Journal of Autism and Developmental Disorders, Vol. 14, No. 4, 1984

Outcomes of Severe Disorders of Language Acquisition[1]

Rhea Paul[2] and Donald J. Cohen
Yale University Child Study Center

Data on speech, language, performance IQ, school placement, and behavior are presented on 18 subjects diagnosed in childhood as "aphasic" and followed through adolescence. Results reveal that slow but steady growth in language is made, with expressive skills showing somewhat more rapid progress than comprehension. Performance IQ is highly correlated with language skills in later childhood and, along with receptive skill, is a good predictor of school placement. The diagnostic and prognostic implications of this information are discussed.

A century of clinical literature in neurology, psychiatry, and language disorders has described a range of children with profound impairments in language acquisition that cannot be accounted for by global mental retardation, serious emotional disorders, or sensory deficits. These children, variously described as suffering from specific language disorders, aphasia, or dysphasia, are identified during their preschool years, and their communication problems are known to persist, with varying degrees of severity, throughout childhood. However, very little systematic research has focused on long-term outcomes of these disorders, or on early predictors of later status.

Several studies present outcome data (King, Jones, & Lasky, 1982; Aram & Nation, 1980; Garvey & Gordon, 1973; Griffiths, 1969; Hall & Tomblin,

[1]We wish to thank Mary Ziltz for help in analyzing the data, and Barbara Caparulo and Bennett Shaywitz for their clinical collaboration. This research was supported by NIMH Mental Health Research Center Grant No. RR00125, NICHD Grant No. 03008, the Merck Foundation, and Mr. Leonard Berger.
[2]Address all correspondence to Dr. Rhea Paul, Yale Child Study Center, P.O. Box 3333, New Haven, Connecticut 06510.

1978). But these outcome reports are often confounded by the fact that subjects included those with both language disorders and milder delays in articulation development. These more mildly impaired children may skew the results of outcome studies. De Ajuriaguerra et al. (1976) examined prognosis for communication in "dysphasic" youngsters and found that they progressed in communicative ability, while their linguistic behavior remained essentially unchanged. Degree of intellectual and linguistic ability at the early evaluation were the best predictors of language progress at the second assessment 2 years later. The subjects in their study were between 10 and 12 years of age at the second evaluation. These results suggest that older adolescents would show little progress in language.

During the past decade, we have followed a cohort of severely language-impaired individuals. In this report, we describe the outcome of early childhood language disorders in this sample in relation to progress seen in several areas: language, behavior, and cognitive functioning. The results of these studies provide information not only on the prognosis of serious language learning difficulties but also on the issues involved in making diagnostic distinction among specific language disorders, mental retardation, and autism.

METHOD

Subjects

The present subjects are a subgroup of children with serious language disorders reported earlier (Paul, Cohen, & Caparulo, 1983). The original group comprised 28 subjects who had received a diagnosis of "childhood aphasia" during the last decade. All showed at least 9 months' delay in receptive language abilities at the time of their initial evaluation, when their average age was 6.5 (*SD* = 3.9, range = 2-21). Most were not speaking at all at that time, while the remaining 43% produced only one- to two-word sentences. All had normal hearing and were screened for the fragile X chromosome abnormality and found to be negative for the syndrome. At the time of their initial evaluation, the subjects received thorough psychological and biological assessments—which included IQ and standardized language testing and clinical observation of language and behavior—and they were rated on a series of behavioral rating scales. These methods have been reviewed in detail elsewhere (Caparulo & Cohen, 1983). Diagnoses were arrived at by having two clinicians, a child psychiatrist and a special educator—both experienced with developmental disabilities—evaluate the patients independently and reach a consensus on diagnosis. In addition, the clinicians' reliability on the Rimland E-2 checklist (Rimland, n.d.) and the Behavior Rating In-

Table I. Subject Ages and Performance IQs at Early (1) and Recent (2) Evaluations

