1992

Validity of discrepancy criteria for identifying children with developmental language disorders

Robin D. Morris
Georgia State University, robinmorris@gsu.edu

Dorothy M. Aram

Nancy E. Hall
University of Michigan, nhall@maine.edu

Recommended Citation
Aram, D., Morris, R. & Hall, N. (1992). Validity of discrepancy criteria for identifying children with developmental language disorders. Journal of Learning Disabilities, 25(9), 549554.

The Validity of Discrepancy Criteria for Identifying Children with Developmental Language Disorders

Dorothy M. Aram, Robin Morris, and Nancy E. Hall

Empirical data from two studies address the clinical validity of discrepancy criteria for identification of children with developmental language disorders (DLD). Study 1 involved 256 preschoolers clinically defined as DLD and meeting inclusionary criteria for normal hearing, intellectual, neurological, and psychiatric status. Application of alternative psychometrically derived discrepancy criteria identified only 40% to 60% of the clinically defined group as language disordered. Study 2 applied nonverbal IQ-language performance discrepancy criteria to 368 eight-year-old, randomly selected control subjects, resulting in over 45% of the controls being identified as DLD. Factors contributing to underidentification in Study 1 and over-identification in Study 2 are discussed, raising questions regarding the validity of discrepancy criteria for identification of DLD children.

The problem of appropriately defining developmental language disorders in children has led to difficulties in reliably identifying such disorders (Johnston, 1988; Stark & Tallal, 1981; Tomblin, 1983). In 1982, the American Speech-Language-Hearing Association (ASHA) developed the following definition for a language disorder:

> A language disorder is the impairment or deviant development of comprehension and/or use of a spoken, written, and/or other symbol system. The disorder may involve (1) the form of language (phonologic, morphologic, and syntactic systems), (2) the content of language (semantic system), and/or (3) the function of language in communication (pragmatic system) in any combination. (p. 949)

However, the applicability of this definition clinically or for research purposes is questionable. This limitation is indicated by the following statement accompanying the definition: "Various definition and eligibility criteria may exist for determining degree of handicap and disability compensation. The definition[s] in this document are not intended to address issues of eligibility and compensation" (p. 949). Thus, there is no universally agreed-upon definition of developmental language disorders; rather, the definitions used typically include outlines of both exclusionary and discrepancy criteria. For example, Carrow-Woolfolk and Lynch (1982) suggested that "clinical language disorder" refers to the "slow, limited, or deviant manner" (p. 284) of development of certain aspects of a native language, which leads to discrepancies among various skills. These authors go on to point out that their definition, which includes the word *clinical* as opposed to *developmental*, excludes those children whose language deficits might be accounted for by "obvious developmental disorders such as intellectual retardation, hearing loss, and emotional disturbance" (p. 284). There appears to be fair agreement regarding factors that should be excluded, even though the specific operational definitions for each criterion may differ somewhat. Most definitions of developmental language disorders exclude peripheral hearing loss, frank neuromuscular disorders, significant emotional disturbance, and mental retardation (e.g., Johnston, 1988; Tallal, 1988).

In addition to exclusionary criteria, it has also been advocated that discrepancy criteria be used (e.g., Stark & Tallal, 1981; Tallal, 1988). Although both clinical judgment and psychometric formulas have been used to determine discrepancy between language and other aspects of development, there appears to be an increasing trend to require the use of psychometric formulas (see McCauley & Demetras, 1990, for examples). For instance, a recent national survey of state educational agencies reported that at least 16 states require the use of a standardized discrepancy formula in the identification of children with language disorders (Nye & Montgomery, 1989). However, the method of determining discrepancy and the degree of discrepancy required among these states

JOURNAL OF LEARNING DISABILITIES
VOLUME 25, NUMBER 9, NOVEMBER 1992
PAGES 549-554

have been inconsistent, with a child's language performance variably compared to grade level, chronological age, mental age, or nonverbal intelligence. The survey indicated that when grade, chronological age, or mental age levels were the reference, discrepancy requirements ranged from V2 to 4 years. When nonverbal intelligence was used, the discrepancy required between IQ and the language measure ranged from 1 to 2 standard deviations (SD).

