

# Trajectories of Autism Severity in Early Childhood

Courtney E. Venker · Corey E. Ray-Subramanian ·  
Daniel M. Bolt · Susan Ellis Weismer

Published online: 2 August 2013  
© Springer Science+Business Media New York 2013

**Abstract** Relatively little is known about trajectories of autism severity using calibrated severity scores (CSS) from the Autism Diagnostic Observation Schedule, but characterizing these trajectories has important theoretical and clinical implications. This study examined CSS trajectories during early childhood. Participants were 129 children with autism spectrum disorder evaluated annually from ages 2½ to 5½. The four severity trajectory classes that emerged—Persistent High ( $n = 47$ ), Persistent Moderate ( $n = 54$ ), Worsening ( $n = 10$ ), and Improving ( $n = 18$ )—were strikingly similar to those identified by Gotham et al. (*Pediatrics* 130(5):e1278–e1284, 2012). Children in the Persistent High trajectory class had the most severe functional skill deficits in baseline nonverbal cognition and daily living skills and in receptive and expressive language growth.

**Keywords** Autism severity · Growth trajectories · Calibrated severity scores · Functional skill trajectories

## Introduction

Autism spectrum disorders (ASD) affect an estimated 1 in 88 children in the United States (CDC 2012). Although autism is currently the focus of a great deal of research, one area that is not well understood is how autism severity changes over the course of early development. Identifying early trajectories of autism severity has both theoretical and clinical implications. From a theoretical perspective, identifying such trajectories, as well as early predictors of these trajectories, can inform our understanding of ASD as a developmental disorder. This type of prospective developmental research is only now possible with more children being diagnosed with ASD at earlier ages. From a clinical standpoint, classifying trajectories of autism severity during early childhood would allow for improved understanding of early intervention outcomes and of the relationships between functional skills (e.g., cognition, language, and adaptive behavior)—often targeted in intervention programs—and autism severity. Identification of early autism severity trajectories would also provide clinicians with empirical information to help address questions regarding a child's short-term prognosis.

Gotham et al. (2012) conducted the first published study of trajectories of calibrated autism severity scores. It was our goal to carry out a replication and extension of this study, with the intention of determining whether similar autism severity trajectory classes emerged in an independent sample of children with ASD assessed during an earlier and narrower window of development: early

---

C. E. Venker (✉)  
Department of Communication Sciences and Disorders  
and Waisman Center, University of Wisconsin-Madison,  
1500 Highland Avenue, Waisman Center Room 475,  
Madison, WI 53705, USA  
e-mail: cgerickson@wisc.edu

C. E. Ray-Subramanian  
Waisman Center, University of Wisconsin-Madison,  
1500 Highland Avenue, Room 441, Madison, WI 53705, USA

D. M. Bolt  
Department of Educational Psychology and Waisman Center,  
University of Wisconsin-Madison, 1082a Educational Sciences,  
1025 W. Johnson St., Madison, WI 53706, USA

S. E. Weismer  
Department of Communication Sciences and Disorders  
and Waisman Center, University of Wisconsin-Madison,  
1500 Highland Avenue, Room 473, Madison, WI 53705, USA

childhood. The current study examined latent classes of autism severity trajectories across early childhood, investigated potential associations between these trajectories and demographic variables and experiential factors, and examined longitudinal trajectories of nonverbal cognition, daily living skills, and receptive and expressive language within each autism severity trajectory class.

### Measuring Autism Severity

Clinicians and researchers have used various measures to quantify the degree of core autism symptomatology in individuals with ASD. Many of these measures are based on self- or parent/caregiver-report, such as the Social Communication Questionnaire (SCQ; Rutter et al. 2003a), the Social Responsiveness Scale (SRS; Constantino and Gruber 2005), Gilliam Autism Rating Scale, Second Edition (GARS-2; Gilliam 2006) and Autism Diagnostic Interview—Revised (ADI-R; Rutter et al. 2003b). Among autism assessments based on direct observation by a clinician, scores from the Childhood Autism Rating Scale (CARS & CARS-2; Schopler et al. 1986, 2010) and the Autism Diagnostic Observation Schedule (ADOS & ADOS-2; Lord et al. 2002, 2012a, c) have been used to quantify autism severity. The variety of metrics used to measure autism severity illustrates the continued attempts of clinicians and researchers to identify valid measures to capture the construct of severity in ASD.

As discussed by Gotham et al. (2009), many measures of autism severity are highly correlated with age, cognitive abilities, and/or language skills. Unfortunately, these correlations suggest that many measures capture aspects of children's developmental levels (e.g., IQ or language skills) in addition to their core autism symptomatology, which calls into question the validity of these measures. To address this issue, as well as the fact that raw algorithm scores from the ADOS are not directly comparable across ADOS modules, Gotham and colleagues developed a standardized metric of ADOS calibrated severity scores (CSS) that are more uniformly distributed across ages and language levels. In the initial validation study (Gotham et al. 2009), the CSS more clearly differentiated diagnostic groups of children with autism, ASD, and non-spectrum developmental disorders and were less influenced by participant characteristics (e.g., age, verbal IQ, and maternal education) than ADOS raw scores.

The CSS developed by Gotham et al. (2009) have been widely influential, in large part because the ADOS is considered to be the gold standard for direct behavioral assessment of ASD symptoms and thus is commonly used in both clinical and research settings. The ADOS and the newly published ADOS-2 consist of modules (the Toddler

Module and Modules 1–4) that are selected on the basis of an individual's chronological age, developmental level, and expressive language level. Ratings are completed for a number of items within the domains of language and communication, reciprocal social interaction, play/imagination, and stereotyped behaviors and restricted interests. Item scores are then entered into module-specific diagnostic algorithms (Gotham et al. 2007) that provide cutoffs for *autism* and *autism spectrum* classifications on Modules 1 through 4 and *ranges of concern* on the Toddler Module. Although ADOS raw algorithm scores are substantially correlated with phenotypic characteristics, CSS have shown relative independence from child-level skills (de Bildt et al. 2011; Gotham et al. 2009; Shumway et al. 2012). The CSS have been incorporated into the updated ADOS-2 as Comparison Scores, which indicate the level of autism spectrum-related symptoms observed during the ADOS administration: High (scores of 8–10), Moderate (5–7), Low (3–4), or Minimal to No Evidence (1–2).

Two recent studies have directly addressed the issue of CSS validity in independent samples of children with ASD. de Bildt et al. (2011) conducted the first large-scale replication of Gotham et al.'s (2009) study by examining CSS in an independent sample of 1,248 Dutch children with a total of 1,455 ADOS administrations (Modules 1, 2, & 3). Although there were some differences between the Gotham et al. (2009) and de Bildt et al. (2011) samples (e.g., in age, verbal IQ, and autism severity levels), the study by de Bildt et al. largely replicated the findings of Gotham et al. Specifically, CSS for Modules 1 and 3 were more comparable across age and language groups; showed improved diagnostic group discrimination; and were relatively independent of verbal and nonverbal cognition and maternal education, compared to raw algorithm scores. Differences in sample characteristics between the two studies may have contributed to the inconsistent results for Module 2 (de Bildt et al. 2011).

