped children. Providence, RI: Author, 1978.

Rhode Island Society for Autistic Children (RISAC) v. Bd. of Regents for Ed. for the State of RI, Civil Action No. 5081 (D.R., filed Jan. 22, 1973.)

Sabatino, D.A. Classification for handicapped children and youth. In Sabatino, D.A & Miller, T.L. (Eds.), Describing learner characteristics of handicapped children and youth. New York: Grune & Stratton, 1979.

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

Samuda, R.J. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead, 1975.

Sander, E.K. When are speech sounds learned? Journal of Speech and Hearing Disorders, 1972, 37, 55-63.

Sandoval, J. & Mille, M. Accuracy of judgments of WISC-R item difficutly for minority groups. Journal of Consulting and Clinical Psychology, 1980, 48, 249-253.

Sandoval, J. The WISC-R and internal evidence of test bias with minority groups. Journal of Consulting and Clinical Psychology, 1979, 47, 919-927.

Savage, D.G. Educating all the handicapped. Arlington, VA: National School Public Relations Association, 1977.

Schissel, R.J., & Flourny, J.E. An investigation of judgments of experienced and inexperienced listeners in their use of a screening test of articulation. Journal of Communication Disorders, 1978, 11, 459-468.

Schissel, R.J., & James, L.B. An investigation of the assumptions underlying the scoring system of the Arizona Articulation Proficiency Scale: Revised. Language, Speech, and Hearing Services in the Schools, 1979, 10, 241-245.

Schumauch, V.A., Panagos, J.M. & Klich, R.J. Syntax influences the accuracy of consonant production in language disordered children. Journal of Communication Disorders, 1978, 11, 315-323.

Schwartz, R.G., Leonard, L.B., Wilcox, M.J., & Folger, M.K. Replication in child phonology. Journal of Child Language, 1980, 7, 75-87.

Segregation of poor and minority children into classes for the mentally retarded by the use of IQ tests. Michigan Law Reporter, 1971, 6, 1212-1250.

Shanks, S., Sharpe, M., & Jackson, B. Spontaneous responses of first grade children to diagnostic picture articulation test. Journal of Communication Disorders, 1970, 3, 106-117.

Shepard, L.A. & Smith, M.L. The identification, assessment, placement, and remediation of perceptual and communicative disordered children in Colorado. Boulder, CO: Laboratory of Educational Research, University of Colorado, 1981.

Shine, R.E., & Freilinger, J.J. Significant factors considered in selecting pupils for speech therapy. Unpublished paper, 1962. Available from J.J. Freilinger, Clinical Speech Svcs., Iowa Dept. of Public Instruction, Des Moines, IA.

Shriner, T.H., Holloway, M.S., & Daniloff, R.G. The relationship between articulatory deficits and syntax in speech defective children. Journal of Speech and Hearing Research, 1969, 12, 319-325.

Silverman, E. & Opens, K.V. An investigation of sex-bias in classroom teachers' speech and language referrals. Language, Speech, and Hearing Services in Schools, 1980, 11, 169-174.

Silversmith, R. A comparison of federal, state and local policies affecting education of handicapped children in Montgomery County Public Schools. Rockville, MD: Montogomery County Public Schools Department of Educational Accountability, 1980.

Stark, R.E. & Tallal, P. Selection of children with specific language deficits. Journal of Speech and Hearing Disorders, 1981, 46, 114-122.

State Department of Public Instruction. Rules governing programs and services for children with special needs. Raleigh, NC: Author, 1979.

Stewart, S.R. and Weybright, G. Articulation norms used by practicing speech-language pathologists in Oregon: results of a survey. Journal of Speech and Hearing Disorders, 1980, 45, 103-111.

Stockman, I.J. & Vaughn-Cooke, F.B. A re-examination of research on the language of black children: the need for a new framework. Journal of Education, 1982, 164, 157-72.

Stockman, I.J. & Vaughn-Cooke, F.B. A re-examination of research on the language of black children: the need for a new framework. Washington, D.C: Center for Applied Linguistics, 1982.