Despite the trend toward using discrepancy criteria to establish eligibility for language services, the practice has been criticized in many quarters (American Speech-Language-Hearing Association, 1989; Fletcher & Morris, 1986; Johnston, 1988; Nation & Aram, 1977; Tomblin, 1983). In March 1989, the American Speech-Language-Hearing Association's Committee on Language Learning Disorders reviewed the practice of basing decisions relative to eligibility for language intervention on discrepancy formulas. The committee concluded that the "exclusive use of discrepancy formula as a required procedure for determining eligibility for language intervention should be viewed with extreme caution and avoided whenever possible" (p. 115).

Although some clinicians appear to be moving away from the exclusive use of discrepancy criteria, the use of psychometrically determined discrepancy criteria to define research subjects appears to be increasing. For example, Tallal (1988), in her comprehensive summary of developmental language disorders and summary of research priorities presented to the National Conference of Learning Disabilities, stated,

> It is essential that in addition to demonstrating that language abilities of a child are significantly below what would be expected based on the child's chronological age, . . . [it is established that] they are also significantly discrepant from what would be predicted based on the child's mental abilities. (p. 211)

Despite the strong advocacy positions that individuals or groups have taken for or against the use of discrepancy criteria for defining developmental language disorders, surprisingly little empirical work has addressed the development, use, or validity of discrepancy formulas for identifying these children. Stark and Tallal (1981) are among the few who have reported their attempts to develop and apply standard exclusionary and discrepancy criteria for defining subjects with specific language disorders. Their investigation of 132 children (ages 4 years to 8 years 6 months) with language impairments included the following exclusionary criteria:

1. Hearing sensitivity greater than 25 dB across the frequencies 250 through 6000 Hz (although the investigators failed to indicate whether unilateral or bilateral results were required);
2. Presence of emotional/behavioral problems;
3. History of frank neurological deficits as reported by parent or teacher;
4. Performance IQ below 85 as measured by the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1963), or the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974);
5. Presence of speech motor deficits, determined by a standard oral peripheral examination administered by a speech-language pathologist.

The discrepancy criteria employed by Stark and Tallal (1981) required the children to have an overall language age that was at least 12 months below their chronological age (or their mental performance age, whichever was lower). In addition, children who demonstrated a speech articulation impairment that was more severe than their language disorder (i.e., greater than 6 months below estimated expressive language age) were excluded, as Stark and Tallal thought that this combination might indicate a mixed language-motor speech disorder. Similarly, children whose reading age was greater than 6 months below their language age were excluded, again assuming that co-occurring language and reading disorders suggested the presence of a mixed disorder.

Not surprisingly, the application of these stringent exclusionary and discrepancy criteria resulted in the identification of only one third of the 132 clinician-defined children with language impairments as meeting criteria for specific language deficits. The rationale underlying Stark and Tallal's (1981) selection of these particular criteria is debatable, but reporting the results of their attempt to apply explicit criteria is commendable. Equally informative are the results of the application of these criteria. This 2:1 mismatch between clinicians' judgments and criteria-based determination of specific language disorders suggests that something is wrong with the current processes used to identify children with language disorders.

The remainder of this article presents data from two studies that provide additional information addressing the validity of the application of discrepancy formulas for identification of developmental language disorders.

Nosology Project

The first study, "Nosology of Higher Cerebral Function Disorders in Children," was a multicenter program project sponsored by the Child Neurology Society and funded by the National Institutes of Health. Its primary aim was the development of empirically based, internally valid classification systems for developmental language disorders and for autism. During the first 3 years, the study attempted to (a) develop objective criteria to differentiate among developmental language disorders, autism, and nonautistic mental deficiencies and (b) identify subtypes of developmental language disorders and autism. A massive data base has been compiled for

each child, including historical and family information; neurological assessments; adaptive behavior measures; psychiatric evaluations; neuropsychological assessments; speech and language assessments; and 30-minute, videotaped play sessions from which measures of spontaneous language, play behavior, and sociability were derived.

