Journal of Consulting and Clinical Psychology
2007, Vol. 75, No. 4, 594–604

Copyright 2007 by the American Psychological Association
0022-006X/07/$12.00 DOI: 10.1037/0022-006X.75.4.594

Patterns of Growth in Verbal Abilities Among Children With Autism Spectrum Disorder

Deborah K. Anderson, Catherine Lord, and Susan Risi
University of Michigan

Pamela S. DiLavore
University of North Carolina, Chapel Hill

Cory Shulman
Hebrew University

Audrey Thurm
National Institute of Mental Health

Kathleen Welch
University of Michigan

Andrew Pickles
University of Manchester

Verbal skills were assessed at approximately ages 2, 3, 5, and 9 years for 206 children with a clinical diagnosis of autism (n = 98), pervasive developmental disorders-not otherwise specified (PDD-NOS; n = 58), or nonspectrum developmental disabilities (n = 50). Growth curve analyses were used to analyze verbal skills trajectories over time. Nonverbal IQ and joint attention emerged as strong positive predictors of verbal outcome. The gap between the autism and other 2 groups widened with time as the latter improved at a higher rate. However, there was considerable variability within diagnostic groups. Children with autism most at risk for more serious language impairments later in life can be identified with considerable accuracy at a very young age, while improvement can range from minimal to dramatic.

Keywords: autism, PDD, growth trajectories, verbal age equivalent, verbal skills

Communication impairment is a core feature of autism (American Psychiatric Association, 1994) and one of the most frequent reasons for referral among children who are later diagnosed with autism spectrum disorder (ASD; Tager-Flusberg, Paul, & Lord, 2005). Moreover, the accumulated evidence from longitudinal studies of young children with and without an ASD indicates that those with impairments in verbal skills are at increased risk for less favorable outcomes later in life. Early language abilities predict social functioning, academic achievement, and psychiatric outcome in late childhood and adulthood (e.g., Beitchman et al., 2001; Gillberg & Steffenburg, 1987; Kobayashi, Murata, & Yoshinaga, 1992; Venter, Lord, & Schopler, 1992). Long-term outcome is more limited for language-impaired children with ASD compared with language-impaired children with nonspectrum disorders (Howlin, Mawhood, & Rutter, 2000). These findings have direct implications for quality of life and the financial costs of services for such individuals. Hence, a growing body of research and theory seeks to better understand the ways in which the verbal skills of children with ASD might be improved or maximized through the identification of key factors affecting verbal outcome.

A handful of prospective studies have examined potential factors believed to facilitate or hamper the development of verbal skills in ASD. Joint attention, probably the most widely studied predictor of outcome, refers to "behaviors used to follow or direct the attention of another person to an event or object to share an interest in that event or object" (Siller & Sigman, 2002, p. 77). Theory suggests that a child's ability to initiate and respond to bids for joint attention is essential to the development of social-cognitive and verbal abilities (Mundy & Neal, 2001). Empirical findings lend support to this claim. In one study, joint attention had the strongest relation to concurrent language ability relative to other social skills such as social orienting and attention to distress (Dawson et al., 2004). Of seven longitudinal studies, all found a positive association between one or more aspects of joint attention (alternating gaze, pointing, showing) and subsequent verbal skills regardless of the measures used (e.g., Charman et al., 2003; Mundy, Sigman, & Kasari, 1990; Sigman & McGovern, 2005). The time between initial and follow-up assessments varied from 1 to 15 years, and from preschool to adulthood. The relationship held

Deborah K. Anderson, Catherine Lord, and Susan Risi, Autism and Communication Disorders Center, University of Michigan; Pamela S. DiLavore, TEAACH Division, University of North Carolina at Chapel Hill; Cory Shulman, School of Social Work and Social Welfare, Hebrew University, Jerusalem, Israel; Audrey Thurm, Pediatric and Developmental Branch, National Institute of Mental Health; Kathleen Welch, Center for Statistical Consultation and Research, University of Michigan; Andrew Pickles, School of Medicine, University of Manchester, Oxford, England.

This study was supported by National Institute of Mental Health Grants MH57167 and MH066469 and National Institute of Child Health and Human Development Grant HD 35482-01 awarded to Deborah K. Anderson. Parts of this work were presented at the International Meeting for Autism Research, Boston, Massachusetts, May 2005. We thank Steve Heinz and Shanping Qiu for data management assistance.

Correspondence concerning this article should be addressed to Deborah K. Anderson, University of Michigan, Autism and Communication Disorders Center, 1111 E. Catherine Street, Ann Arbor, MI 48109-2054. E-mail: debcarl@umich.edu

regardless of initial language level (e.g., Sigman & McGovern, 2005).