Taylor, J.S. Public school speech-language certification standards: Are they standard? Asha, 1980, 22, 159-165.

Templin, M.C. Certain language skills in children, their development and interrelationships. Institute of Child Welfare Monograph Series 26. Minneapolis: University of Minnesota Press, 1957.

Templin, M.C. & Darley, F.L. The Templin - Darley tests of articulation, second edition. Iowa City: University of Iowa Bureau of Educational Research and Service, 1979.

Thurlow, M. & Ysseldyke, J. Current assessment and decision-making practices in LD programs. Learning Disability Quarterly, 1979, 2, 15-24.

Tobias, S., Cole, C. Zibrin, M., & Bodlakova, V. Teacher-student ethnicity and recommendations for special education referrals. Journal of Educational Psychology, 1982, 74, 72-76.

Tomlinson, J.R., Acker, N., Canter, A., & Lindborg, S. Minority status, sex, and school psychological referrals. Psychology in the schools, 1977, 14, 457-460.

Toronto, A.S. Screening test of Spanish grammar. Journal of Speech and Hearing Disorders, 1976, 41, 150-171.

Tucker, J.A. Ethnic proportions in classes for the learning disabled: issues in nonbiased assessment. The Journal of Special Education, 1980, 14, 93-106.

U.S. Congress, Public Law 93-38, Section 613(a), Education Amendments of 1974, August 21, 1974.

U.S. Congress, Public Law 94-142, Part B, Education for All Handicapped Children Act, November 29, 1975.

U.S. Department of Education. Second annual report to Congress on the implementation of Public Law 94-142. Washington, D.C., 1980.

U.S. Department of Health and Human Services. The status of handicapped children in Head Start programs. Washington, D.C., 1980.

Vance, H.B. & Wallbrown, F.H. The structure of intelligence for black children: A hierarchical approach. The Psychological Record, 1978, 28, 31-39.

Vaughn-Cooke, F.B. Evaluating the language of Black English speakers: implications of the Ann Arbor decision. Unpublished manuscript, 1982. (Available from author: University of the District of Columbia, Washington, D.C.)

Walters, P. Evaluation of the process of mainstreaming handicapped children into Project Head Start: final report. Unpublished manuscript, HEW 103-76-1113, 1978.

Weatherly, R. & Lipsky, M. Street level bureaucrats and institutional innovations: Implementing special education reform. Harvard Education Review, 1977, 47, 171-197.

Webster, E.J., Perkins, W.H., Bloomer, H.H., & Pronovost, W. Case selection in the schools. Journal of Speech and Hearing Disorders, 1966, 31, 352-358.

Weintraub, F., Abeson, A., & Braddock, D. State law and education of handicapped children: Issues and recommendations. Arlington, VA: The Council for Exceptional Children, 1971.

Wellman, B., Case I, Mengert, I. & Bradbury, D. Speech sounds of young children. University of Iowa Studies of Child Welfare, 1931, 5, 1-82.

Wertz, R.T. & Mead, M.D. Classroom teacher and speech clinician severity ratings of different speech disorders. Language, Speech and Hearing Services in the Schools, 1975, 6, 119-124.

West Virginia Department of Education. Standards for the education of exceptional children. Charleston, WV: Bureau of Learning Systems, 1981.

Whitehead, R.L., & Mullen, P.A. A comparison of the administration times of two tests of articulation. Language, Speech, and Hearing Services in Schools, 1975, 6, 150-153.

Wingate, M.E. Criteria for stuttering. Journal of Speech and Hearing Research, 1977, 20, 396-607.

Winitz, H. Articulatory acquisition and behavior. New York: Appelton-Century-Crofts, 1969

Wolfram, W. Speech pathology and dialect differences. Dialects and educational equity. washington, D.C.: Center for Applied Linguistics, 1979.

Young, M.A. Articulation effort: transivity and observor agreement. Journal of Speech and Hearing Research, 1981, 24, 224-232.