Subject	Age at first evaluation (years-months)	Age at follow-up (years-months)	Years elapsed	Performance IQ at first evaluation (PIQ_1)	Performance IQ at recent evaluation (PIQ_2)
3	19-0	20-0	1.0	61	30
8	13-0	21-2	8.2	67	43
9	2-6	7-6	5.0	117	93
11	6-5	12-5	6.0	—	54
12	9-11	17-3	7.4	—	37
13	4-2	9-6	5.3	92	66
14	13-0	17-2	4.2	44	28
15	4-6	16-9	12.3	100	33
16	6-4	12-3	6.0	95	98
17	6-4	15-8	9.4	—	23
18	3-8	11-8	8.0	—	75
20	4-5	13-3	8.9	51	105
21	11-1	18-1	7.0	—	16
22	6-7	13-8	7.1	91	25
24	11-3	22-5	11.1	28	33
27	2-11	7-1	4.2	—	67
28	5-0	10-0	5.0	—	48
29	2-3	9-8	7.4	—	23
Mean	7.4	14.2	6.9		
SD	4.6	4.6	2.7		

strument for Autistic and Atypical Children (BRIACC; Ruttenberg, Dratman, Fraknoi, & Wenar, 1966) was assessed (see Cohen et al., 1978, for details). All subjects were enrolled in intensive special education programs. They have also been involved in a series of studies of the biological correlates of neuropsychiatric disorders of childhood (Cohen, Caparulo, Shaywitz, & Bowers, 1977; Waldo et al., 1978; Caparulo et al., 1981; Young, Kavanagh, Anderson, Shaywitz, & Cohen, 1982; Cohen, 1982). As part of their involvement in these research studies, subjects were evaluated periodically for language and cognitive performance. For the present study, all families in the original cohort who could be located were contacted and invited to bring their children back to participate in a follow-up evaluation. Eighteen of the original 28 families could be reached and agreed to participate. There were 6 females and 14 males in this sample. Their average age at the time of the follow-up was 14.2 (SD = 4.6, range = 7.1-22.4). Their mean age at the time of their first evaluation was 7.4 (SD = 7.4, range = 2.3-19) (see Table I). The average time lapse between their initial and most recent evaluation was 6.9 years (SD = 2.7, range = 1-12.3). T tests showed no differences in either age ($t = 1.35$, $p < .2$) or IQ ($t = .2$, $p < .3$) at the time of the initial evaluation between those subjects who participated in the follow-up and those from the original subject pool who did not.

The subjects in the earlier study were classified as having either developmental language disorders (DLD) without social deficits or "atypical" DLD (ADLD). The latter evidenced social withdrawal, poor or fleeting social relations, and some of the sensory and motor symptoms of autism, including rocking, stereotypic behavior, and unusual responses to stimuli. None of the ADLD subjects satisfied full diagnostic criteria for infantile autism, on either the Behavior Rating Instrument for Autistic and Atypical Children (Ruttenberg et al., 1966) or the Rimland E-2 checklist. Of the original 28 subjects, 14 showed developmental language disorders only (DLD group), and 14 showed this atypical profile in addition to their language deficits (ADLD group). In the follow-up sample, 11 DLD subjects and 7 ADLD subjects participated.

Procedure

The subjects were given a language and cognitive evaluation consisting of a series of standardized tests. A sample of conversation was recorded on videotape. Parents completed the revised Conners Parent Questionnaire (Goyette, Conners, & Ulrich, 1978) in order to rate the patient's behavior. (This questionnaire requests parents to rate their children's behaviors in terms of conduct, learning, hyperactivity, and anxiety on a 4-point scale from "not at all" to "very much." Norms are provided for making comparisons to the ratings given to normal children by their parents.) An assessment of speech-motor functioning and a test of visual-motor integration were given, and each subject's educational placement was noted. A variety of other information was also gathered from EEGs, brainstem auditory evoked responses, neurological examination, structured psychiatric interviews, and the Vineland Adaptive Behavior Scales, a recent revision (Sparrow, Balla, & Cicchetti, 1984) of Doll's Vineland Social Maturity Scale (1965). Results of these latter measures will be discussed in subsequent reports.

Procedures included the following:

Cognitive. The Leiter International Performance Scale (LIPS) (Authur, 1952) was given except in cases where records indicated the subject was likely to have a mental age greater than 13, since IQ for the Arthur Adaptation of the LIPS is normed only up to 13. In these cases, Raven's Standard Progressive Matrices (Raven, 1960) were used to assess cognitive level. These scores are referred to as the subjects' Performance IQ (PIQ).