Controlling for initial language or verbal level makes it possible to assess both the unique contributions of other key factors that may be related to early verbal skills as well as the stability of verbal scores over time. When initial verbal IQ scores are not available, many studies have used nonverbal IQ as a control because it is usually highly correlated with concurrent verbal skills (e.g., Bono, Daley, & Sigman, 2004). Prospective studies have been fairly consistent, despite the use of different measures and outcomes, in documenting a strong positive correlation between initial nonverbal or verbal competence and subsequent verbal skills 2–8 years later (e.g., Charman et al., 2005; Stevens et al., 2000; Stone & Yoder, 2001; Turner, Stone, Pozdol, & Coonrod, 2006) to more than 15 years later (Mawhood, Howlin, & Rutter, 2000; Sigman & McGovern, 2005), though there are a few counter examples (e.g., Charman et al., 2003; Mundy et al., 1990). The lack of association in some studies may have been due to small sample size, measurement error, or real changes in verbal skills among very young children, thus making initial measurements less reliable predictors of various long-term outcomes (Lord & Schopler, 1989).

Some research has tested the effect of symptom severity at early ages on subsequent verbal skills. For children with autism, greater impairment at age 3 in the areas of restricted and repetitive behaviors and socialization have been associated with poorer language outcomes at age 7 (Charman et al., 2005). Moreover, a diagnosis of autism, compared with the broader diagnosis of pervasive developmental disorders-not otherwise specified (PDD-NOS) for preschool-age children, predicted poorer language outcome 2–3 years later (Charman et al., 2003), even after accounting for initial nonverbal and language scores (Thurm, Lord, Lee, & Newschaffer, in press). In yet another study (Mawhood et al., 2000), a high-functioning autism group was more impaired on language and verbal IQ measures at age 7.5 years than children with receptive language disorder only. Although the autism group enjoyed gains in their verbal skills by 24 years of age, they had significantly more problems in the use and understanding of language than the non-ASD group with a language disorder. Studies with nonspectrum comparison groups most always report less optimal outcomes for children with ASD than those with nonspectrum disabilities, such as language and other developmental delays (e.g., Mawhood et al., 2000; Sigman & Ruskin, 1999).

Other studies have tested the effects of various types of interventions on verbal skills in ASD. Although it is not always clear whether improvement is due to the treatment, maturation, or other factors, research has reported positive effects for interventions that vary in type and scope, such as those specifically targeting joint attention, communication skills, as well as more general education programs (e.g., Howlin & Rutter, 1989; McGee, Krantz, & McClannahan, 1985; Rogers, 2005; Smith, Groen, & Wynn, 2000; Turner et al., 2006). However, there is a need for better information on the characteristics of different subgroups and trajectories of development so that early intervention efforts can be optimally formulated and targeted. The emphasis placed on various skills may need to vary considerably for different 2- and 3-year-olds with ASD.

Regardless of the predictors considered, prospective studies generally report overall gains in abilities for ASD children (e.g., Mawhood et al., 2000; Sigman & Ruskin, 1999), with considerable variability in outcomes, from large gains to decreases in verbal skills over time relative to age norms (e.g., Mawhood et al., 2000; Stevens et al., 2000; Turner et al., 2006). Although these studies provide a foundation for studying the development of verbal abilities in ASD populations, some methodological challenges are evident, most notably, sample size restrictions. Much of the research to date involves sample sizes well under 50, making it difficult to identify subgroups within ASD and to assess the relative importance of multiple predictor variables.

In addition, few studies of verbal skill development in autism populations involve samples of children less than age 3 at the initial assessment. Theories concerned with the neuroplasticity of a child's brain from birth to preschool contend that this is a dynamic and critical period of development (e.g., Mundy & Neal, 2001). From this perspective, typically developing children are born with relatively unorganized synaptic connections and a predisposition toward orienting to social stimuli. In the first few years of life, these connections are pruned and shaped by input from the outside environment. At the same time, the child's preferences influence the kind of response he or she receives from the social environment. For a child with ASD, an initial deficit in joint attention may set in motion a cumulative, negative feedback cycle in which an "impoverishment" of social information prevents healthy neurological development that would ordinarily provide the foundation for subsequent social cognition, behavior, and verbal skills. However, the high degree of give-and-take between biological and environmental factors most typical of early childhood may allow greater responsivity to the influence of the environment (i.e., greater development and growth). Hence, the ability to detect and prevent potential long-term impairments in early childhood may have important implications for intervention.

The unique characteristics of the data for the present study allowed for the simultaneous consideration of multiple factors affecting long-term verbal skills among various diagnostic groups, beginning when the children were still in the early stages of acquiring language. Specifically, we prospectively examined the rate and pattern of growth in verbal abilities between the ages of 2 and 9 years for children who differed with respect to early diagnosis (autism, PDD-NOS, or nonspectrum disabilities). This study had two main objectives. The first was to contrast the differences *between* diagnostic groups in the development of verbal skills. We were interested in the degree to which age 2 predictors were related to outcome through age 9. Children diagnosed with autism at age 2 were expected to have fewer verbal skills than PDD-NOS and nonspectrum children at intake, with group inequalities increasing over time. Nonetheless, all groups were expected to show gains with respect to verbal age equivalence scores even though relative improvement for the children with fewer verbal abilities would not likely keep pace with age norms for typically developing children. Age 2 nonverbal age equivalent (NVAE) and joint attention were expected to contribute positively to the level and rate of improvement in subsequent verbal skills.