All of the children referred to the study were considered to exhibit deficits in language development as determined by expert clinicians in the field of developmental language disorders (e.g., speech-language pathologists, neuropsychologists, neurologists, and developmental psycholinguists). To be included as subjects, individuals had to meet the following criteria:

a. Be between 3 years and 5 years 11 months of age;
b. Possess normal hearing acuity (defined as equal to or less than 20 dB at 1000 and 2000Hz bilaterally, and equal to or less than 25 dB at 500 and 4000Hz bilaterally);
c. Have been raised in a predominantly monolingual English-speaking home;
d. Exhibit the absence of frank neurological or orofacial deficits;
e. Exhibit normal nonverbal IQ, defined as greater than 80 on the Stanford-Binet Intelligence Scale-Revised (Thorndike, Hagen, & Sattler, 1986) Quantitative or Abstract/Visual Reasoning subtests;
f. Have not been identified as autistic by DSM-lli-R criteria (Wing, 1985).

The determination of discrepancy criteria was more problematic. Inclusion of a broadly defined population of children with developmental language disorders was seen as central to the purposes of the study. Recall that all children were referred to the study by an expert in language disorders and were considered to present developmental language disorders. Nonetheless, from a research perspective, an explicit operational definition for developmental language disorders was also considered to be essential for purposes of interinvestigator communication, objectivity, and replicability. Consistent with Tallal's (1988) recommendation, the application of objectively derived psychometric formula was planned. Because subjects were not yet school-aged, and grade-level comparisons could not be used, nonverbal intelligence (or mental age) and chronological age references provided alternatives against which to compare language performance. The use of absolute cutoff scores on specified language measures permitted a third approach to quantification of language deficit.

Nonverbal IQ-Language Discrepancy

The first approach to objective definition of developmental language disorders involved establishing a discrepancy between nonverbal IQ and a language measure. Following a review of the tools available, this approach was operationalized a priori and required a difference of at least 1 standard deviation between the score on one of the nonverbal subtests of the Stanford-Binet (which, as previously noted, had to be above 80), and the standard score on the Test of Early Language Development (TELD) (Hresko, Reid, & Hammill, 1981). Application of this Stanford-Binet-TELD discrepancy resulted in 131 (51%) of the 256 clinically defined children with language disorders being identified.

To further assess the use of a discrepancy approach in identifying children with language disorders, several alternative criteria using varying degrees of discrepancy between the Stanford-Binet and the TELD were examined using the current data (see Table 1). The most liberal alternative, based on the standard error of measurement of each measure, resulted in a minimal discrepancy requirement of greater than or equal to 7 points and identified 155 (61%) of the 256 clinically defined children with developmental language disorders. In contrast, the most restrictive definition used the individual residual score from a regression approach, which corrected for the high correlation (r = .77) between the Stanford-Binet and the TELD scores, and for the standard error of the residual. This approach required a 17-point discrepancy to classify children, and identified 101(39%) of the 256 clinically defined children with developmental language disorders.

In contrast to using standard scores, the use of mental-age (MA) and language-age (LA) equivalents derived from the Stanford-Binet and TELD, respectively, provided a further alternative. Requiring a discrepancy of more than 1 year, this approach resulted in the identification of 111(43%) of the 256 children as language disordered, thus falling between the least and the most restrictive criteria using standard scores. Therefore, given the least restrictive (155) and the most restrictive (101) criteria, there is nearly a 25% difference in the number of children identified as language disordered. Even with the least restrictive

TABLE 1
Alternative Discrepancy Criteria in Nosology Study

Stanford-Binet-TELD	Discrepancy	*n*	%of 256
1 standard deviation	15 pt.	131	51
Standard error	7 pt.	155	60
Regression	> 17 pt.	101	39
Mental age	> 1 yr.	111	43
Chronological/language age	> 1 yr.	129	50
Language cutoff TELD standard score	< 85	131	51

Note. TELD = Test of Early Language Development.

criteria, almost 40% of the clinically defined sample would not be identified as language disordered.

Chronological Age-Language Discrepancy

A second approach required at least a 1-year difference between the TELD LA and chronological age at time of test. This discrepancy criteria identified 129 (50%) of the 256 clinically defined group as language disordered.