Language Comprehension. Three measures of receptive language were obtained. The Peabody Picture Vocabulary Test (Dunn & Dunn, 1981) measures single word vocabulary. The Grammatical Understanding subtest of the Test of Language Development (Newcomer & Hammill, 1971) was used to assess syntactic comprehension. In cases where mental age was below 4 years, the Assessment of Children's Language Comprehension (Foster, Giddan, & Stark, 1973), which is appropriate for lower-functioning subjects, was given. The Auditory Reception subtest of the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968) was used as a measure of general receptive ability. All three of these measures yield age-equivalent scores. The age scores for the three were averaged to produce Language Reception Age, then divided by the subject's present chronological age to yield a Language Reception Quotient (LRQ). (This procedure for indexing receptive and expressive language levels is advocated by Stark & Tallal, 1981, to avoid basing summary scores on the results of one instrument, which may assess a limited area of language function. Data from Stark & Tallal, 1981, indicate that weighting scores from the various tests may give better predictions of mental age, but since at the time of this writing, no information on the validity of any particular weights in the prediction equation is available, a simple mean appears the safest way to synthesize the data. Admittedly, though, more sophisticated statistical techniques would be preferable.) This LRQ can be compared to the LRQ derived similarly from data from the earlier evaluations (see Paul et al., 1983, for a description of earlier data).

Language Production. Three standardized measures of expressive language were obtained. The Expressive One-Word Picture Vocabulary Test (Gardiner, 1979) measures naming ability. The Grammatical Completion subtest of the Test of Language Development (Newcomer & Hammill, 1971) measures the ability to add grammatical markers to words (I have a dress, and she has a dress. We have two — — — [dresses]). The Word Articulation subtest of the Test of Language Development gives a general measure of articulatory maturity. These measures also yield age-equivalent scores. The age scores for the three measures were again averaged to yield a Language Production Age, then divided by chronological age to produce a Language Production Quotient (LPQ). LPQ from the current evaluation can then be compared to LPQ at previous evaluations. When subjects had no speech but used sign instead, the vocabulary measure was administered in sign, but articulation and grammatical closure measures could not be administered in this mode; nor did the subjects ever combine more than two signs in one utterance, so that grammatical measures were not really relevant. For children with utterance lengths in spoken language of only one to two words as well, the grammatical measure was inappropriate since grammatical markers could not be expected to appear in utterances at this level. For 50% of the subjects, then, LPQ represents a measure of expressive vocabulary only, since this measure is the only available way to evaluate productive language level in children who do not speak, or speak so little as to make grammatical analysis irrelevant. Data from speaking and signing subjects were combined because sign is seen in this study as an alternative but analogous system of symbolic representation that may provide extended opportunities for expression to subjects whose oral-motor dysfunctions make speech extremely difficult to master. We believe underlying productive language competence can be manifested in either speech or signing behavior, and the two modes can be evaluated meaningfully using the same set of criteria.

Communication. In addition to standardized tests of language performance, a rating of communicative competence was made for each subject. This rating was made on the basis of an observation of a 10-minute sample of spontaneous conversation that was videotaped in which subjects interacted with their parents and the examiners. The rating consisted of dichotomous judgment, either "normal" or "atypical." "Normal" ratings were given in cases in which the child expressed three or more communicative intentions, such as requesting objects or actions, spontaneously naming objects, answering questions, or relating past events. In these communication ratings, the form of the intention—gesture, sign, single word, or complete sentence—was not considered. The only criterion for receiving a rating of "normal" communicative competence was that the subject spontaneously expressed several intentions to convey a meaningful message to the listener. Subjects received ratings of "atypical" if they failed to produce at least three meaningful communications during the 10-minute interview—that is, if they showed no interaction or if their responses consisted entirely of imitations of words or signs addressed to them. Subjects were also, however, given ratings of "atypical" if their conversation contained three or more bizarre or irrelevant remarks, despite the fact that they may have responded appropriately to some aspects of the interactions.

Two trained raters, blind to the subjects' diagnoses, made these judgments independently by viewing the videotapes. Interrater agreement for the 18 subjects was 83%.

Behavior. Parents of all subjects completed the Revised Conners Parent Questionnaire (Goyette et al., 1978). This measure requires parents to rate their children's behaviors on a 4-point scale from 0 (not at all) to 3 (very much). Individual behaviors are then grouped into five factors: conduct disorders, learning disorders, psychosomatic complaints, impulsivity-hyperactivity, and anxiety. Scores on each factor are compared to norms adjusted for age and sex. Scores falling above the cutoff point for the norming sample are considered positive for that factor.

Oral Structure and Function. The structure and function of the oral mechanism, including lips, tongue, and palate, and the ability to repeat series of syllables ([bʌ], [dʌ], [gʌ]) were examined by a certified speech-language pathologist. Although formal norms are not available for these ratings, the tasks used are those generally expected to be mastered by normal children 7 or more years of age. Functioning in each of the three articulators was rated on a 3-point scale. Syllable repetition was rated on a 4-point scale. A speech-motor rating was derived by adding the subscores for each of the four areas, so that the highest possible score was 13. Interrater reliability for this score was determined by having a second speech-language clinician rate recordings of the subjects' performance. Reliability between the two clinicians, counting discrepancies of 1 point as agreement, was 100% on a 29% randomly chosen sampling of the ratings. Four of the subjects refused to cooperate with this examination, so data were unavailable for 22% of the sample.