The second, more exploratory objective was to longitudinally examine differences *within* diagnostic groups in verbal abilities. We hypothesized considerable variability in outcomes within the autism and PDD-NOS groups, including a narrowing of the gap from typical development for a substantial proportion of ASD children from ages 2 to 9. In other words, we expected diagnosis

or symptom severity and other characteristics to be predictive but not definitive of later outcome. We expected the least variability in the autism group on the basis of previous analyses of this sample up to age 5 (Taylor, Pickering, Lord, & Pickles, 1998). In light of theory and research suggesting substantial change in early childhood verbal abilities, we used change itself as a predictor of age 9 outcome for children with ASD. More specifically, we hypothesized that changes in key predictors (i.e., verbal and nonverbal skills, symptom severity) between the ages of 2 and 3 would contribute to the age 9 verbal age equivalent over and above cross-sectional data obtained at age 3.

Method

Participants

Eligible participants were consecutive referrals younger than 37 months of age from agencies across North Carolina and metropolitan Chicago serving very young children with developmental delays. All 221 families agreed to participate in the study initially. One later withdrew and six other families became ineligible for inclusion when the children reached the age of 36 months before the first assessment could be scheduled. With the exception of the children's age, these seven families did not differ demographically from the other 97% of families who participated in the study. Participants consisted of 192 children (162 males, 30 females), referred for evaluation for possible autism, and 22 nonautistic developmentally delayed (DD) children (10 males, 12 females) recruited in North Carolina. The autism referral group comprised children under age 3 from four North Carolina state-funded autism centers ($n = 111$) or a Chicago autism clinic within a private university hospital ($n = 81$). Exclusion criteria included moderate to severe sensory impairments or cerebral palsy, known genetic abnormalities, or poorly controlled seizures. Nearly one half the 214 participants (47%) received a diagnosis of autism at age 2, whereas the other half was divided between children with PDD-NOS (28%) and those without ASD (25%). The nonspectrum group consisted of children with some degree of mental retardation or a language delay (91%), whereas the remainder had other disabilities such as attention-deficit/hyperactivity disorder (ADHD) or a medical condition. There was a mix of children from rural and urban locations. Ethnic minorities, the vast majority of whom were African American, accounted for a sizable proportion of the sample (33%). A third of the children had mothers with a high school education or less, whereas parent education for the remainder of the sample ranged from some college (29%) to completion of a college degree (38%).

All children were first seen at approximately age 2, with a mean age of 29 months ($SD = 5.17$). The number of subsequent assessments varied by site and referral status. The North Carolina autism referral group was assessed at approximately ages 2, 3, 5, and 9, whereas the other two groups were seen at three of the four time points (i.e., the DD group was not seen at age 3, and Chicago referrals were not seen at age 5). Of the original 214 participants, 5 were lost to follow-up after the initial assessment; another 37 were lost by age 9 because of geographical location, unreachable status, or refusal to participate. Although African American families with less education were lost to the study at a significantly higher rate than Caucasians and families with more education, attrition was not related to original diagnosis, gender, or verbal or nonverbal IQ. The present study includes the 206 children who were seen multiple times (at least twice) and were administered diagnostic and psychometric instruments from the list described below.

Measures

Diagnostic instruments. The Autism Diagnostic Instrument-Revised (ADI-R; Lord, Rutter, & LeCouteur, 1994) is a comprehensive, standardized parent interview designed to distinguish children with ASD from non-ASD and DD populations. Algorithm scores are totaled for each of three domains, consistent with those outlined in the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (American Psychiatric Association, 1994): social behaviors, communication, and repetitive interests. Parents were administered the ADI-R in the final two assessments when the children were ages 5 and 9. The ADI-R distinguishes between participants who have acquired sufficient verbal skills to be scored on language usage questions (daily, functional use of three-word phrases that sometimes include a verb for a code of "0") and those who have not (mostly single words for a code of "1"; fewer than five words used on a daily basis for a code of "2"). A toddler version of the ADI-R, which includes a number of additional items specific to the first 2 years of life, was given when the children were 2- and 3-years-old (see Lord, Shulman, & DiLavore, 2004).

The Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) and one of its predecessors, the Pre-Linguistic Autism Diagnostic Observation Schedule (PL-ADOS; DiLavore, Lord, & Rutter, 1995), acquire diagnostic information through direct observation of the child by a trained clinician. An administration consists of a series of socially constructed play tasks administered in a semistructured manner. An algorithm calculates summary scores for the social and communication domains. Children in the present study were given the PL-ADOS at ages 2 and 3, which was scored using the algorithm for the Module 1 ADOS (for children without phrase speech). The ADOS was administered in subsequent assessments. Children were observed with one of three modules, depending on whether they used single or no words (M1), phrase speech (M2), or fluent speech (M3).

Each member of the research clinical team established interrater reliability exceeding 90% exact agreement ($\kappa > 0.70$) for all items on the ADI-R and 80% exact agreement ($\kappa > 0.60$) on codes for the PL-ADOS and ADOS for three consecutive administrations before the study began. Reliability was maintained through consensus coding approximately every sixth administration with a second rater who was blind to referral status.

Psychometric instruments. The Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984) is a standardized parent interview of the child's everyday adaptive functioning, which yields domain scores in the areas of communication, daily living skills, social skills, and motor development, as well as an adaptive behavior composite score. Adequate validity and reliability have been established for both the adaptive behavior composite and the individual subscales of the VABS (Sparrow et al., 1984).