Young, M.A. Observor agreement for marking moments of stuttering. Journal of Speech and Hearing Research, 1975, 18, 530-540.

Ysseldyke, J.E. & Algozzine, B. Bias among professionals who erroneously declare students eligible for special services. The Journal of Experimental Education, 1982, 50, 223-228.

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

Ysseldyke, J.E. Monograph #1: Research and development plan. Minneapolis, MN: Institute for Research on Learning Disabilities, 1978.

Ysseldyke, J.E., & Thurlow, M.L. The special education assessment and decision-making process: Seven case studies. Minneapolis, MN: Institute for Research on Learning Disabilities, 1980.

APPENDIX I

DIAGNOSTICIAN AND CHILD-SPECIFIC INTERVIEW QUESTIONS

DIAGNOSTICIAN INTERVIEW QUESTIONS (DGI)

Hello, my name is _____. I am working for Project Head Start which is conducting a study of speech impaired children served by Head Start. The Head Start Handicapped Coordinator scheduled this time with you to conduct an interview. I hope this time is still convenient for you.

This interview is scheduled in two parts, corresponding to the interview guides which you received from the Head Start Handicap Coordinator.

First I'll ask you some general questions about your experience and background in the area of speech.
Second, I'll ask some questions relating to your diagnosis for a specific Head Start child.

Do you have any questions before we begin?

1. What is your current professional title? If you have more than one, answer only in terms of services you provide to Head Start.

2. How many years have you been a (FILL IN TITLE FROM ABOVE), including this year?

3. In addition to this position have you had any other professional positions?

4. How many years of experience in other professional positions have you had?

5. What academic degree(s) do you have, in which field(s), and when did you receive each one?

6. Do you hold any current certificatiion(s) and/or license(s)?

7. What is your sex?

8. What is your race?

9. Did you receive the diagnostic criteria for speech impairment? (Transmittal notice #75.11 "Announcement of Diagnostic Criteria for Reporting Handicapped Children in Head Start")

CHILD SPECIFIC INTERVIEW QUESTIONS (CSI)

The following questions should be answered concerning one of the Head Start children whom you diagnosed as speech impaired. Specifically, child _____ (GIVE FIRST NAME AND CODE #). Your answers should concern this specific child only.

1. Give me a total picture of the diagnostic process you used to determine that this child was speech impaired. What diagnostic techniques did you use? Techniques is meant to refer to any informaiton gathering procedures you might have used with this child, including tests, observations, informal inventories, checklists and nonstandard procedures.

2. Do you feel that the test results reflect the child's actual abilities? If not, why not?

3. What were the speech problems you found? Describe the nature and severity level of the child's speech impairment as completely as possible. (RECORD COMPLETE NARRATIVE RESPONSE. REQUEST THAT ALL CASE FILE RECORDS BE SENT HERE.)

4. Did you consider any factors in addition to the test results in making your final diagnosis?

5. Does this child have other diagnosed handicaps?
   5a. What other diagnosed handicaps does this child have?

6. What is the race of this child?

7. What is the child's language or dialect preference? If the child has insufficient intelligible language to make this judgment then to which language/dialect is the child primarily exposed?

(REPEAT CSI FOR EACH CHILD IN THE SAMPLE WHO WAS DIAGNOSED BY THIS PERSON.)