Visual-Motor Skill. The Test of Visual-Motor Integration (VMI; Beery & Buktenica, 1967) was administered to all subjects. This test requires the subject to copy figures increasing in complexity.

Educational Placement. The school setting in which each child was placed was rated on a 5-point scale from least to most restrictive as follows: 1 = regular classroom, 2 = regular classroom with tutoring, support services; 3 = self-contained resource room in a public school; 4 = special day school for the handicapped; 5 = special school with residential placement.

RESULTS AND DISCUSSION

Behavioral Ratings

Conduct disorders, psychosomatic complaints, and anxiety were noted by parents on the Connors Questionnaire in 0%, 10%, and 20% of the subjects, respectively. Since these areas were cited as concerns relatively rarely,. they were not pursued in any further analyses. A majority of the parents reported difficulties in learning and hyperactivity, but because 78% of the subjects function in the retarded range, the fact that they have learning difficulties cannot be surprising. Hyperactivity, though, does seem to represent a distinct component of the language-disordered child's difficulties. Hyperactivity scores on the Connors Questionnaire were above the normal cut off point in 55% of the subjects and were, then, the only aspect of these behavior ratings included in the present analysis.

The remaining data were compared in three ways. First, differences in PIQ, LRQ, and LPQ scores from the early and recent evaluations were examined. Second, correlations among measures derived from the recent evalution were calculated. Finally, differences among the subgroups within the sample were explored.

Comparisons Over Time

Paired t tests were used to look for changes in PIQ, LRQ, and LPQ from early to later evaluations. These comparisons are presented in Table II. (Individual PIQ scores, including missing values, can be seen in Table I. There were no missing LRQ and LPQ values.) Neither PIQ nor LRQ showed significant changes ($t = 1.73$, $p < .12$; $t = 1.17$, $p < .26$, respectively), but LPQ increased significantly in the interval between evaluations ($t = 2.68$, $p < .02$). It has been reported in a cross-sectional study of children between 3 and 9 years of age (Eisenson, 1972) that performance IQs of young, school-aged "aphasic" subjects were significantly lower than those of preschool aphasics. These data imply a decrement in PIQ over time, but our longitudinal data fail to confirm this observation. Average IQ scores in the current study are lower at the more recent evaluation, but not significantly so. It should be noted that PIQ scores from early evaluations were available for only 10 of the subjects (see Table I), so the failure to find significant differences may be due to the small size of the sample. Also, the difference in the age between Eisenson's cross-sectional sample (3-9 years)

Table II. Paired t Tests of Differences in Scores at Early (1) and More Recent (2) Evaluations

Measure	Mean score (and *SD*) $Time_1$	Mean score (and *SD*) $Time_2$	*N*	t	Significance level
PIQ	74.6 (28.6)	49.8 (28.0)	10	1.73	<.12 *n.s.*
LRQ	47.7 (22.2)	41.2 (26.6)	18	1.17	<.26 *n.s.*
LPQ	24.9 (13.5)	42.3 (31.9)	18	2.68	< .02

and the present group of subjects (7-22 years) may be a factor in this discrepancy in findings.

The fact that LRQ scores do not change significantly over time indicates that children are making progress in receptive skills, and that this progress is commensurate with increments in age. These data are encouraging in that it appears that growth in receptive language continues, even into adolescence, since the average age of the sample is over 14 years.

The significant increase in LPQ in the sample is even more encouraging. These subjects are making progress in expressive language at a rate exceeding their growth in age. It seems quite likely that these accelerated rates of growth in expressive skills are related to educational intervention, since expressive skills are more amenable to intervention than is comprehension, an essentially private event. Some of the improvement may be related to increased expressive vocabulary and sentence length in those subjects taught to communicate in sign. Twenty-eight percent of the subjects made some use of sign in their expressive language. These data argue for the continuation of language training in adolescence and perhaps beyond since language-disordered subjects can continue to make important gains. They differ from the findings of de Ajuriaguerra et al. (1976), who saw little change in linguistic behavior in their subjects in the 10- to 12-year age range. Still, for subjects in the present study, language skills remain seriously deficient. Only 11% of the sample had language quotients above 75. Linguistic skills remain for the most part primitive, but growth does continue from year to year.