The Infant Mullen Scales of Early Learning (Mullen, 1985) is a normed measure of verbal and nonverbal cognitive abilities for children below the 36-month level in the Gross Motor, Fine Motor, Nonverbal Cognition (i.e., "Visual Reception"), Receptive Lan-

guage, and Expressive Language domains. The standard version of the Mullen (Mullen, 1989) for children below the 60-month level was administered to older children who were unlikely to achieve a basal score on cognitive tests of a higher skill level such as those described below. (For specific procedures used to determine the appropriate test for a child, please see the *Verbal and nonverbal test selection* section.)

The Differential Ability Scales (DAS; Elliott, 1990), another measure of verbal and nonverbal cognitive abilities, was normed on a national sample of children between the ages of 2 years, 6 months and 17 years, 11 months. If children in the present study met age requirements and seemed able to achieve a basal score, then they were administered the school-age version. This version includes the Conceptual Similarities and Word Definition subtests of verbal ability along with four nonverbal reasoning and spatial abilities subscales. If the school-age version (DAS-S) was not appropriate, then children were given the preschool version (DAS-P). Its verbal cluster includes Verbal Comprehension and Naming Vocabulary subtests along with three additional subtests for measuring nonverbal ability.

The Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991) was only administered to the highest functioning children at age 9. The WISC-III measures cognitive abilities for children who score developmentally between the ages of 6 years and 16 years, 11 months. Five subtests are mandatory for computation of summary verbal scores, including information about common events, places, and people; similarities in word pairs; comprehension; vocabulary; and arithmetic problems answered orally. Like the other psychometric instruments previously described, the WISC-III measures "language" skills with respect to syntax and semantics as well as other verbal skills not typically included in linguistic studies (e.g., tests of memory, general knowledge, and conceptual and quantitative abilities expressed orally). Five additional subtests measure different facets of nonverbal cognitive skills.

Joint attention measure. This measure was based on factor analyses of the ADOS and consisted of summed scores from several items depending on developmental level (see Gotham, Risi, Pickles, & Lord, 2007). For children with some speech, the composite variable includes the gestures, pointing, showing, and spontaneous initiation of joint attention items. For nonverbal children, the measure comprises the gestures, use of other's body to communicate, response to name, and response to joint attention items. Item scores ranged from 0 to 2 or 3. The composite variable was reverse coded so that higher scores indicate greater skill.

Procedure

The full battery of diagnostic and psychometric instruments was administered at each assessment free of charge. Verbal feedback and a written report were provided to families. Written informed consent was obtained from parents prior to each assessment, as indicated by the University's Institutional Review Board.

Diagnostic classification. The ADI-R parent interview was typically administered in the home, whereas direct observation of the child took place at a nearby clinic or, in later assessments, at school with the parent observing. Following the two-part evaluation, an overall best estimate diagnosis of autism, PDD-NOS, or other nonspectrum disability was jointly determined by two trained staff members, including doctoral-level research associates, child psychiatrists, or clinical psychologists. The decision for the best estimate diagnosis was based on the clinicians' summary reports, psychometric and diagnostic algorithm scores, as well as videotapes of the direct observation of the child. (For a more detailed description of the procedures, see Lord et al., 2006.) *Diagnosis* in the present study refers to the age 2 best estimate classification.

Verbal and nonverbal test selection. The "preferred" verbal age equivalent at each measurement varied according to the child's developmental level and was selected from the list of psychometric instruments described above. In order to determine the most appropriate cognitive ability test, children's level of functioning was initially estimated from verbal and adaptive skill scores on the VABS parent interview, administered prior to the direct testing of the child. Test selection followed a hierarchy from the most to least difficult to complete: WISC-III, DAS-S, DAS-P, and the Mullen. Children were administered the test with the greatest degree of difficulty appropriate to their estimated abilities. If either basal or ceiling scores could not be obtained, then participants were given the next test up or down in the hierarchy. Verbal ability was then constructed by averaging the age equivalence scores from the various verbal skill subtests of the chosen cognitive test in the hierarchy. Age 2 verbal abilities were scored exclusively from the Mullen ($n = 206$). The remaining age cohorts were split between various tests in the hierarchy: 179 Mullens and 1 DAS-P for age 3; 82 Mullens, 41 DAS-Ps, and 3 DAS-Ss for age 5; 55 Mullens, 41 DAS-Ps, 29 DAS-Ss, and 39 WISC-IIIs for age 9. We used verbal age equivalents because a majority of the children were language delayed and many could not be assigned a standard score. Moreover, age equivalents are more appropriate indices of change compared with standard scores, which are intended to show stability (Charman & Howlin, 2003). Because we elected to use verbal scores from developmental and intelligence tests, and collinearity within domains was very high (Elliott, 1990; Taylor et al., 1998), we did not attempt to separate receptive and expressive language abilities. NVAEs were scored from the Mullen for all 206 participants at age 2.