CASE FILE REVIEW FORM

CASE FILE REVIEW

ID

TEST CODE

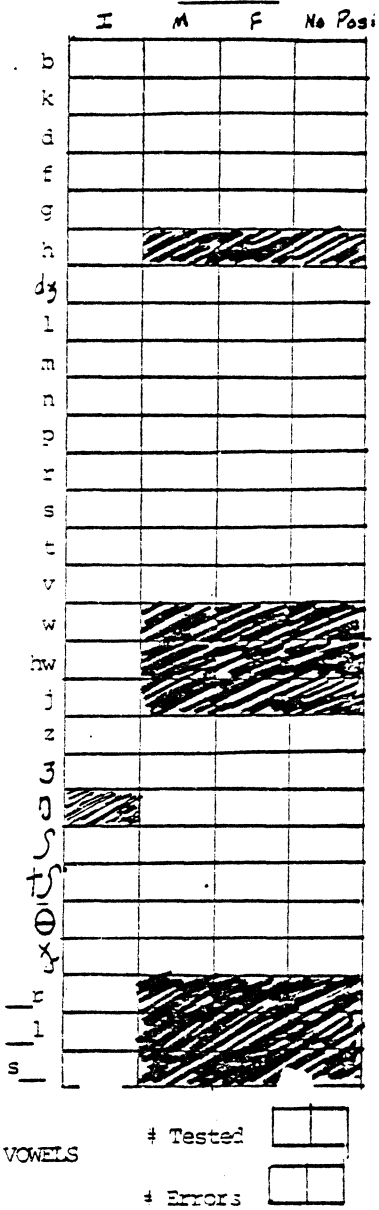
EVAL DATE

DOB

SEX M=1 F=2

Articulation Scores

0=correct production
1=substitution, _____
2=distortion
3=omission
9=not tested



VOWELS

Tested

Errors

CODES

1=normal
2=mild
3=mild to moderate
4=moderate
5=moderate to severe
6=severe
0=not judged

Severity

ARTIC

LANG

LANG. AGE

C.A (for L.A.)

ADDITIONAL INFORMATION

APPENDIX II

FREQUENCY COUNTS FOR ORIGINAL CATEGORIES OF EACH VARIABLE

ARTICULATION SEVERITY

ARTICSEV

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
MILD	2	32	9.3	9.3	9.3
MILD TO MODERATE	3	19	5.5	5.5	14.8
MODERATE	4	79	22.9	22.9	37.7
MODERATE TO SEVERE	5	59	17.1	17.1	54.8
SEVERE	6	156	45.2	45.2	100.0
	TOTAL	345	100.0	100.0	

MODE	6.000	RANGE	4.000	MINIMUM	2.000
MAXIMUM	6.000				
VALID CASES	345	MISSING CASES	0		

LANGUAGE SEVERITY

LANGSEV

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
NOT RATED	0	34	9.9	9.9	9.9
NORMAL	1	64	18.6	18.6	28.4
MILD	2	37	10.7	10.7	39.1
MILD TO MODERATE	3	45	13.0	13.0	52.2
MODERATE	4	50	14.5	14.5	66.7
MODERATE TO SEVERE	5	44	12.8	12.8	79.4
SEVERE	6	71	20.6	20.6	100.0
	TOTAL	345	100.0	100.0	

MODE	6.000	RANGE	6.000	MINIMUM	0.0
MAXIMUM	6.000				
VALID CASES	345	MISSING CASES	0		

ARTICULATION TEST

ARTICTES

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
ARIZONA ARTIC PS	2	42	12.2	12.2	12.2
DENVER SCREENING	15	2	0.6	0.6	12.8
HEJNA ARTIC TEST	18	4	1.2	1.2	13.9
FISHER LOGEMANN	24	6	1.7	1.7	15.7
GOLDMAN FRISTOE	26	189	54.8	54.8	70.4
PHOTO ARTIC TEST	46	52	15.1	15.1	85.5
ZIMMERMAN PLS	49	1	0.3	0.3	85.8
FLUHARTY SCREENING	57	3	0.9	0.9	86.7
TEMPLIN DARLEY	59	12	3.5	3.5	90.1
BAYLOR ARTIC TEST	358	3	0.9	0.9	91.0
WEISS ARTIC TEST	363	4	1.2	1.2	92.2
IOWA PRESSURE	364	1	0.3	0.3	92.5
BRINGLESON ARTIC TST	367	2	0.6	0.6	93.0
INFORMAL:TCHR-MADE	400	24	7.0	7.0	100.0
	TOTAL	345	100.0	100.0	