Language Performance Cutoff Scores

A third approach used the TELD screening instrument as a measure of language performance and specified an absolute level, or cutoff score, below which a child's performance would be classified as language disordered. Using a cutoff level of lower than 85 (i.e., greater than **1** standard deviation) on the TELD, 131subjects (51%) would be classified as language disordered; a cutoff level of lower than 70 (i.e., greater than 2 standard deviations) identified 60 subjects, or 23% of the clinically defined sample.

Underidentification Using Discrepancy Formulas

Depending on the discrepancy formula applied, between 100 and 150, or approximately 40% to 60%, of the clinically defined sample of children with developmental language disorders fail to be identified using these criteria. Numerous factors would appear to contribute to this percentage of mismatch between clinical and research definitions of developmental language disorders, including limitations to professional judgment, measurement issues, and differences in the conceptualization of developmental language disorders. For instance, it may be that the referring professionals were simply wrong 40% to 60% of the time in defining children with language disorders. Alternatively, the professionals' clinical judgments may have been more sensitive than the psychometrically derived discrepancy formulas. A second factor that could account for these differences might involve measurement issues. For example, these results could relate to the specific tools used and their limitations in measuring nonverbal IQ and language, as well as statistical concerns regarding comparison of scores based on standardization derived from different groups. Further, theoretical concerns having to do with the relationship between language and other aspects of conceptual development, and the adequate measurement and interpretation of such, are inherently involved in the task of defining developmental language disorders. Finally, there may be a fundamental flaw with viewing children who simply perform on the lower end of the normal continuum on some language measure as having developmental language disorders (Leonard, 1991). **It** may be that. alternative definitions that use a discrepancy between different aspects of language, or variation in style or rate of language learning, may better identify developmental language disorders in children than comparisons between global language scores and performance IQ, mental age, or chronological age.

The Very l.Dw Birthweight Study

The second set of data come from a follow-up study of very low birthweight (VLBW, defined as < 1,500 grams at birth) children at 8 years of age (Aram, Hack, Hawkins, Weissman, & Borawski-Clark, 1991). The language assessment was only one small portion of an extensive outcome study, requiring an entire day of testing. The areas of outcome assessed included physical growth, neurological evaluation, vision and hearing, intelligence, academic achievement, speech and language, and behavioral adjustment. The subjects in this study included 249 VLBW 8-year-olds, which represented 79% of the original surviving birth cohort admitted to the neonatal intensive care unit of Rainbow Babies and Children's Hospital in Cleveland, Ohio, over a 3-year period, from 1977 through 1979.

The VLBW children were compared to a randomly selected, geographically based sample of normal birthweight children born between 1977 and 1979. The control subjects were selected from the total population of children born in the eastern part of Cuyahoga County (Cleveland, Ohio) who were enrolled in public and Catholic diocese schools. The original plan was to select 10 children from each of 50 randomly selected schools assigned to six strata on the basis of racial composition and median family income of the schools' catchment areas. However, because of extensive busing in Cleveland, the stratification scheme was used only for children enrolled in suburban and Catholic schools. A simple, random sample of children attending public schools on the east side of Cleveland was employed. A total of 643 children were thus randomly selected. One hundred twenty-four families declined to participate, and 156 either were born out of Cuyahoga County or were preterm. The control population thus included 363 normal birthweight, term children. Demographic characteristics for the two groups are given in Table 2. The only significant difference between the groups other than gestational age and birthweight was age, with the corrected age for the VLBW children being, on the average, 4 months younger than the normal birthweight controls.

As part of the more extensive testing protocol, the entire WISC-R and a language battery were administered. The language battery included the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981), the Token Test for Children (DiSimoni, 1978) (Parts 4 and 5), the Rapid Automatized Naming Test (RAN) (Denckla & Rudel, 1974), and the Sentences subtest of the Clinical Evaluation of Language Functions (Wiig & Semel-Mintz, 1980). To address the question of

whether VLBW is a risk factor for developmental language disorders, a child was identified as having a language disorder if a discrepancy of more than 1 standard deviation occurred between WISC-R Performance IQ and any of the five language comprehension or production measures provided the child also had normal intelligence (Performance IQ > 85), normal hearing in at least one ear, and the absence of major neurological deficits.