Shumway et al. (2012) examined the stability of ADOS CSS over a period of 12–24 months in a sample of 368 children, ages 2–12. They also assessed the relationship between verbal and nonverbal developmental quotients and language abilities, and raw and calibrated ADOS scores. Within diagnostic groups (i.e., autism, PDD-NOS, non-spectrum delay, typically developing), CSS were more uniformly distributed across modules than raw algorithm scores. Verbal developmental quotients and language skills (e.g., receptive and expressive vocabulary) were found to influence raw algorithm scores more than CSS, and CSS were relatively stable in children with autism over a 12- to 24-month period. The authors identified the need for continued research on the stability of the CSS over longer periods.

## Identifying Trajectories of Autism Severity

Gotham et al. (2012) conducted the first investigation of trajectories of autism severity based on the ADOS CSS. Participants were 345 children (2–15 years of age) who contributed between two and eight data points, for a total of 1,026 ADOS assessments; these participants were part of the original CSS calibration sample (Gotham et al. 2009) and were diagnosed with ASD on at least one occasion. The vast majority of children fell into one of four latent trajectory classes: Persistent High, Persistent Moderate, Worsening, or Improving. Gotham et al. (2012) found that a variety of factors, including gender, race, history of language regression, participation in intensive therapy, and initial nonverbal IQ, were not significantly associated with autism severity trajectory class membership. Children with higher initial verbal IQ scores, however, were more likely to be in the Persistent Moderate, Worsening, and Improving classes as compared to the Persistent High class, which was designated as the reference class in these analyses.

In their examination of verbal IQ and daily living skill trajectories within the latent severity classes, Gotham et al. (2012) found that the Improving class had higher baseline verbal IQ levels than the Persistent High and Persistent Moderate classes but did not differ from the Worsening class. Additionally, children in the Improving class tended to have relatively higher verbal IQ and daily living skills than the other groups by age 6, and their rates of growth in verbal IQ were highest. Nonverbal IQ trajectories were not tested, presumably because baseline nonverbal IQ was not significantly related to trajectory class membership. These findings suggest that differences in language and daily living skills are meaningfully associated with differential trajectories of autism severity.

A recent study by Lord et al. (2012b) examined trajectories of ADOS raw algorithm scores in young children at risk for ASD, who were assessed, on average, 5–7 times between 18 and 36 months. The study by Lord and colleagues is relevant to the current study because of its focus on latent trajectory classes in toddlerhood and on relationships between trajectory classes and child-level variables. It is important to note, however, that differences in findings may be at least partially explained by their use of ADOS raw algorithm scores as opposed to CSS. Four trajectory classes emerged: Severe Persistent, Worsening, Improving, and Non-Spectrum. These classes overlap with those identified by Gotham et al. (2012), with the exception of the non-spectrum class, which was comprised of a subset of participants who never received ASD diagnoses. For the children ever diagnosed with ASD, trajectory classes did not differ on the basis of gender, maternal education, treatment, or report of skill regression. Trajectory class differences emerged for verbal IQ and verbal and nonverbal

mental age, but not nonverbal IQ. For example, children in the Severe Persistent class had slower gains in receptive and expressive language skills than children in the Improving class.

Although both Gotham et al. (2012) and Lord et al. (2012b) examined the relationships between trajectories of autism severity and a variety of child-level factors, additional studies are required to determine whether these findings hold in independent samples of children at different ages. In other words, a single study on CSS trajectories in children and adolescents with ASD and a single study on ADOS raw algorithm trajectories in toddlers cannot provide definitive evidence regarding these issues. To continue the work in this area, we also examined the relationship between demographic variables and experiential factors (e.g., history of language loss and participation in intensive behavioral intervention), and autism severity trajectories during early childhood. An estimated 20–30 % of parents of children with ASD report that their child experienced a loss of previously acquired language skills (e.g., Jones and Campbell 2010; Meilleur and Fombonne 2009), but research findings on differential outcomes for children who experience language regression have been inconsistent (e.g., Jones and Campbell 2010; Meilleur and Fombonne 2009; Rogers 2004; Shumway et al. 2011). Although Gotham et al. (2012) and Lord et al. (2012b) explored the relationship between general skill regression and autism severity trajectories, we were specifically interested in examining whether clear language loss was predictive of severity trajectory class membership.

Given that most ASD intervention research has focused on outcomes such as IQ, social communication skills, adaptive behavior, and educational placement (e.g., Dawson et al. 2010; Lovaas 1987; Yoder and Stone 2006), we were also interested in examining whether participation in intensive behavioral intervention during early childhood was associated with autism severity trajectories. Recent intervention studies have included ADOS raw scores or CSS as outcome measures for young children (e.g., Dawson et al. 2010; Green et al. 2010) but have not demonstrated clear support for intervention effects on these scores.

Following Gotham et al. (2012) and Lord et al. (2012b), we also examined how CSS trajectories related to trajectories of three associated but separable functional skills: nonverbal cognition, daily living skills, and language. First of all, multi-level growth models (described below) have the potential to identify differences in rates of growth even when baseline ability levels are similar—meaning that our analysis might reveal qualitatively different relationships between severity trajectory classes and each of these functional skill trajectories in intercept, slope, or both. In fact, Gotham et al. (2012) identified a relationship between autism severity trajectory class and baseline daily living

skills, but not nonverbal cognition, supporting separate examination of these functional skill trajectories.

Second, although we would expect cognition and daily living skills to be related in young children with ASD, they are distinct constructs that warrant separate examination. Cognitive and daily living skills have been shown to be only moderately correlated ( $r = 0.47$ ) in 2-year-olds with ASD (Ray-Subramanian et al. 2011), meaning that they capture information about different skills. The Daily Living Skills domain on the Vineland-II measures skills such as independent feeding, safety awareness, and participation in household routines. Nonverbal cognitive skills, such as visual discrimination, memory, and visual-motor ability, likely contribute to the development of daily living skills, but research has shown that there may be a gap between nonverbal IQ and daily living skills for some individuals with ASD (Kanne et al. 2011).

Further, structural language skills are an area of marked variability in children with ASD. Investigating the relationship between trajectories of autism severity and trajectories of language skills may help shed light on underlying causes of this heterogeneity. Examining trajectories of CSS may be particularly advantageous because these scores were designed to limit the impact of verbal IQ. This type of work, in turn, may lead to empirically motivated study of the mechanisms related to autism symptomatology that lead to decreased language abilities in children with ASD. We were interested in the independent trajectories of receptive and expressive language because these abilities may follow distinct patterns of development in children with ASD. For example, receptive language may be even more severely impaired than expressive language in some young children with ASD (Charman et al. 2003; Ellis Weismer et al. 2010; Volden et al. 2011; but see Kover et al. 2013, for role of nonverbal cognition), which underscores the importance of examining differences in receptive and expressive language during early childhood. Additionally, ADOS modules specifically account for differences in children's spoken language levels, but potential differences in receptive language are not explicitly addressed.