Analyses. Growth curve analysis with SAS Proc Mixed (SAS for Windows release 9.1.3) was used to examine longitudinal growth in verbal age equivalents from approximately age 2 to age 9. In the present study, a random (separate) intercept and slope were calculated for each child as a control for the high correlations among repeated measures on the same individuals over time. We used the growth models to compare the three diagnostic groups with respect to (a) the relative starting points at 24 months of age (i.e., the intercept); (b) the rate of change from age 2 to age 9 (i.e., the slope); and (c) the pattern of change from age 2 to age 9 (i.e., linear vs. quadratic). Because data for more than three time points were available, higher order polynomials were also considered in order to determine the shape of the curve. Covariates were added as fixed effects to examine how much they accounted for variation in the intercept and slope: age at testing, gender, ethnicity, mother's education, site, and age 2 nonverbal and joint attention skills. We tested the effects of these variables in three steps. Controlling only for basic demographic variables, we first wanted to examine group differences in early verbal skills and rate of change over time by diagnostic category before adding the other covariates of interest. In the second step, we added joint attention and NVAE to the model to see whether they accounted for group differences in

verbal abilities at the first assessment. Age was centered at 24 months so the intercept could be interpreted as the mean verbal age equivalent at 2 years old. Finally, we added the time interactions with these two covariates to see how the slope or rate of change in verbal skills over time was affected.

The estimates for both the covariance and beta parameters were obtained by restricted maximum likelihood methods (REML) so that the results would be less biased (Verbeke & Molenberghs, 2000). To test for group differences in slopes and intercepts, we used *t* statistics reported for each parameter, calculated as the ratio of the parameter estimate divided by the standard error. To examine whether rate of change in the verbal age equivalents over time differed significantly from zero, we used *t* tests for linear combinations of variables representing the slopes.

A power analysis was carried out for a comparison of the growth curve slopes for the diagnostic groups from age 2 to age 9 (Diggle, Heagerty, Liang, & Zeger, 2002). For an effect size of .15 with 80% power, and sample sizes of 50 children in each of the PDD and nonspectrum groups and 100 participants with autism, we estimate that we can detect a difference in slopes of .33 units per month between groups (i.e., the difference in slopes between the autism and PDD groups in our data), assuming a residual variance of 25, a correlation among observations on the same participant of .68 (again generated from our data), and an alpha of .05.

A mixture modeling procedure called TRAJ (Jones, Nagin, & Roeder, 2001) was used to focus more specifically on the variability in outcomes within the ASD subsample. Intended to complement growth curve analysis, PROC TRAJ is an exploratory procedure written for use in SAS that identifies linear and nonlinear patterns in longitudinal data and classifies the sample into groups. We ran the procedure using an uncensored, normal distribution for the age equivalent scores to see whether distinct subgroupings would emerge within ASD. We compared the absolute value of the Bayesian information criterion (BIC) between respective models (smaller indicates a better fit) to determine the optimal number of groups. (See Jones et al., 2001, for use of the BIC for nonnested model selection.) The contribution of factors related to assignment to groups with varying growth trajectories was also assessed. Because group membership changed slightly, depending on which covariates or "risk factors" were included, we could not compare nested models. For covariates, we therefore used *t* tests for the individual parameter estimates and change in the absolute value of the overall BIC as criteria for improvement of model fit. Odds ratios (ORs) were calculated for the parameter estimates so that the relative contributions of risk factors could be assessed.

Finally, least square regressions were conducted in order to examine whether changes in key predictor variables (i.e., verbal and nonverbal age equivalents, restricted and repetitive behaviors, ADOS social-communication total algorithm score, and joint attention) between the ages of 2 and 3 contributed significantly to the age 9 verbal age equivalent over and above cross-sectional data obtained at age 3. The change variables were added into the second block.

Results

Table 1 is a general approximation of expressive language level by diagnosis for children who were seen at age 9, where language level was measured by which ADOS module a child was able to take and whether the child had sufficient language to be administered questions about language deviance in the ADI-R. Almost one quarter of the autism group and more than half the PDD-NOS and nonspectrum children were using complex sentences to talk about topics outside of the immediate physical context by age 9 (the ADOS definition of *fluent*). Another 24%–31% of the children in each group were using sentences but were not fluent speakers. Finally, 29% of the autism group and fewer than 10% of the PDD-NOS and nonspectrum children were still nonverbal (i.e., using fewer than five words daily according to the ADI-R). The analyses that follow examined how children within each diagnostic group arrived at their age 9 or most recent level of verbal ability as measured at the time of their last assessment.

Between-Group Differences

Reduced model. Table 2 includes the growth curve models that tested for changes in verbal abilities from 2 to 9 years of age. Coefficients of special interest for each model are in bold. First, note in Model 1 that the intercept of 9.4 months is the average verbal age equivalent score for the sample at age 2. Also, the positive and significant effect of age indicates gains in verbal age equivalents over time for the sample as a whole. Significant unexplained variance in the random slopes and intercepts remains after accounting for age. This reduced model serves as a basis for comparison with subsequent models as we consider the contribution of other factors.