Table 3 summarizes the number of children in the VLBW and control groups who were identified as having a primary language disorder based on the discrepancy criteria, in comparison to those who demonstrated a discrepancy in conjunction with other abnormalities (e.g., Performance IQ < 85, hearing impairment, or neurological deficits) or no discrepancy. A discrepancy of 1 standard deviation or more between Performance IQ and language measures classified one third (33.7%) of the VLBW and nearly half (45.7%) of the randomly selected, normal birthweight controls as language disordered.

Given the extraordinarily high percentage of children classified as presenting language disorders by these criteria, a greater than 2 standard deviation discrepancy criterion between Performance IQ and any of the five language measures was applied, resulting in 14.5% of the VLBW children and 18.7% of the normal birthweight children being classified as language disordered. Even the use of a strict criterion of greater than 2 standard deviations discrepancy identified an astonishingly high number of both the VLBW and normal birthweight children as language disordered. This percentage far exceeds general estimates of the incidence of developmental language disorders, which are not greater than 3% to 5% (Leske, 1981).

TABLE 2
Demographic Characteristics of VLBW Children and Matched Controls

Demographics	VLBW	Geographic matched controls
Number tested	249	363
Age at test (years, months)*	8.6a	8.9
Sex (male)	50%	47%
Race (black)	56%	62%
Birthweight	1,177 grams	Sibs.
Gestational age (weeks)	29.7	37
Maternal education		
> high school	28%	32%
high school	55%	51%
< high school	17%	17%

Note. VLBW = very low b1rthweight.
**p<.05.*
acorrected age.

TABLE 3
Subjects Identified Using 1 or 2 Standard Deviation Discrepancy Criteria

	1 *SD* Performance IQ-language measures				2 *SD* Performance IQ-language measures			
	VLBW		Controls		VLBW		Controls	
Demographics	*n*	%	*n*	%	*n*	%	*n*	%
Language impairment	84	33.7	116	45.7	36	14.5	68	18.7
Discrepancy and other factors (IQ, neurological, sensory)	66	26.5	53	14.5	33	13.3	25	6.9
No discrepancy	99	39.8	144	39.7	180	72.3	270	74.4

Note. VLBW = very low birthwe1ght.

Overidentification Using Discrepancy Formulas

Several factors may have contributed to the high identification rate of both the VLBW and control subjects as language disordered. First, the incidence of a range of developmental problems generally is found to be substantially higher among inner city, minority children, who were disproportionally represented in the VLBW study. Second, it could be argued that requiring a discrepancy on only one of the five language measures results in an overinflation of the number identified as language disordered. On the other hand, frequently in practice, the identification of children with language disorders is based on the score of a single language measure (Nye & Montgomery, 1989). Furthermore, given the heterogeneous nature of developmental language disorders and the fact that each measure was selected to assess a different aspect of language, there is no reason to believe that a child need be deficient in more than one aspect of language to be considered language disordered. Irrespective of what factors may account for this high rate of identification, the usefulness of a definition of developmental language disorders that identifies almost half of the control subjects as disordered must be questioned.

In summary, data from these two studies further bring into question the clinical validity of definitions of developmental language disorders based on IQ-language discrepancy criteria. In the first study, 40% to 60% of the children considered clinically to be language disordered could not be classified as language disordered using these criteria. In the second study almost half of the randomly selected

control subjects were identified as language disordered. These data add to the growing recognition of the long-overdue need to rethink our operational definitions of developmental language disorders, and, in particular, the use of discrepancy criteria.