### The Current Study

With the exception of the study by Gotham et al. (2012), little is known about how standardized levels of autism severity change over the course of development—in large part because a standardized severity metric based on objective clinical observations was only recently made available. The current study is an investigation of longitudinal trajectories of autism severity in a well-characterized sample of young children with ASD from toddlerhood to early school age. Broadly, our aims were to identify

trajectories of autism severity during early childhood and to determine how demographic variables, experiential factors, and functional skill trajectories differ by autism severity trajectory class. Our specific objectives were to (1) identify latent classes of autism severity trajectories across early childhood in a heterogeneous group of over 100 children with ASD; (2) determine whether demographic variables or experiential factors were associated with trajectories of autism severity; and (3) examine between-class differences in baseline levels (intercepts) and rates of growth (slopes)—of cognition, daily living skills, and receptive and expressive language. CSS were selected as the standardized measure of autism severity in the current study because these scores are comparable across different developmental and language levels and are less influenced by age, nonverbal cognition, and language skills than ADOS raw algorithm scores—which supports their validity as a measure of autism severity (de Bildt et al. 2011; Gotham et al. 2009; Shumway et al. 2012). Additionally, CSS have demonstrated stability over a 1- to 2-year period (Shumway et al. 2012).

As indicated by Gotham et al. (2012), replication of their study is required to better understand how differential trajectories of autism severity may inform research or clinical practice. This study both replicates the investigation by Gotham et al. (2012) and extends it in several ways. First, their participants were a subset of the CSS standardization sample, leading to an acknowledged potential for circularity in findings; the current study is the first study of CSS trajectories in an independent sample. Second, our participants were diagnosed with ASD more recently and at a relatively younger age than many of the participants in Gotham et al. Given that all but one of the children in the current study was diagnosed no earlier than 2007, the present sample is likely to better represent the broader population of children currently diagnosed with ASD. For example, the current sample has somewhat higher language scores than the participant sample in Gotham et al. but nonetheless represents a heterogeneous group of children with ASD. Identification of similar autism trajectory classes across these two studies would suggest that such findings can be generalized to broader samples and are not due entirely to age or cohort effects.

Third, this study speaks to a specific time point in development: early childhood. Because early childhood is a period of rapid development for all children, it is possible that trajectories of autism severity during middle childhood or adolescence differ from those in early childhood. Additionally, functional skills, such as meaningful speech at 5 years of age, are associated with long-term outcomes in individuals with ASD (e.g., Howlin et al. 2004). Understanding trajectories of autism severity from toddlerhood to early school age may help explain why some children attain

age-appropriate cognitive and language skills by school age, but others do not. Fourth, we examined between-class differences in trajectories of receptive language and expressive language development independently. Although Lord et al. (2012b) examined separate effects of receptive and expressive language in their sample of toddlers at risk for ASD, no studies have yet examined these skills separately in relation to trajectories of CSS.

## Methods

### Participants

Participants were 129 children enrolled in a longitudinal study of early language development in children with ASD. Children between 24 and 36 months of age with suspected or diagnosed ASD were initially recruited from local early intervention programs, developmental medical clinics, and from the community. Children participated in an initial visit at age 2½ and annual follow up visits over the next 3 years. Participant demographics are presented in Table 1. The participants in this study overlap with the participant samples in (references omitted for purposes of blind review).

Most children ( $n = 101$ ) contributed data at three or four time points. A subset of children ( $n = 65$ ) was not evaluated at the third time point because of a change in study protocol. In addition, a number of families ( $n = 26$ ) withdrew from the study at some point over the 4 years. In the full sample, 12 participants contributed data at a single time point. All participants were included—regardless of the number of data points they contributed—because they all helped to characterize variability in CSS. Children whose families did and did not complete the full study did not differ on initial age, maternal education, cognitive and language scores, or CSS,  $p$ 's = 0.15–0.91. Children who were not seen at Time 3 had significantly lower maternal education ( $p = .009$ ), CSS ( $p = .049$ ), and PLS-4 Auditory Comprehension standard scores ( $p = .01$ ) than children who were seen at Time 3. The reason for these differences is unclear since the decision of which children to evaluate at Time 3 was based simply on the timing of their initial visits. It should be noted that the magnitudes of these differences were small to moderate (Cohen's  $d = 0.47$  for maternal education; Cohen's  $d = 0.35$  for CSS; Cohen's  $d = 0.45$  for PLS-4 Auditory Comprehension).

### Procedure

Comprehensive evaluations were conducted at age 2½, 3½, 4½, and 5½ (Time 1–4). Parents or legal guardians provided signed informed consent for their child to participate.

**Table 1** Sample description ( $n = 129$ )

|  | <i>n</i>      | %      |
|--|---------------|--------|
| Gender   |               |        |
| Female   | 17            | 13     |
| Male   | 112           | 87     |
| Race/ethnicity                                 |               |        |
| Caucasian                                      | 111           | 86     |
| African American                               | 2             | 2      |
| Hispanic                                       | 4             | 3      |
| Multiracial or other                           | 12            | 9      |
| Maternal education ( $n = 128$ )               |               |        |
| 11–12 years                                    | 43            | 34     |
| 13–15 years                                    | 39            | 30     |
| 16 or more years                               | 46            | 36     |
| Language loss ( $n = 111$ )                    |               |        |
| Yes  | 31            | 28     |
| No   | 80            | 72     |
| Intensive intervention ( $n = 107$ )           |               |        |
| Yes  | 71            | 66     |
| No   | 36            | 34     |
| ADOS module                                    |               |        |
| Time 1 ( $n = 127$ )                           |               |        |
| Module 1(or toddler)                           | 115           | 91     |
| Module 2                                       | 12            | 9      |
| Time 4 ( $n = 103$ )                           |               |        |
| Module 1                                       | 32            | 31     |
| Module 2                                       | 51            | 50     |
| Module 3                                       | 20            | 19     |
|  | Mean (SD)     | Range  |
| Chronological age                              |               |        |
| Time 1   | 30.82 (4.07)  | 23–39  |
| Time 4   | 66.59 (5.00)  | 57–79  |
| ADOS CSS                                       |               |        |
| Time 1 ( $n = 127$ )                           | 7.60 (1.91)   | 1–10   |
| Time 4 ( $n = 103$ )                           | 7.15 (1.81)   | 3–10   |
| Mullen developmental quotient                  |               |        |
| Time 1 ( $n = 111$ )                           | 76.39 (14.46) | 38–115 |
| Time 4 ( $n = 103$ )                           | 76.29 (18.89) | 33–108 |
| Vineland-II daily living skills standard score |               |        |
| Time 1 ( $n = 125$ )                           | 80.09 (9.83)  | 50–104 |
| Time 4 ( $n = 102$ )                           | 79.55 (10.78) | 48–111 |
| PLS-4 auditory comprehension standard score    |               |        |
| Time 1 ( $n = 125$ )                           | 60.14 (12.34) | 50–117 |
| Time 4 ( $n = 100$ )                           | 81.69 (26.46) | 50–129 |
| PLS-4 expressive communication standard score  |               |        |
| Time 1 ( $n = 124$ )                           | 72.92 (11.66) | 50–110 |
| Time 4   | 78.76 (25.86) | 50–133 |

ADOS CSS calibrated autism severity scores on the Autism Diagnostic Observation Schedule, PLS-4 Preschool Language Scale, 4th Edition

All study procedures were approved by the university Institutional Review Board.