Correlations Among Measures

Table III presents the matrix of correlations among the measures derived from the current evaluation. (Correlations of hyperactivity and school placement with the other measures are negative since high scores on these measures represent poorer functioning.) PIQ correlates highly with LRQ ($r = .91, p < .001$) and LPQ ($r = .82, p < .001$), as these latter measures do with each other ($r = .91, p < .001$). PIQ and LRQ also relate moderately to school placement ($r = -.60, p < .03$; $r = -.64, p < .02$, respectively). Hyperactivity scores do not correlate significantly with any of the other measures. Scores on oral motor functioning correlate moderately with scores on the visual-motor assessment ($r = .57, p < .04$) but not significantly with any other measures.

Since scores on formal language and performance IQ tests are highly correlated, the present data support de Ajuriaguerra et al. (1976) in the notion that intellectual capacity is a good predictor of language status. PIQ and receptive skill seem to be more important determinants of school placement than is speech. Children with good understanding of language can func-

Table III. Correlations Among Measures at the Recent Evaluation

	LRQ	LPQ	Hyper-activity	School placement	Speech motor rating (n = 14)
PIQ (n = 18)	.91[a]	.82[a]	−.22	−.60[a]	.43
LRQ (n = 18)		.91[a]	−.25	−.64[a]	.51
LPQ (n = 18)			−21	−.50	.46
Hyperactivity (n = 18)				.23	.07
School placement (n = 18)					−.33
VMI (n = 18)					.57[a]

[a]$p < .05$.

tion in less restrictive settings even when their expressive skills are less advanced.

Hyperactivity did not relate significantly to any of the other measures, although it was more frequently noted by parents than any other behavior problem. The Conners scale, geared toward normally developing children, may not be sensitive enough to behavioral disorders in this population to reveal the other relations among language and behavioral disturbances. Results of analysis of the more intensive structured psychiatric interviews with patients and their parents may shed more light on this question.

The failure to find correlations between speech-motor functioning and other measures, on the other hand, does seem to indicate a relatively independent oral-motor component in the speech problems of these subjects. The significant correlation between speech-motor rating and scores on the VMI may indicate that speech-motor deficits are associated with a general difficulty in fine motor development that might be thought of as a form of developmental dyspraxia. Previous reports on this sample (Paul et al., 1983) indicate a preponderance of fine motor difficulties. The role of this dyspraxic component in disorders of language learning warrants further investigation, particularly in children whose expressive skills lag far behind their comprehension of language.

Comparison Among Groups

Stability of Diagnoses. Communication ratings, based on the videotaped interactions, were used to determine whether subjects had remained within the DLD and ADLD categories to which they were originally assigned. All 11 subjects originally diagnosed as DLD received communication ratings of "normal" from at least one of the diagnosis-blind raters (82% received normal ratings from both). All 7 of those originally diagnosed as ADLD received "atypical" ratings from at least one of the raters (86% received atypical

Table IV. Mean Scores (and Standard Deviations) for Subgroups on Each Measure

	Measure					
Subgroup	PIQ	LRQ	LPQ	Hyper-activity	School placement	Speech motor rating
High IQ DLD (n = 4)	92.8 (12.8)	80.8 (17.1)	84.3 (30.5)	1.6 (1.0)	2.3 (1.5)	12.3 (1.0)
Low IQ DLD (n = 7)	38.4 (16.6)	34.9 (16.7)	31.6 (20.7)	1.2 (.7)	4.4 (.8)	7.0 (4.5)
ADLD (n = 7)	36.7 (16.6)	24.9 (10.7)	29.0 (21.6)	2.0 (.6)	4.4 (.5)	9.8 (2.1)

ratings from both). These data indicate that the subjects' social-communication skills remained relatively stable over time. Those seen as atypical, or autisticlike, during childhood continued to exhibit sparse or deviant patterns of communication. On the other hand, those whose difficulties were more restricted to language, rather than social communication, remained more likely to initiate some form of meaningful communication and did not develop bizarre uses of speech or gestures. On the basis of these data, then, the subgroups identified by the original diagnoses were retained.

Formation of Subgroups. A t test between the DLD and ADLD groups showed significant differences in terms of performance IQ ($t = 1.9$, $p < .04$). Since any differences observed between these two groups might then be attributable simply to IQ differences, the DLD subjects with IQs in the normal range were separated out. Thus, three subgroups were formed: DLD with high IQs ($n = 4$), DLD with low IQs ($n = 7$), and ADLD subjects ($n = 7$). There were no significant differences in IQ between the low IQ DLD group and the ADLD subject ($t = .2$, $p < .3$). The mean scores for each of these three groups on each of the measures are presented in Table IV.