MODE	26.000	RANGE	398.000	MINIMUM	2.000
MAXIMUM	400.000				
VALID CASES	345	MISSING CASES	0		

OTHER FACTORS CONSIDERED IN THE DIAGNOSIS

OTHFACTR

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
CHILD'S MATURITY	1	20	5.8	5.8	5.8
WITHDRAWN BEHVAIOR	2	16	4.6	4.6	10.4
NO OTH SVCS AVAILAB	3	4	1.2	1.2	11.6
CHILD'S AGE	4	1	0.3	0.3	11.9
PARENT-TCHR PRESSURE	5	3	0.9	0.9	12.8
CHILD'S APPEARANCE	6	1	0.3	0.3	13.0
HOME SITUATION	7	22	6.4	6.4	19.4
SPCH FAMILY MEMBERS	8	13	3.8	3.8	23.2
STIGMA OF LABELS	13	2	0.6	0.6	23.8
NONE CONSIDERED	14	241	69.9	69.9	93.6
1 & 7	17	1	0.3	0.3	93.9
BOTH 1 & 8	18	2	0.6	0.6	94.5
BOTH 1 & 9	19	1	0.3	0.3	94.8
CASELOAD	22	1	0.3	0.3	95.1
BOTH 2 & 7	27	6	1.7	1.7	96.8
BOTH 2 & 8	28	1	0.3	0.3	97.1
BOTH 5 & 7	57	1	0.3	0.3	97.4
BOTH 7 & 8	78	9	2.6	2.6	100.0
	TOTAL	345	100.0	100.0	

MODE	14.000	RANGE	77.000	MINIMUM	1.000
MAXIMUM	78.000				
VALID CASES	345	MISSING CASES	0		

DXSEX

DIAGNOSTICIAN'S SEX

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
MALE	1	36	10.4	10.4	10.4
FEMALE	2	309	89.6	89.6	100.0
	TOTAL	345	100.0	100.0	

MODE	2.000	RANGE	1.000	MINIMUM	1.000
MAXIMUM	2.000				
VALID CASES	345	MISSING CASES	0		

DXRACE

DIAGNOSTICIAN'S ETHNICITY

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
BLACK	3	21	6.1	6.1	6.1
WHITE	5	324	93.9	93.9	100.0
	TOTAL	345	100.0	100.0	

MODE	5.000	RANGE	2.000	MINIMUM	3.000
MAXIMUM	5.000				
VALID CASES	345	MISSING CASES	0		

DIAGNOSTICIAN'S HIGHEST SPEECH-LANGUAGE DEGREE

DXDEGREE

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
NO SPEECH DEGREE	1	9	2.6	2.6	2.6
BACHELORS	2	65	18.8	18.8	21.4
MASTERS	3	255	73.9	73.9	95.4
DOCTORATE	5	16	4.6	4.6	100.0
	TOTAL	345	100.0	100.0	

MODE	3.000	RANGE	4.000	MINIMUM	1.000
MAXIMUM	5.000				
VALID CASES	345	MISSING CASES	0		

KNOWCRIT

AWARENESS OF HEAD START DIAGNOSTIC CRITERIA

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
UNAWARE OF CRITERIA	1	225	65.2	65.2	65.2
AWARE OF CRITERIA	2	120	34.8	34.8	100.0
	TOTAL	345	100.0	100.0	

MODE	1.000	RANGE	1.000	MINIMUM	1.000
MAXIMUM	2.000				
VALID CASES	345	MISSING CASES	0		

DIAGNOSTICIAN'S CERTIFICATION

CERTIFIC

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
CCC	1	89	25.8	25.8	25.8
GENERAL TEACHING	3	31	9.0	9.0	34.8
SPECIAL ED TEACHING	4	65	18.8	18.8	53.6
PRIVATE PRACTICE	5	9	2.6	2.6	56.2
CCC & SPECIAL ED	6	79	22.9	22.9	79.1
CCC & GENERAL TCHG	7	11	3.2	3.2	82.3
CCC & PRIVATE PRAC	8	40	11.6	11.6	93.9
SP ED & PRIV PRAC	9	5	1.4	1.4	95.4
NO SPEECH CERTIFIC	10	16	4.6	4.6	100.0
	TOTAL	345	100.0	100.0	