At Time 1, best estimate clinical DSM-IV diagnoses were made using all available information and assessment results, including the ADOS (Lord et al. 2002) and a toddler research version of the ADI-R (Rutter et al. 2003b). In the full sample, 91 % ( $n = 117$ ) of participants received an initial diagnosis of Autistic Disorder/autism and 9 % ( $n = 12$ ) received a PDD-NOS diagnosis. The ADOS was administered at each subsequent time point, and best estimate clinical diagnoses were made again based on all available information. Among the 103 participants who remained in the study through Time 4, four received a different diagnosis than their initial Time 1 best estimate diagnosis. Specifically, three children with an initial PDD-NOS diagnosis were given an Autistic Disorder/autism diagnosis at Time 4, and one child with a Time 1 diagnosis of Autistic Disorder/autism received a PDD-NOS diagnosis at Time 4.

All measures outlined below were administered annually. Demographic and treatment information was collected via parent questionnaires. Maternal education (range = 11–20 years of formal education) was classified as 11–12 years, 13–15 years, or 16+ years; one family did not report this information.

## Measures

### *Autism Diagnostic Observation Schedule*

The ADOS (Lord et al. 2002) is a semi-structured, standardized assessment of social interaction, communication, and behaviors relevant to ASD. Modules are selected based on an individual's expressive language and developmental level. A preliminary research version of the Toddler module (Luyster et al. 2009) was used for participants under 30 months of age at Time 1.

A raw score was calculated for each ADOS administration, based on the revised algorithms (Gotham et al. 2007). Each ADOS raw algorithm score was then converted to a CSS between 1 and 10 based on the child's age and the ADOS Module he or she received (i.e., the respective calibration cell for each data point; see Gotham et al. 2009). For participants who received the Toddler module at Time 1, we followed the same procedure as Gotham et al. (2009) by recording the corresponding items to Module 1 algorithms. Scores of 1–3 indicate a non-spectrum classification; scores of 4–5 indicate an autism spectrum classification; and scores of 6–10 indicate an autism classification. CSS ranged from 1 to 10 (see Table 1).

### *Mullen Scales of Early Learning*

The Mullen Scales of Early Learning (Mullen 1995) is comprehensive developmental measure designed for

children between birth and 68 months of age. The Mullen is comprised of five scales (Receptive Language, Expressive Language, Gross Motor, Fine Motor, & Visual Reception); only the Visual Reception and Fine Motor scales were administered. The Visual Reception scale measures visual discrimination and visual memory and includes items that require children to remember pictures and match objects and letters. The Fine Motor scale measures visual-motor ability, including object manipulation and writing readiness. This scale includes items that require children to imitate block structures, copy shapes, and cut with scissors. It was not possible to obtain T-scores for all participants at each time point, either because children's raw scores were too low or because their ages were outside the range for which the Mullen manual provides normative data. For this reason, a developmental quotient was derived by averaging age equivalent scores from the Visual Reception and Fine Motor scales, dividing by the child's chronological age, and multiplying by 100 (see Bishop et al. 2011).

### *Vineland Adaptive Behavior Scales, Second Edition*

The Survey Interview Form of the Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow et al. 2005), is a semi-structured caregiver interview that assesses an individual's adaptive behaviors in four broad domains: Communication, Daily Living Skills, Socialization, and Motor Skills. The Vineland-II was designed for use with individuals from birth through age 90. Domain-level standard scores and subdomain-level age equivalent scores are available. An overall Adaptive Behavior Composite score can also be obtained. Because we were interested specifically in daily living skills, the standard score from this domain was used in the analyses.

### *Preschool Language Scale, Fourth Edition*

The Preschool Language Scale, Fourth Edition (PLS-4; Zimmerman et al. 2002) is an omnibus measure of receptive and expressive language skills for children between birth and 6 years, 11 months. The PLS-4 Auditory Comprehension subscale and Expressive Communication subscale measure receptive and expressive language, respectively. The PLS-4 assesses a variety of language skills, including vocabulary and grammar. The Auditory Comprehension and Expressive Communication subscales provide raw scores, age equivalent scores, and standard scores; a total language score that combines the receptive and expressive scores is also available. The standard scores from the Auditory Comprehension and Expressive Communication subscales were used in the analyses unless otherwise noted.

## Other Variables of Interest

### Language Loss

This variable was created based on parent responses on the ADI-R (Rutter et al. 2003) and represents whether the child had a parent-reported language loss of three or more words for at least 1 month at some point during development. Only “definite” losses (i.e., coded a “2” on the ADI-R) were included. Among the participants for whom language loss data was available ( $n = 111$ ), 28 % were reported to have had a definite language loss.

### Intensive Behavioral Intervention

Parents completed questionnaires about children’s intervention services at each visit and at 6-month intervals between visits. Because the available information about intervention services was somewhat limited and highly variable, a broad, dichotomous variable was derived that differentiated children who had ever received intensive autism intervention (i.e., 20 or more hours per week) over the course of the larger longitudinal study from those who had not. Among the participants for whom complete intervention data were available ( $n = 107$ ), 66 % received 20 or more hours per week of intensive, in-home autism-specific therapy at some point over the course of the longitudinal study.

### Analysis Plan

To identify trajectory classes of autism severity, a series of latent class growth models (LCGMs; Muthén and Muthén 2000) allowing for 2, 3, 4, and 5 latent classes was estimated using the Mplus software, Version 6.12 (Muthén and Muthén 1998–2011). The analysis assumed a fixed occasion design, with time (Time 1–4) as the independent variable and CSS as the dependent variable. In an LCGM, the intercept and linear growth parameters are allowed to vary between, but not within, the latent classes. We also explored models with a quadratic term added, but such models failed to converge in most instances, likely due to the limited number of measures per child (maximum of 4). Models were estimated using restricted maximum likelihood estimation with robust standard errors, and the residual variance of CSS was constrained to equality across the four time points both within and across classes. Models allowing for different numbers of classes were compared using Akaike’s Information Criterion (AIC) and the sample-size adjusted Bayesian Information Criterion (SSBIC)<sup>1</sup>;

lower AIC and SSBIC values are indicative of a relatively better fit.