Analysis of Variance. Significant differences among the three subgroups were found in scores on PIQ ($F = 18.70$, $p < .001$), LRQ ($F = 19.48$, $p < .001$), LPQ ($F = 8.34$, $p < .003$), and school placement ($F = 5.4$, $p < .02$). There were no significant differences among the subgroups on hyperactivity ($F = .86$, $p < .44$) or speech motor ratings ($F = 3.44$, $p < .07$) (see Table V). Because each of these measures represents a different scale, it was not possible to test for interactions in this analysis.

Duncan's Multiple Range Test, which controls error rates at different levels depending on the number of means between the members of each pair being compared, was used as a post hoc instrument to examine pairwise differences among the three subgroups for the measures on which F tests were significant.

As the t test showed, the high IQ DLD group differed on PIQ scores from the low IQ DLD and the ADLD, using Duncan's procedures, but the two low IQ groups did not differ from each other ($p < .05$). There were significant differences in LRQ and LPQ between the high IQ DLD and the other two groups, but neither receptive nor expressive language skills differed significantly ($p < .05$) between the low IQ DLD and ADLD groups, and scores in these subgroups are seriously depressed, with means between 24 and 34. (It should be noted, too, that PIQ scores for the low IQ DLD and ADLD groups averaged only 36-39.) Similarly, school placement scores differed significantly between the high IQ DLD and the two low IQ groups, but these latter two were not different from each other ($p < .05$). Eighty-six percent of the low IQ DLD group and all the ADLD subjects found placement in highly restrictive special day or residential settings, while only 50% of the high IQ groups were placed outside of public schools.

IQ, then, appears to account for a good deal of the differences observed among the three diagnostic groups. The low IQ DLD group differs from the

Table V. Summary of Results of Analysis of Variance Examining Differences Among the Three Subgroups (High IQ DLD, Low IQ DLD, ADLD)

Variable	N	F value	Significance level
PIQ	18	18.70	<.001[a]
LRQ	18	19.48	<.001[a]
LPQ	18	8.34	<.004[a]
Hyperactivity	18	0.86	<.44
School placement	18	5.40	<.02[a]
Speech motor rating	14	3.44	<.07

[a] $p < .05$.

ADLD only in terms of greater degree of the communicative intent expressed by the DLD group. Subjects who were identified as ADLD at the earlier evaluations persisted in evidencing weak communicative intentions or bizarre uses of language, and made few gains in appropriate, spontaneous communication, despite steady growth in sentence length and vocabulary, as evidenced by their LPQ scores.

Diagnostic and Prognostic Issues

All subjects in this study were originally diagnosed as "aphasic," yet clear and important differences emerge among the three subgroups that fall out of the follow-up. If diagnosed today, without knowledge of their history, few, if any, would be given this diagnostic label. The high IQ DLD group would probably be called "learning-disabled," since they have school learning problems, evidenced by their school placement scores, in spite of normal PIQs and relatively intact expressive and receptive language scores. It is interesting to note that the four subjects in the high IQ DLD group all displayed the deficits in language organization and word finding so typical of the learning-disabled population (Wiig & Semel, 1980). Their narratives were perseverative and disorganized, and their speech frequently included word substitutions ("Then Papa Porridge came in"). One of these subjects was completely unable to construct a connected narrative, despite relatively good performance on formal language tests. When asked to explain how to play his favorite game, he replied, after several tries, "I can't explain things."

The ADLD group would not be easily distinguished from adolescents diagnosed in childhood as classic infantile autistics, since they show the oddities of communication, failure of communicative intent, language deficits, and social withdrawal that are characteristic of residual autism in adolescence. The low IQ DLD group's most notable trait now would be their intellectual retardation, since their present LRQ and LPQ scores are very close to their PIQs and are no longer dramatically depressed relative to nonverbal mental age level. Two factors in this sample, then, appear to have had great predictive significance: PIQ and social competence. Children with normal PIQs generally developed near-normal language, at least in terms of basic grammatical processes, while children with low PIQs eventually showed language performance that approximated their PIQ scores.

Do these data imply that PIQ can be taken as a sufficient prognostic indicator in early childhood? Probably not. Since 44% of this sample was untestable on PIQ measures at the early evaluations, it is difficult to know the direction of causality in the finding that language scores generally come to resemble PIQs. Results of PIQ testing may be confounded by the fact that language disabilities place a ceiling on achievement on performance IQ tests because of the verballike nature of the classification, sequencing, and coding tasks involved in even nonverbal test items. Because of the small size of the subgroup in this sample with valid PIQ scores from the early evaluations, the course of change in PIQ and its role in the limitation of eventual language status cannot be fully described.