MODE	1.000	RANGE	9.000	MINIMUM	1.000
MAXIMUM	10.000				
VALID CASES	345	MISSING CASES	0		

YEAR DIAGNOSTICIAN RECEIVED LATEST SPEECH-LANGUAGE DEGREE

CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
54	1	0.3	0.3	0.3
57	1	0.3	0.3	0.6
58	8	2.3	2.3	2.9
60	3	0.9	0.9	3.8
62	2	0.6	0.6	4.3
63	9	2.6	2.6	7.0
64	3	0.9	0.9	7.8
65	2	0.6	0.6	8.4
67	10	2.9	2.9	11.3
68	11	3.2	3.2	14.5
69	7	2.0	2.0	16.5
70	1	0.3	0.3	16.8
71	40	11.6	11.6	28.4
72	20	5.8	5.8	34.2
73	23	6.7	6.7	40.9
74	13	3.8	3.8	44.6
75	32	9.3	9.3	53.9
76	45	13.0	13.0	67.0
77	24	7.0	7.0	73.9
78	10	2.9	2.9	76.8
79	29	8.4	8.4	85.2
80	22	6.4	6.4	91.6
81	27	7.8	7.8	99.4
82	2	0.6	0.6	100.0
TOTAL	345	100.0	100.0	

MODE	76.000	MINIMUM	54.000
MAXIMUM	82.000		
VALID CASES	345	RANGE	28.000
		MISSING CASES	0

CHILD'S SEX

CHILDSEX

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
MALE	1	230	66.7	66.7	66.7
FEMALE	2	115	33.3	33.3	100.0
	TOTAL	345	100.0	100.0	

MODE	1.000	RANGE	1.000	MINIMUM	1.000
MAXIMUM	2.000				
VALID CASES	345	MISSING CASES	0		

CHETHNIC

CHILD'S ETHNICITY

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
HISPANIC	1	14	4.1	4.1	4.1
BLACK	3	93	27.0	27.0	31.0
NATIVE AMERICAN	4	9	2.6	2.6	33.6
WHITE	5	229	66.4	66.4	100.0
	TOTAL	345	100.0	100.0	

MODE	5.000	RANGE	4.000	MINIMUM	1.000
MAXIMUM	5.000				
VALID CASES	345	MISSING CASES	0		

BEHAVIOR

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
NOT A PROBLEM	1	295	85.5	85.5	85.5
HYPERACT OR AGRESSIV	2	50	14.5	14.5	100.0
	TOTAL	345	100.0	100.0	

MODE	1.000	RANGE	1.000	MINIMUM	1.000
MAXIMUM	2.000				
VALID CASES	345	MISSING CASES	0		

CHLANG

CHILD'S PRIMARY LANGUAGE OR DIALECT

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
STANDARD ENGLISH	1	279	80.9	80.9	80.9
BLACK DIALECT	11	56	16.2	16.2	97.1
APPALACHIAN DIALECT	12	10	2.9	2.9	100.0
	TOTAL	345	100.0	100.0	

MODE	1.000	RANGE	11.000	MINIMUM	1.000
MAXIMUM	12.000				
VALID CASES	345	MISSING CASES	0		

CHILD'S OTHER HANDICAPS

OTHNDCP

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
VISUAL	2	1	0.3	0.3	0.3
HEARING	4	6	1.7	1.7	2.0
PHYSICAL	5	3	0.9	0.9	2.9
HEALTH	6	3	0.9	0.9	3.8
EMR	7	5	1.4	1.4	5.2
ED	8	6	1.7	1.7	7.0
LD	9	1	0.3	0.3	7.2
NONE	10	320	92.8	92.8	100.0
	TOTAL	345	100.0	100.0	

MODE	10.000	RANGE	8.000	MINIMUM	2.000
MAXIMUM	10.000				
VALID CASES	345	MISSING CASES	0		

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2