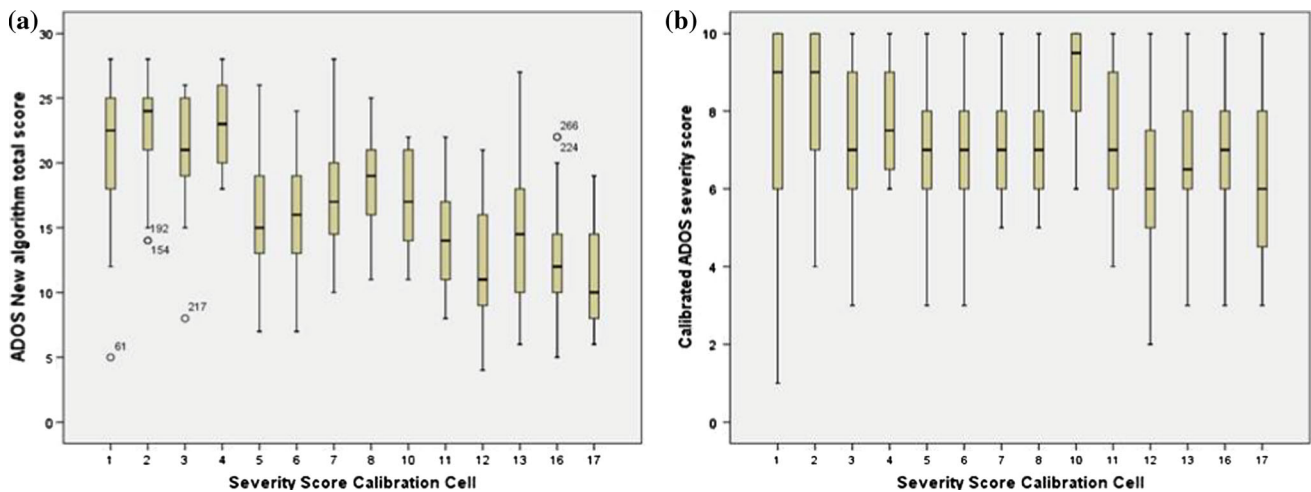
Following model selection, children were assigned to an autism severity trajectory class based on their posterior probabilities of class membership. Posterior probability values range from 0 to 1 and represent the likelihood that each child belongs to a particular class; values close to 0 indicate a low likelihood that a child would be assigned to a particular class, and values close to 1 indicate a high likelihood of being assigned to that class. For example, a child might have a posterior probability of 0.002 for belonging to one class and 0.998 for belonging to another class. Children were placed in the class with the highest posterior probability.

Our second objective was to determine whether demographic variables and experiential factors were predictive of class membership. IBM SPSS Statistics for Windows, Version 21 (IBM Corp 2012), was used to perform multinomial logistic regression analyses in which class membership was the categorical outcome variable and each factor of interest was a predictor. Each predictor was tested in a separate model because we were interested in the zero order associations between each factor and class membership. Strength of prediction was evaluated using McFadden’s  $R^2$ , a Pseudo  $R^2$  value, with higher values indicating better prediction. McFadden’s  $R^2$  represents the relative goodness-of-fit of a model, or its substantive significance; unlike linear regression, it should not be interpreted as the proportion of variance in the outcome variable explained by the predictor(s).

Our third objective was to determine whether trajectories of nonverbal cognition, daily living skills, receptive language, and/or expressive language differed within each trajectory class. To address this aim, we estimated a series of multi-level linear growth models predicting each of the four outcomes of interest using Hierarchical Linear and Nonlinear Modeling (HLM) software, Version 7 (Raudenbush et al. 2010). A multi-level approach allowed us to investigate class differences in the random intercepts and random slopes of these outcomes, while appropriately handling the longitudinal nature of the data (i.e., repeated measures across children). In each of the four models, time (Time 1–4) was a Level-1 predictor and latent class membership was a Level-2 predictor of both intercept and slope. Time was centered at Time 1, when children were approximately 2½ years of age. In each model, we first tested the main effect of between-class differences in intercept and slope. If an omnibus  $\chi^2$  test indicated significant class differences, planned pairwise contrasts were

<sup>1</sup> AIC is calculated as:  $-2 \log \text{likelihood} + 2p$ , where  $p$  is the number of parameters in the model. SSBIC is calculated as:  $-2 \log \text{likelihood} + p \ln([N + 2]/24)$ , where  $p$  is the number of parameters in the model

Footnote 1 continued  
and  $N$  is the sample size. SSBIC takes sample size into account and is more appropriate than unadjusted BIC for limited sample sizes.



**Fig. 1** The distribution of calibrated severity scores (a) and raw algorithm scores (b) on the Autism Diagnostic Observation Schedule, separated by calibration cells based on age and language level

conducted. Type 1 error was controlled using the Bonferroni-Holm method.<sup>2</sup> Effect sizes of significant between-class differences in intercept and slope were quantified using a measure analogous to Cohen’s *d*.<sup>3</sup>

**Results**

**Initial CSS Validation**

Because the CSS metric has undergone independent validation (de Bildt et al. 2011; Shumway et al. 2012), we examined the issue of validity prior to conducting our primary analyses. Combining all data points, we first compared distributions of CSS and ADOS raw scores across calibration cells based on age and language level (see Fig. 1). Consistent with prior work (Gotham et al. 2009; de Bildt et al. 2011; Shumway et al. 2012), the CSS represented a more uniform distribution than the raw scores across the calibration cells.

Second, regression analyses revealed that CSS were consistently more weakly associated with a number of

demographic variables and experiential factors than raw algorithm totals, confirming their relative independence from phenotypic and demographic characteristics (Gotham et al. 2009; Shumway et al. 2012; de Bildt et al. 2011). Regression analyses predicting raw algorithm scores and CSS were conducted with nonverbal cognition (Mullen developmental quotient) and language (PLS-4 total language standard score) in the first block, and demographics (gender, race/ethnicity, maternal education, and age) in the second block. At Time 2, the full model explained 39 % of the variance in raw algorithm scores, but only 14 % of the variance in CSS. This pattern was consistent across all four time points; it confirmed the intended properties of the CSS and supported their use in subsequent analyses.

**Trajectory Classes of Autism Severity**

LCGMs were estimated containing 2, 3, 4, and 5 classes. In each model, the dependent variable was CSS, and the independent variable was time (Time 1–4). The four-class model had the lowest AIC and SSBIC, indicating that it provided the best fit to the data (see Table 2). The four latent trajectory classes that emerged are presented in Fig. 2. Interestingly, the four classes closely resembled the four primary classes identified by Gotham et al. (2012). To maintain consistency, each class was named on the basis of its qualitative and quantitative features—Persistent High, Persistent Moderate, Worsening, and Improving—using the same terminology adopted by Gotham et al. Children were assigned to the class with the highest posterior probability. The average posterior probabilities and the number of children assigned to each latent class are presented in Table 3, along with intercept and slope values for each class. Most children were assigned to either the Persistent High class (36 %) or the Persistent Moderate class

<sup>2</sup> Alpha level for the pairwise contrasts was controlled using the Bonferroni-Holm method as follows. First,  $\alpha$  was set at a standard level of  $\alpha = 0.05$ . The contrast with the lowest *p* value was tested against  $\alpha/6$ , or  $\alpha = 0.008$ , since there was a total of six planned contrasts among classes. If the first contrast was significant, the contrast with the next lowest *p* value was tested at  $\alpha/5$ , or  $\alpha = 0.01$ , since there were five remaining contrasts. This process continued until a contrast was non-significant.

<sup>3</sup> The effect size variant of Cohen’s *d* used for the pairwise comparisons was calculated as the difference in the intercepts (or slopes) between classes, divided by the residual standard deviation of the random intercept (or slope), as estimated using the HLM software. Following Cohen’s classification, effect sizes of  $d > 0.8$  are considered indicative of large effects.



(42 %), with fewer children assigned to the Worsening class (8 %) and Improving class (14 %).

Table 4 presents descriptive statistics of CSS by class and by time point. At all four time points, the mean CSS was 7 for the Persistent Moderate class and 9 for the Persistent High class, indicating the general stability of the CSS means in these two classes. The range of mean CSS in the Worsening class was 4–6 across the four time points; the range of mean CSS in the Improving class was 5–6. Although their names suggest definite patterns of change in CSS over time, it is important to note that the Worsening

and Improving classes were both characterized by mean CSS in the mild to moderate range.