The predictive role of early social skills for eventual language development in this sample is not very strong, since both the ADLD and DLD subjects with low IQs showed similarly poor language outcomes. It does seem clear from these data, though, that early social competence is a good predictor of later social and communicative development. The fact that the ADLD group remained seriously deficient in communication throughout the decade during which they were followed also argues for the prognostic value of early social skills. These results suggest that any child showing both language and social impairments in the preschool period should be given early, intensive educational intervention, focusing especially on social-communicative skill, even if the social deficits are mild compared to those seen in classic autistics.

The findings of this study raise several nosological issues. First, how are the children called "ADLD" here to be classified? According to current DSM-III criteria (American Psychiatric Association, 1980), they would probably be labeled Atypical Pervasive Developmental Disorder since they show "distortions in the development of multiple basic psychological functions that are involved in the development of social skills and language and that cannot be classified as either Infantile Autism or Childhood Onset Pervasive Developmental Disorder." This classificatory scheme would capture their resemblance to autistic children with regard to the outcome of the disorder, but it leaves many questions unanswered. For example, why should there be such a large group (half our original sample) who "fall through the cracks" of the current diagnostic system and need to be placed in this grab-bag category? Second, are these patients autistic or not? Should they be included in the studies of biological correlates of autism, for example? One reason for these problems may be a lack of information about the natural history of these disorders. The present data suggest that outcomes for children with social deficits showing a broad range of severity, including the relatively mild ones seen here, look quite similar in later childhood and adolescence. Outcomes do not appear to vary much whether a child shows clear infantile autism, "Childhood Onset PDD," or a form of the disorder too mild to be considered either. A system, then, that treated this disease as a spectrum with degrees of severity may be more valid than the current scheme, which rests primarily on an age of onset criterion ("infantile" vs. "childhood onset").

A second nosological issue raised by the present data concerns the DLD category. DSM-III criteria for specific developmental language disorders state that they are "not due to Mental Retardation," and yet most of the DLD subjects in this sample did function in the retarded range on IQ measures. There are some children with sustained, profound specific language disorders in the presence of normal IQs, but as Eisenson (1972) points out, this condition is quite rare, especially after the preschool period. DSM-III criteria will need to change inorder to capture this more common situation, in which children with cognitive delays show deficits in language that are more severe than their general mental retardation. One solution would require that DLD children show receptive or expressive language skills that were 1 or more years below mental age level, as measured by a nonverbal intelligence test. Using this criterion would allow both retarded and cognitively normal children with deficits in language to be considered within the DLD category. Retarded children would then receive MR as a diagnosis on Axis I and SDLD on Axis II. This system would more realistically reflect the scope of disorders seen not only in the current sample but also by researchers such as Miller, Chapman, and McKenzie (1981), who find that 50% of their heterogeneous retarded subjects show language delays in excess of their cognitive limitations.

The current data suggest that the overall prognosis for this population is quite guarded. Children with strong intellectual endowment can eventually master basic syntax, but some language and learning problems remain. Others of these subjects remain severely impaired in language, cognitive, and in some cases, social development. Outcomes in psychiatric and adaptive status and neurological predictors of outcome will be discussed in forthcoming reports. These current results argue for the efficacy of special education, since slow but steady progress is seen in language skills even in adolescence. But the serious degree of impairment that persists in many cases, despite intensive intervention, argues for the necessity of greater understanding of the organic roots of these disorders, so that treatments that impinge more directly on these underlying mechanisms can be ultimately be identified.

REFERENCES

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, D.C.: Author.

Aram, D., & Nation, J. (1980). Preschool language disorders and subsequent language and academic difficulties. *Journal of Communication Disorders, 13,* 159-170.

Arthur, G. (1952). *Arthur Adaptation of the Leiter International Performance Scale.* Washington, D.C.: Psychological Services Center.

Beery, K., & Buktenica, N. (1967). *Developmental Test of Visual Motor Integration.* Chicago: Follett.

Caparulo, B. K., & Cohen, D. J. (1983). Developmental language studies in the neuropsychiatric disorders of childhood. In K. E. Nelson (Ed.), *Children's language* (Vol. 4, pp. 423-463). Hillsdale, New Jersey: Erlbaum.