Differences by diagnosis. Model 2 of Table 2, depicted graphically in Figure 1A, tested for differences in verbal abilities over the 7-year period according to age 2 diagnosis. As expected, the PDD-NOS and nonspectrum groups differed from children with

Table 1
Expressive Language Level at Age 9 by Age 2 Diagnosis: Percentage of 172 Participants

Language level	Autistic (n = 84)	PDD-NOS (n = 46)	Nonspectrum (n = 42)
Complex sentences (ADOS Module 3)	23.8	58.7	54.8
Sentences but not fluent (ADOS Module 2)	23.8	26.1	31.0
Words but not sentences (ADOS Module 1; ADI-R = 1)	23.8	10.9	7.1
No or few consistent words (ADI-R = 2)	28.6	4.3	7.1

Note. Four children were not administered ADOSs; level of language was inferred from ADI, Vineland, and best verbal IQ scores. PDD-NOS = pervasive development disorders-not otherwise specified; ADOS = Autism Diagnostic Observation Schedule; ADI-R = Autism Diagnostic Instrument-Revised

Table 2
Growth Models for Changes in Verbal Age Equivalents from Age 2 to Age 9 (n = 206)

Predictor	Model 1		Model 2		Model 3		Model 4	
	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Fixed effects								
Intercept	9.37***	0.50	6.72***	0.91	8.55***	0.93	8.38***	0.93
Age at testing	0.59***	0.03	0.45***	0.06	0.45***	0.06	0.50***	0.06
Age 2 diagnosis								
Autism		—	—	—	—	—	—	—
PDD-NOS				1.18	2.10	1.24	2.31	1.23
Nonspectrum				1.44	2.11	1.57	2.37	1.56
Site								
N.C. possible autism referral			−1.47	0.94	−1.37	0.88	−1.44	0.88
N.C. developmentally delayed			4.67**	1.73	4.67**	1.60	4.58**	1.60
Chicago			—	—	—	—	—	—
Non-White			−1.38	0.92	−1.20	0.84	−1.20	0.84
Mother's education			−0.26	0.37	−0.33	0.34	−0.35	0.34
Nonverbal AE at age 2							0.08	0.06
Joint attention							0.90***	0.19
Linear slopes								
Age × autism			—	—	—	—	—	—
Age × PDD-NOS				0.10	0.35***	0.10	0.28**	0.09
Age × Nonspectrum				0.11	0.33**	0.10	0.23*	0.10
Age × NVAE								
Quadratic slopes								
Age^2			−0.00	0.00	−0.00	0.00	−0.00	0.00
Age^2 × Autism			—	—	—	—	—	—
Age^2 × PDD-NOS					−0.02	0.07	−0.03	0.07
Age^2 × Nonspectrum					−0.20*	0.07	−0.20*	0.07
Random effects								
Intercept variance	32.64***		14.77***		11.83***		11.10***	
Slope variance	0.23***		0.21***		0.21***		0.15***	
Intercept/slope covariance	0.30		−0.14		−0.58**		−0.38*	

Note. Quadratic term coefficients are multiplied by 100. PDD-NOS = pervasive development disorders-not otherwise specified; N.C. = North Carolina; AE = age equivalent; NVAE = nonverbal age equivalent; dashes indicate reference category; boldface indicates fixed effects of interest in each model.
* $p < .05$. ** $p < .01$. *** $p < .001$.

autism in a number of respects. First, the main effect coefficients for diagnosis show that differences in verbal abilities were already significant by age 2, with the nonspectrum and autism groups scoring the highest and lowest, respectively. Second, the Group × Time interactions were also significantly different, with the PDD-NOS and nonspectrum children improving at a faster rate over time compared with the children with autism. In other words, group inequalities increased with time.

It was expected that group differences would remain after controlling for other factors such as ethnicity, mother's education, and site. Results supported this prediction. Gender was always nonsignificant when included. Not predicted, but interesting, was the finding that improvements in verbal abilities began to level off for the nonspectrum group before age 9, whereas both ASD groups continued to improve without slowing through age 9. This decelerating pattern for the nonspectrum group was both significantly different from zero (not shown in Table 2), $t(254) = -3.40$, $p < .001$, and from the other two groups.

Covariates affecting the starting point. Model 3 adds several covariates in order to account for the observed variability in the intercepts, or diagnostic group differences in verbal age equivalents at 24 months. NVAE and joint attention at the first assessment were both significant ($p < .001$) and associated with superior verbal abilities at 24 months. When the effects of these two covariates were accounted for, the diagnostic group differences in verbal age equivalents at 24 months became nonsignificant. Closer inspection revealed that group differences became nonsignificant with the addition of joint attention alone. This was not the case for NVAE, though its effects remained positive and significant with or without the inclusion of joint attention in the model. Moreover, when either joint attention or NVAE were excluded from Model 3, the model with joint attention but not NVAE had a better fit than the model with NVAE but not joint attention. (BICs were 4853.2 and 4886.0, respectively.) Hence, consistent with our expectations, NVAE, and joint attention in particular, reduced group inequalities in age 2 verbal age equivalents.