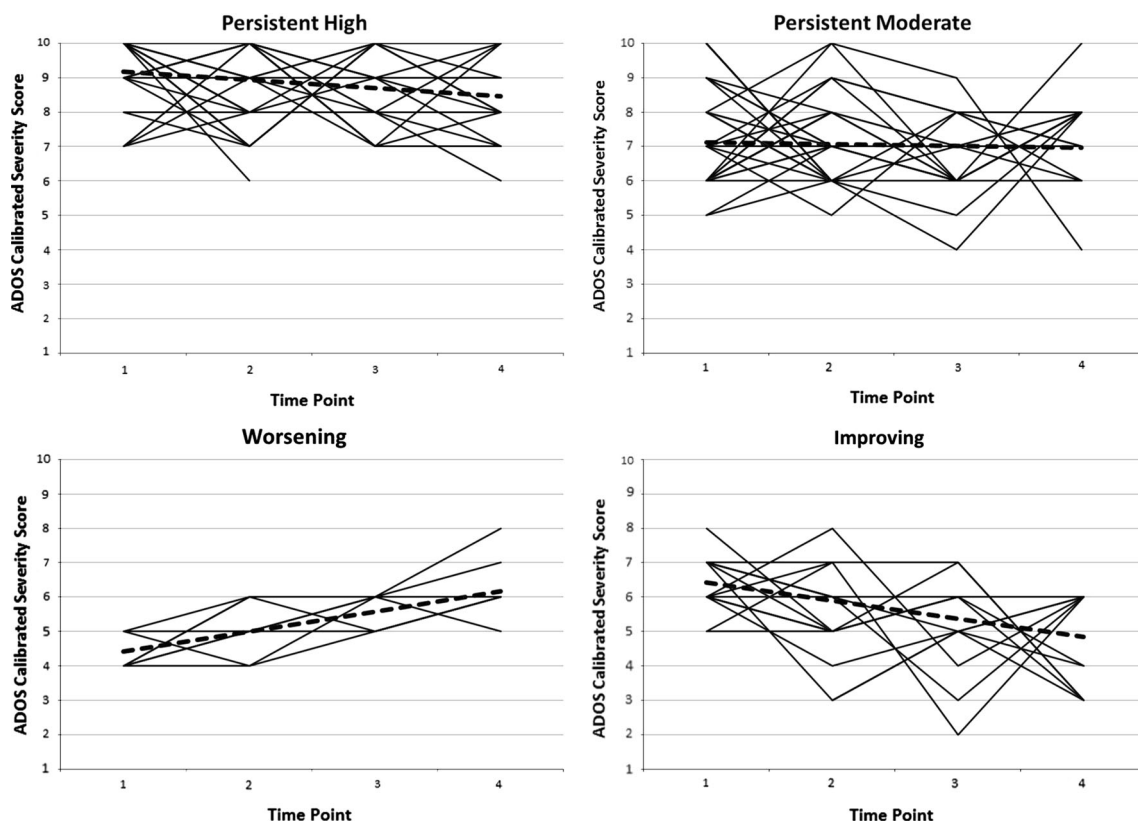
To further characterize the trajectory classes, we examined the number of children in each group whose final CSS decreased, increased, or stayed the same, compared to their initial CSS. (Note that children with data at only one time point ( $n = 12$ ) could not be categorized in this way). Based on the mean trajectories, we expected that most children in the Worsening class would have final CSS that exceeded their initial CSS, and vice versa for children in the Improving class. We also anticipated that there would be roughly similar numbers of children who worsened or improved slightly in the Persistent High and Persistent Moderate classes, since mean trajectories for these classes were generally stable. The majority of children in the Worsening class (80 %) had higher CSS at their final visit than their initial visit; no children showed improving CSS in this group. As would be expected, the majority of children in the Improving class (61 %) had lower CSS at their final visit than their initial visit; approximately one-third of children in this class had the same CSS at both visits, and only one child had an increased CSS at the final visit. The proportions of children with increased, decreased, or identical CSS were generally similar in the Persistent

**Table 2** Latent trajectory class model comparison

|               | AIC             | SSBIC           |
|---------------|-----------------|-----------------|
| 2-Class model | 1,572.24        | 1,570.43        |
| 3-Class model | 1,564.60        | 1,561.88        |
| 4-Class model | <b>1,564.33</b> | <b>1,560.70</b> |
| 5-Class model | 1,566.79        | 1,562.25        |

AIC Akaike's Information Criterion, SSBIC sample-size adjusted Bayesian Information Criterion

The four-class model (indicated in bold) had the lowest AIC and SSBIC values, indicating the best fit to the data



**Fig. 2** Individual trajectories of calibrated severity scores for children assigned to the Persistent High trajectory class (**a**;  $n = 47$ ), the Persistent Moderate class (**b**;  $n = 54$ ), the Worsening class (**c**;

$n = 10$ ), and the Improving class (**d**;  $n = 18$ ). The dashed line indicates the mean trajectory within each class

**Table 3** Autism severity trajectory classes

|                     | <i>n</i> (%) | Mean posterior probability | Intercept | Slope |
|---------------------|--------------|----------------------------|-----------|-------|
| Persistent high     | 47 (36.4)    | 0.90                       | 9.18      | −0.24 |
| Persistent moderate | 54 (41.8)    | 0.78                       | 7.12      | −0.05 |
| Worsening           | 10 (7.8)     | 0.77                       | 4.42      | 0.58  |
| Improving           | 18 (14.0)    | 0.73                       | 6.43      | −0.53 |

Children were assigned to the class with the highest posterior probability. Mean Posterior Probability values are presented for classifying children into each of the four classes. Intercept is the mean calibrated severity score at Time 1 (age 2½). Slope is the mean change in calibrated severity score per year

**Table 4** CSS characteristics by trajectory class and by time point

|                     | Time 1<br>Mean (SD)<br>Range | Time 2<br>Mean (SD)<br>Range | Time 3<br>Mean (SD)<br>Range | Time 4<br>Mean (SD)<br>Range |
|---------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| Persistent high     | 9.33 (0.92)<br>7–10          | 8.93 (1.19)<br>6–10          | 8.57 (1.17)<br>7–10          | 8.63 (1.17)<br>6–10          |
| Persistent moderate | 7.19 (1.30)<br>5–10          | 6.98 (1.10)<br>5–10          | 6.59 (1.14)<br>4–9           | 7.07 (1.24)<br>4–10          |
| Worsening           | 3.90 (1.20)<br>1–5           | 5.13 (0.84)<br>4–6           | 5.57 (0.54)<br>5–6           | 6.25 (0.89)<br>5–8           |
| Improving           | 6.41 (0.80)<br>5–8           | 5.44 (1.34)<br>3–8           | 5.15 (1.46)<br>2–7           | 4.71 (1.45)<br>3–6           |

CSS represent calibrated autism severity scores on the Autism Diagnostic Observation Schedule

Moderate and Persistent High classes; roughly 40 % of children in these two classes had decreased CSS, and 23–33 % of children had increased CSS, with the remainder receiving the same CSS at both time points.