Caparulo, B. K., Cohen, D. J., Rothman, S. L., Young, J. G., Katz, J. D., Shaywitz, S. E., & Shaywitz, B. A. (1981). Computed tomographic brain scanning in children with developmental neuropsychiatric disorders. *Journal of the American Academy of Child Psychiatry, 20,* 338-357.

Cohen, D. J. (Ed.). (1982). Clinical research in child psychiatry [Whole issue]. *Schizophrenia Bulletin, 8.*

Cohen, D. J., Caparulo, B., Gold, J. R., Waldo, M. C., Shaywitz, B. A., Ruttenberg, B. A., & Rimland, B. (1978). Agreement in diagnosis. *Journal of the American Academy of Child Psychiatry, 17,* 589-603.

Cohen, D. J., Caparulo, B., Shaywitz, B., & Bowers, M. (1977). Dopamine and serotonin metabolism in neuropsychiatrically disturbed children: CSF homovanillic acid and 5-hydroxy-indoleacetic acid. *Archives of General Psychiatry, 34,* 545-550.

de Ajuriaguerra, J., Jaeggi, A., Guignard, F., Kocher, F., Maquard, M., Roth, S., & Schmid, E. (1976). The development and prognosis of dysphasia in children. In D. Morehead & A. Morehead (Eds.), *Normal and deficient child language* (pp. 345-386). Baltimore: University Park Press.

Doll, E. A. (1965). *Vineland Social Maturity Scale.* Circle Pines, Minnesota: American Guidance Services.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test.* Circle Pines, Minnesota: American Guidance Services.

Eisenson, J. (1972). *Aphasia in children.* New York: Harper & Row.

Foster, R., Giddan, J., & Stark, J. (1973). *Assessment of children's language comprehension.* Palo Alto: Consulting Psychologists Press.

Gardiner, M. (1979). *Expressive One-Word Picture Vocabulary Test.* California: Academic Therapy Publications.

Garvey, M., & Gordon, N. (1973). A follow-up study of children with disorders of speech development. *British Journal of Disorders of Communication, 8,* 17-28.

Goyette, C. H., Conners, C. K., & Ulrich, R. T. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology, 6,* 221-236.

Griffiths, C. P. S. (1969). A follow-up study of childen with disorders of speech. *British Journal of Disorders of Communication, 4,* 46-56.

Hall, P., Tomblin, B. (1978). A follow-up study of children with articulation and language disorders. *Journal of Speech and Hearing Disorders, 43,* 227-241.

King, P. R., Jones, C., & Lasky, E. (1982). In retrospect: A fifteen year follow-up report of speech-language disordered children. *Language, Speech and Hearing Services in the Schools, 13,* 24-32.

Kirk, S., McCarthy, J., & Kirk, W. (1968). *Illinois Test of Psycholinguistic Abilities* (Rev. ed.). Urbana: University of Illinois Press.

Miller, J., Chapman, R. S., & McKenzie, H. (1981). *Individual differences in the language acquisition of mentally retarded children.* Paper presented at the Second International Congress for the Study of Child Language, Vancouver, August.

Newcomer, P. L., & Hammill, D. D. (1971). *Test of language development.* Austin, Texas: Empiric Press.

Paul, R., Cohen, D. J., & Caparulo, B. K. (1983). A longitudinal study of patients with severe developmental disorders of language learning. *Journal of the American Academy of Child Psychiatry, 22,* 525-534.

Raven, J. C. (1960). *Standard progressive matrices.* London: H. K. Lewis.

Rimland, B. (n.d.). Form E-2. *Diagnostic Checklist for Behavior Disturbed Children.* San Diego: Institute for Child Behavior.

Ruttenberg, B. A., Dratman, M., Fraknoi, S., & Wenar, C. (1966). An instrument for evaluating autistic children (BRIACC). *Journal of American Academy of Child Psychiatry, 5,* 453-478.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, Minnesota: American Guidance Service.

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

Waldo, M., Cohen, D. J., Caparulo, B., Young, J., Prichard, S., & Shaywitz, B. (1978). EEG profiles of neuropsychiatrically disturbed children. *Journal of the American Academy of Child Psychiatry, 18,* 656-668.

Wiig, E. H., & Semel, E. M. (1980). *Language assessment and intervention for the learning-disabled.* Columbus, Ohio: Charles E. Merrill.

Young, J. G., Kavanagh, M. E., Anderson, G. M., Shaywitz, B. A., & Cohen, D. J. (1982). Clinical neurochemistry of autism and associated disorders. *Journal of Autism and Developmental Disorders, 12,* 147-166.