ABOUT THE AUTHORS

Dorothy M. Aram received her PhD in *speech pathology from Case Western Reserve Univer*sity in 1972. *She currently* is *a professor* in *the* Division *of Communication Disorders, Emerson College. Her research interests include language and learning sequelae of acquired lateralized brain lesions in children, and classification issues involving children with developmental language disorders. Robin Morris received his PhD in clinical psychology from the University of Florida in 1982.* His *research interests are* in *the area of developmental neuropsychology. Nancy E. Hall received her PhD in communication* dis*orders from Case Western Reserve University* in *1992. Her research interests focus on child language disorders and stuttering. Address: Dorothy* M. *Aram, Division of Communication Disorders, Emerson College, 168 Beacon St., Boston, MA 02116.*

AUTHORS' NOTE

The data reported here were derived from two studies supported by the National Institutes of Health: HD 20057 *and NS 20489.*

REFERENCES

American Speech-Language-Hearing Association. (1982). Definitions: Communicative disorders and variations. *ASHA, 24,* 949-950.

American Speech-Language-Hearing Association. (1989). Issues in determining eligibility for language intervention. *ASHA, 31,* 113-118.

Aram, D.M., Hack, M., Hawkins, S., Weissman, B.M., & Borawski-Clark, E. (1991). Very low birthweight children and speech and language development. *Journal of Speech and Hearing Disorders, 34,* 1169-1179.

Carrow-Woolfolk, E., & Lynch, J.l. (1982). *An integmtive approach to language disorders in children.* New York: Grune & Stratton.

Denckla, M.B., & Rudel, R.R. (1974). Rapid automatized naming of pictured objects, colors, letters, and numbers by normal children. *Cortex, 10,* 186-202.

DiSimoni, F. (1978). *The token test for children.* Allen, TX: DLM.

Dunn, L.M., & Dunn, L.M. (1981). *Peabody picture vocabulary test-Revised.* Circle Pines, *MN:* American Guidance Service.

Fletcher, J.M., & Morris, R.D. (1986). Oassification of disabled learners: Beyond exclusionary definitions. In S. Ceci (Ed.), *Handbook of cognitive, soc al, and neuropsychological aspects of learning disabilities* (Vol. 1, pp. 55-80). Hillsdale, NJ: Erlbaum.

Hresko, W.P., Reid, K.D., & Hammill, D.D. (1981). *Test of early language development.* Austin, TX: PRO-ED.

Johnston, J.R. (1988). Specific language disorders in the child. In N.J. Lass (Ed.), *Handbook of speech-language pathology and audiology* (pp. 685-715). St. Louis, MO: C.V. Mosby.

Leonard, L.B. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services* in *Schools,* 22, 66-74.

Leske, M.C. (1981). Prevalence estimates of communicative disorders in the U.S.: Speech disorders. *ASHA,* 23, 217-225.

McCauley, R.J., & Demetras, M.J. (1990). The identification of language impairment in the selection of specifically language-impaired subjects. *Journal of Speech and Hearing Disorders,* 55, 468-475.

Nation, }.E., & Aram, D.M. (1977). *Diagnosis of speech and language disorders.* St. Louis, MO: C.V. Mosby.

Nye, C., & Montgomery, J.K. (1989). Identification criteria for language disordered children: A national survey. *Hearsay: The Journal of the Ohio Speech and Hearing Association, 4,* 26-33.

Stark, R.E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46,* 114-180.

Tallal, P. (1988). Developmental language disorders. In J.F. Kavanagh & T.J. Truss, Jr. (Eds.), *Learning disabilities: Proceedings of the national conference* (pp. 181-272). Parkton, MD: York Press.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *Stanford-Binet intelligence scale-Revised.* Chicago: Riverside.

Tomblin, }.B. (1983). An examination of the concept of disorder in the study of language variation. *Proceedings From the Symposium on Research* in *Child Language Disorders, 4,* 81-109.

Wechsler, D. (1963). *Wechsler preschool and primary scale of intelligence.* San Antonio, TX: Psychological Corp.

Wechsler, D. (1974). *Wechsler intelligence scale for children-Revised.* San Antonio, TX: Psychological Corp.

Wiig, E.H., & Semel-Mintz, E. (1980). *Clinical evaluation of language functions.* San Antonio, TX: Psychological Corp.

Wing, L. (1985). *DSM-III-R: Autistic* dis*orders.* Unpublished manuscript.