**Impact of Demographic Variables and Experiential Factors on Trajectory Class**

Next, a series of multinomial logistic regression analyses was conducted to determine which demographic variables and experiential factors were related to latent trajectory class. In each model, class membership was the categorical dependent variable, and a demographic variable or experiential factor was the predictor variable. The Persistent High class was designated as the reference category.

Consistent with our initial hypotheses, autism severity class membership was not significantly associated with gender,  $\chi^2(3) = 3.94, p = .27$ , McFadden’s  $R^2 = 0.01$ , race/ethnicity (Caucasian vs. other),  $\chi^2(3) = 4.58, p = .21$ , McFadden’s  $R^2 = 0.02$ , maternal education,  $\chi^2(6) = 6.29, p = .39$ , McFadden’s  $R^2 = 0.02$ , or Time 1 chronological age,  $\chi^2(3) = 4.98, p = .17$  McFadden’s  $R^2 = 0.02$ . These results indicate that children were not more or less likely to be placed within a particular trajectory class of autism

severity on the basis of these factors. Additionally, language loss was not a significant predictor of class membership,  $\chi^2(3) = 0.77, p = .86$ , McFadden’s  $R^2 < 0.01$ , meaning that children’s autism trajectory class assignment appears largely unrelated to whether their parents reported a loss of language ability early in life.

Next we tested the association between intensive intervention (i.e., having ever received intensive autism services vs. having never received them) and trajectory class membership. Results indicated a significant main effect of intensive intervention services on class membership,  $\chi^2(3) = 24.43, p < .01$  McFadden’s  $R^2 = 0.09$ . Specifically, children who had ever received intensive autism intervention services were more likely to be assigned to the Persistent High class than each of the other classes: Persistent Moderate,  $b = 1.35$ , Wald  $\chi^2(1) = 4.72, p = .03$ ; Improving,  $b = 3.07$ , Wald  $\chi^2(1) = 16.87, p < .01$ ; and Worsening,  $b = 2.62$ , Wald  $\chi^2(1) = 8.44, p < .01$ .

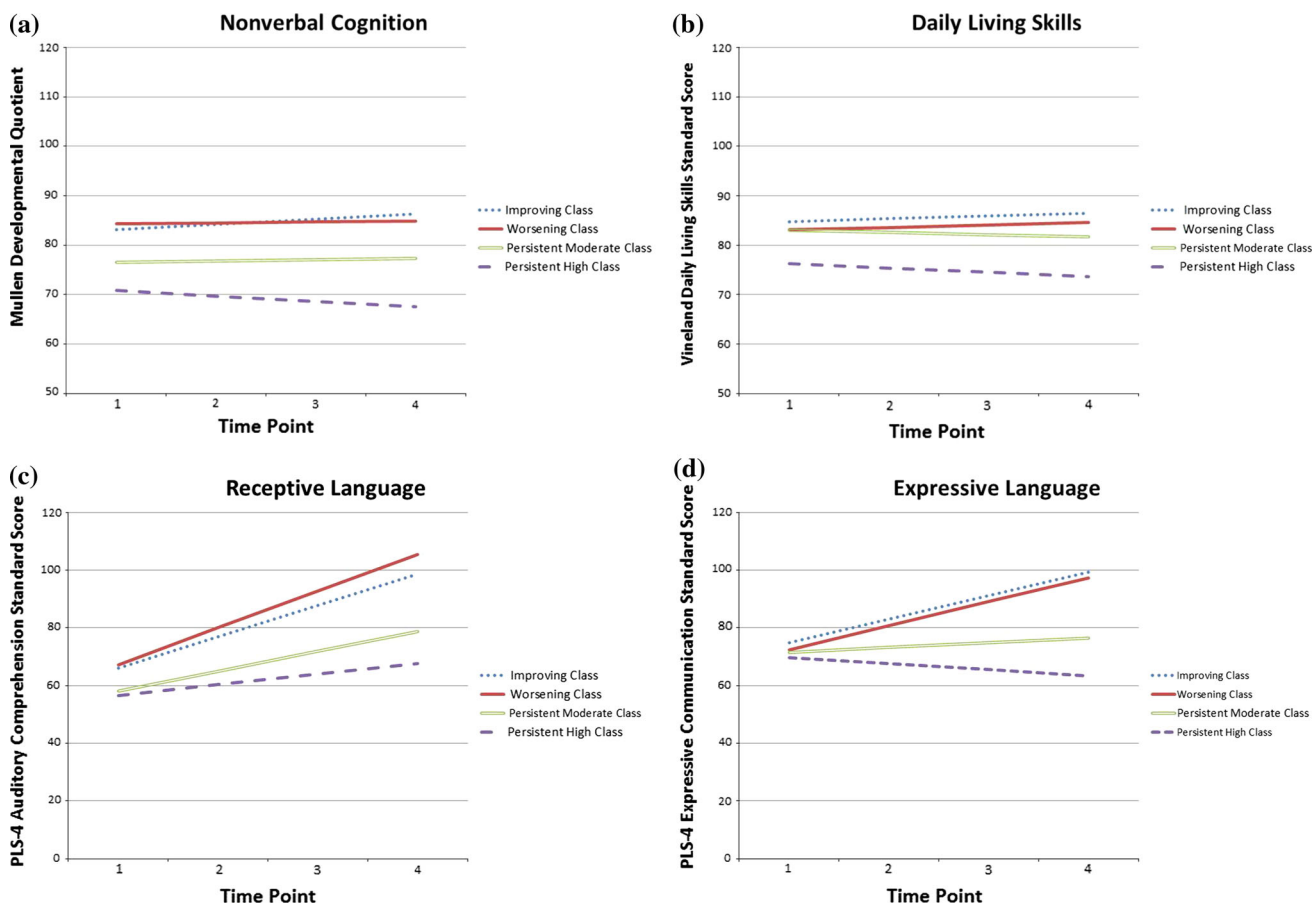
Thus, the results of the second objective have shown that children who ever received intensive autism intervention were more likely to show a Persistent High trajectory of autism severity than any other pattern. Class membership was not significantly related to gender, race/ethnicity, maternal education, age, or language loss.

**Skill Trajectories Across Autism Trajectory Classes**

Finally, a series of multi-level models was estimated to determine whether trajectories of nonverbal cognition, daily living skills, receptive language, and/or expressive language differed across the four trajectory classes (see Fig. 3). Tests were conducted to identify main effects of class membership on intercept and slope for each variable of interest; if a main effect was significant, it was followed with pairwise contrasts, using the Bonferroni-Holm method to control Type 1 error rate. The intercepts and slopes of the functional skill trajectories for each class are presented in Table 5.

The main effect of class membership on baseline nonverbal cognition (intercept) was significant,  $\chi^2(3) = 11.17, p = .01$ . Pairwise comparisons revealed that Time 1 nonverbal cognition was significantly lower in the Persistent High class than in the Improving class,  $\chi^2(1) = 7.81, p < .01, d = 0.92$ . The difference between initial nonverbal cognition in the Persistent High and Worsening classes was marginal,  $\chi^2(1) = 6.46, p = .01, d = 1.00$ . There were no significant differences in slope of nonverbal cognition across the four classes,  $\chi^2(3) = 3.42, p = .33$ .