Covariates affecting the rate of change. Model 4 tested the hypothesis that the same covariates accounting for intercept differences would also account for variation in the slopes or rate of growth in verbal abilities over time by diagnosis. Although the

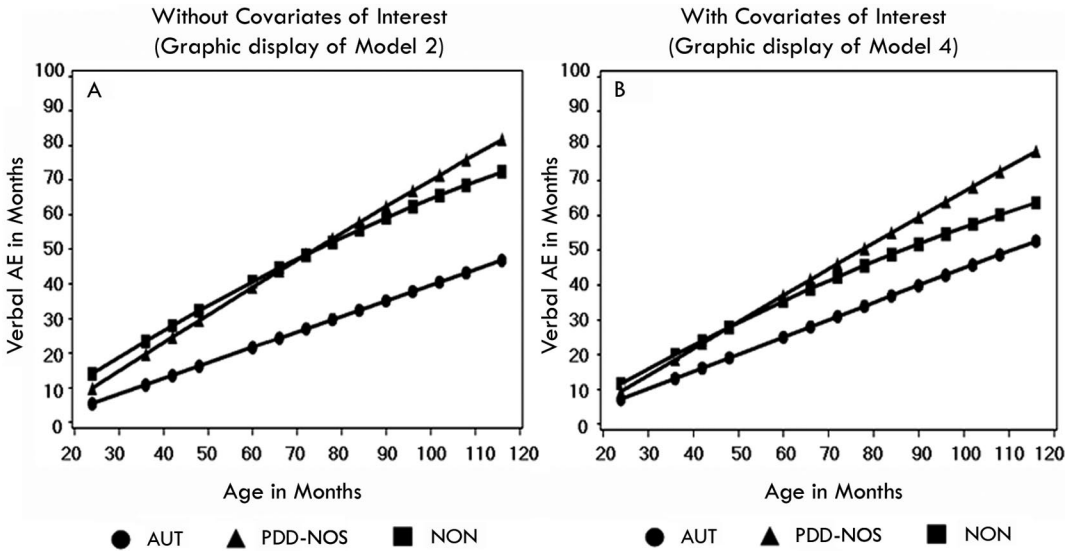

Figure 1. Predicted growth trajectories in verbal age equivalents (AE) from age 2 to age 9 for children with autism (AUT; $n = 98$), pervasive developmental disorders-not otherwise specified (PDD-NOS; $n = 58$), and nonspectrum disabilities (NON; $n = 50$). A: test for differences in verbal abilities over the 7-year period according to age 2 diagnosis. B: includes same fixed effects as those in Panel A, with the addition of nonverbal AE, joint attention, and a nonverbal AE × Time interaction.

interaction between joint attention and time was nonsignificant (not shown in Table 2), the interpretation of this finding was not entirely clear, given the moderate correlation between joint attention and diagnostic category ($r = .59$; $p < .001$). However, when diagnostic category was excluded from the model, the Joint Attention × Time interaction still did not reach significance, although there was a nonsignificant trend ($p > .10$) for those with better joint attention skills at age 2 to progress more rapidly over time. However, NVAE had a clear positive and significant effect on the slope so that children with higher IQs made greater gains in verbal ability than their peers. Cognitive abilities accounted for considerable variability in the growth rate over time, as can be seen graphically in the contrast between Figures 1A (Model 2) and 1B (Model 4). Whereas inequalities between diagnostic groups assume a fanlike spread over time in Figure 1A, this pattern is somewhat minimized in Figure 1B after controlling for the time interactions with NVAE.

Although the residual variance in Model 4 was relatively normally distributed, there were a few outliers for whom the model fit poorly. Seven children (10 observations) had particularly large residuals (with a chance occurrence of $p < .01$). These were children who made unusually good progress, as reflected by a change diagnosis over time, either from autism to PDD-NOS or from PDD-NOS to nonspectrum, and general improvements across multiple domains, including verbal, nonverbal, and general adaptive skills. When these cases were removed from the analysis, the main findings remained unchanged, although the decelerating pattern of growth in verbal skills that was observed in the nonspectrum sample also became significant for the PDD-NOS group. There was still a trend ($p < .10$) for the nonspectrum group's growth rate to slow down over time relative to that of the autism group.

In summary, symptom severity related to diagnostic grouping and lack of joint attention skills, along with age 2 NVAE, account for a substantial amount of cross-sectional and longitudinal variation in verbal age equivalents from age 2 to 9. Overall, the random effects at the bottom of Table 2 show a reduction in the unexplained variance in the intercepts (by 66%) and slopes (by 35%) between growth curve Model 1 and Model 4. However, these results reveal little about the variability within the respective diagnostic groups, particularly for the children with ASD.

Within-Group Differences

Variability in outcomes. We predicted considerable variability in outcomes, with marked improvements in verbal abilities between the ages of 2 and 9 years for a substantial proportion of ASD children, particularly for the PDD-NOS group, though we were uncertain what proportion would show substantial improvement and the extent of the improvement. Figures 2A–C display the trajectories for individuals within each diagnostic group, where the dotted line represents typical development (i.e., age equivalent matches chronological age). Consistent with expectations, a noticeably larger proportion of children with autism clustered at or near the bottom, showing little improvement with time compared with children in the other groups. However, contrary to expectations, the *range* of outcomes, including variability in the slopes and intercepts, was as great or greater for the autism sample despite the much lower *mean* verbal ability score at age 9 (48 months) than the PDD-NOS (82 months) or nonspectrum (67 months) groups. The distributional pattern for the autism group was somewhat bimodal, with more individuals clustered at the high and low ends. At the high end, a larger number of children with autism than with nonspectrum disabilities, for example, were

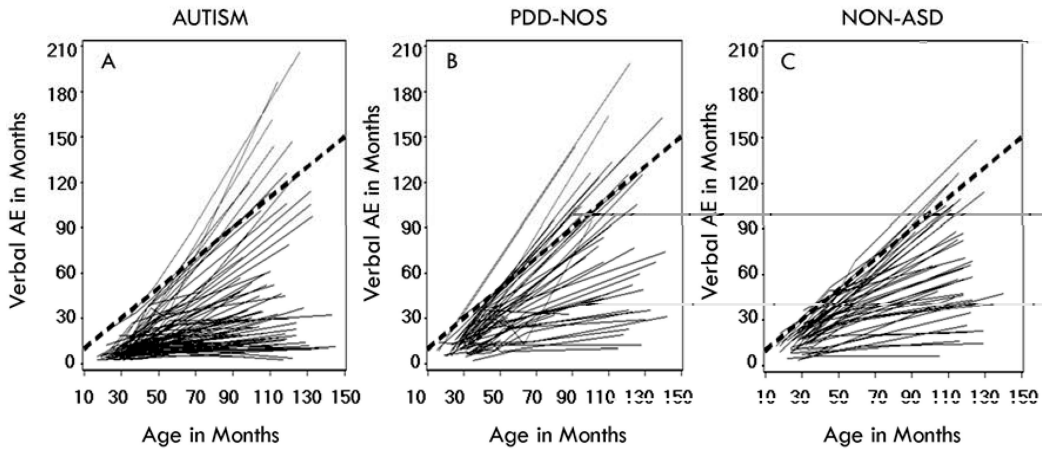

Figure 2. Growth in verbal age equivalents (AEs; solid lines) from age 2 to age 9 for individual children grouped by age 2 diagnosis. A–C: display of the trajectories for individuals within each diagnostic group. The dotted line represents typical development (i.e., AE score = chronological age). Descriptive statistics for each diagnostic group are as follows: Autism, $n = 98$, mean verbal AE at 9 years old = 48 months, intercept $SD = 3.40$, slope $SD = .43$; pervasive developmental disorders-not otherwise specified (PDD-NOS), $n = 58$, mean verbal AE at 9 years old = 82 months, intercept $SD = 3.87$, slope $SD = .42$; nonspectrum autism spectrum disorder (NON-ASD), $n = 50$, mean verbal AE at 9 years old = 67 months, intercept $SD = 3.15$, slope $SD = .27$.

near or above age norms by age 9. There was a great deal of variation within the PDD-NOS group as well, though outcomes were more evenly distributed.

Within-group variability for the children with ASD was further examined with growth mixture modeling in PROC TRAJ. The results produced four distinct subgroups on the basis of the level of improvement in verbal abilities over time. The four-group model fit the data best, with the smallest BIC (2175) as compared with those for the two-, three-, and five-group solutions (2316, 2236, and 2340, respectively). With little ambiguity, the mean probability of an individual's assignment to one group over the other three ranged from .91 to .97, depending on the group. Further evidence of good fit is shown visually in Figure 3, in which the discrepancies between the observed (solid lines) and expected values (elongated, dashed lines) are minimal. Again, growth trajectories, were mostly linear and fan shaped with a slight, positive quadratic effect for the most improved group ($p < .001$).

Notably, when the data were divided into four ASD groups, the higher and lower extremes became more visible. As expected, a substantial minority of the ASD children—about one third ($n = 53$)—showed dramatic improvement and were doing quite well at the time of the last measurement. By age 9, the average score for the two most improved groups was close to age norms or above, despite having scored below age level at the first assessment. However, the two least improved groups showed more modest average gains of about 12 and 37 months (Groups 1 and 2, respectively) over the 7-year period. Group 1 consisted mainly of nonverbal children with few or no words in the last assessment; 82% of this group were administered a Module 1 ADOS at age 9. Most of the children in Group 2 were speaking in phrases and some sentences but were not yet fluent speakers when last seen; 70% of these children were given a Module 2 ADOS at age 9. Nevertheless, 54% of the children initially diagnosed with PDD-NOS were in one of the two most improved groups. And, notably, just under 20% of the children initially diagnosed with autism were in one of the two most improved groups.

We also considered a number of potential "risk" factors that could make assignment to higher or lower functioning groups more likely. Covariates were added to the model one at a time to see whether the resulting model was a better fit, as indicated by the significance of the parameter estimates and a corresponding decrease in the BIC. When NVAE was added to the no-risk model,

Figure 3. Predicted growth trajectories in verbal age equivalents (AEs; solid lines) from age 2 to age 9 for all children ever diagnosed with autism or pervasive developmental disorders-not otherwise specified ($n = 166$) by level of improvement. PROC. TRAJ procedure generated 4 subgroupings: Group 1 (Gr. 1), $n = 70$ (56 with autism at age 2); Group 2 (Gr. 2), $n = 43$ (23 with autism at age 2); Group 3 (Gr. 3), $n = 30$ (11 with autism at age 2); Group 4 (Gr. 4), $n = 23$ (7 with autism at age 2). The dotted line represents typical development (i.e., AE score = chronological age).