The main effect of class membership on Time 1 daily living skills (intercept) was significant,  $\chi^2(3) = 15.71, p < .01$ . Pairwise comparisons revealed that Time 1 daily living skills were significantly lower in the Persistent High class as compared to each of the other classes: Improving,



**Fig. 3** Mean functional skill trajectories for nonverbal cognition (a), daily living skills (b), and receptive language (c) and expressive language (d) on the Preschool Language Scale, 4th Edition, within each autism severity trajectory class

**Table 5** Intercepts and slopes of functional skill trajectories by autism severity trajectory class

|  | Persistent high<br>(n = 47) | Persistent moderate<br>(n = 54) | Worsening<br>(n = 10) | Improving<br>(n = 18) |
|--|-----------------------------|---------------------------------|-----------------------|-----------------------|
| Mullen developmental quotient                  |                             |                                 |                       |                       |
| Intercept                                      | 70.79                       | 76.46                           | 84.31                 | 83.18                 |
| Slope  | -1.11                       | 0.29                            | 0.18                  | 1.02                  |
| Vineland-II daily living skills standard score |                             |                                 |                       |                       |
| Intercept                                      | 76.28                       | 82.97                           | 83.08                 | 84.76                 |
| Slope  | -0.89                       | -0.44                           | 0.49                  | 0.56                  |
| PLS-4 auditory comprehension standard score    |                             |                                 |                       |                       |
| Intercept                                      | 56.64                       | 58.15                           | 67.24                 | 65.97                 |
| Slope  | 3.66                        | 6.88                            | 12.78                 | 10.93                 |
| PLS-4 expressive communication standard score  |                             |                                 |                       |                       |
| Intercept                                      | 69.58                       | 71.61                           | 72.52                 | 74.84                 |
| Slope  | -2.04                       | 1.64                            | 8.29                  | 8.20                  |

PLS-4 Preschool Language Scale, 4th Edition

$\chi^2(1) = 8.43, p < .01, d = 1.15$ ; Worsening,  $\chi^2(1) = 11.93, p < .01, d = 0.92$ ; and Persistent Moderate,  $\chi^2(1) = 12.06, p < .01, d = 0.91$ . There were no significant differences in slopes across the four classes,  $\chi^2(3) = 2.92, p > .50$ .

The main effect of class membership on Time 1 receptive language (intercept) was marginal,  $\chi^2(3) = 7.33, p = .06$ ; no pairwise contrasts were significant. The main effect of class membership on receptive language growth (slope) was significant,  $\chi^2(3) = 21.22, p < .01$ . Pairwise

contrasts revealed that the Persistent High class had a significantly lower slope than the Improving class,  $\chi^2(1) = 16.09$ ,  $p < .01$ ,  $d = 1.55$ , and the Worsening class,  $\chi^2(1) = 9.54$ ,  $p < .01$ ,  $d = 1.95$ . This means that children in the Improving and Worsening classes showed significantly higher rates of growth in receptive language than children in the Persistent High class.

The main effect of class membership on Time 1 expressive language (intercept) was nonsignificant,  $\chi^2(3) = 1.97$ ,  $p > .50$ , but there was a significant main effect of class membership on slope,  $\chi^2(3) = 53.74$ ,  $p < .01$ . Pairwise contrasts revealed that the Persistent High Class had a significantly lower slope than each of the other classes: Improving,  $\chi^2(1) = 49.94$ ,  $p < .01$ ,  $d = 2.08$ ; Worsening,  $\chi^2(1) = 9.35$ ,  $p < .01$ ,  $d = 2.10$ ; and Persistent Moderate,  $\chi^2(1) = 6.82$ ,  $p < .01$ ,  $d = 0.75$ . In addition, the Persistent Moderate class had a significantly lower slope than the Improving class,  $\chi^2(1) = 18.41$ ,  $p < .01$ ,  $d = 1.33$ . These results indicate that children in the Persistent High class showed significantly slower rates of growth in expressive language than children in all other classes; children in the Persistent Moderate class also showed significantly slower expressive language growth than children in the Improving class.

In summary, the results of the third objective indicated that children in the Persistent High Class tended to have lower functional skills than children in the other classes, either in baseline level (intercept) or in rate of growth over time (slope). There were significant class differences in baseline levels of nonverbal cognition and daily living skills, but not in rates of growth over time. We found precisely the opposite case for language skills—namely, that there were trajectory class differences in receptive and expressive language growth, but not in baseline language levels. With one exception (Persistent High vs. Persistent Moderate expressive language slope), all pairwise comparisons had values for  $d$  above 0.9, indicating large effects.

## Discussion

### Trajectory Classes of Autism Severity

This study identified four distinct trajectory classes of autism severity—Persistent High, Persistent Moderate, Worsening, and Improving—in a heterogeneous sample of young children with ASD seen at four time points across early childhood. Despite differences in study design and participant characteristics—namely the younger age and higher language level of the current sample—these trajectory classes are very similar to those identified by Gotham et al. (2012) in a group of children with ASD from

ages 2–15. Although the current sample was recruited more recently than the sample in the study by Gotham et al., the descriptive characteristics of the CSS (i.e., mean, median, standard deviation, and ranges) in each sample were quite similar (K. Gotham, personal communication, December 18, 2012), which may help to explain the similarities in the latent trajectory classes that emerged. Three of the classes (Persistent High, Worsening, and Improving) are also similar to trajectory classes identified by Lord et al. (2012b) in toddlers with ASD, using ADOS raw algorithm scores. The fact that these studies have identified similar trajectory classes of autism severity in different age groups of children with ASD provides strong continuity within the literature and demonstrates the robustness of these developmental trajectories, regardless of whether children were assessed during toddlerhood, from toddlerhood to early school age, or through adolescence. Importantly, the participant sample in Gotham et al. was a subset of the original CSS calibration sample; this study replicates and extends their findings in an independent sample of children with ASD.

In both the current study and the study by Gotham et al. (2012), approximately 80 % of children were assigned to either the Persistent High or Persistent Moderate trajectory class, with fewer children assigned to the Worsening or Improving classes (8 and 14 % of the current sample, respectively). In conjunction, these findings suggest that a vast majority of children with ASD present with levels of autism severity that are consistently moderate or severe, with little change in overall severity level during early development. Although individual children's CSS varied somewhat across repeat ADOS administrations, the mean CSS within the Persistent High and Persistent Moderate Classes changed very little over the 4-year period (see Fig. 2a, b). This points to relatively consistent autism symptom presentation within these classes and supports the stability of the CSS scoring metric, despite considerable changes in children's ages and language levels. Note, however, that consistent presentation of autism symptoms does not mean that children are also showing consistent delays in other developmental domains; in fact, as discussed below, many children gained considerable functional skills over this 3-year period.

Gotham et al. (2012) hypothesized that a persistent, stable, and mild trajectory class of autism severity may emerge in studies of children who were diagnosed with ASD more recently, and at a younger age—like those in the current study. We found, however, no evidence of a persistent mild class of autism severity. Instead, children with more mild CSS fell primarily within the Worsening or Improving classes. At Time 4, no children in the Worsening class had an improved CSS, and only one child in the Improving class worsened, suggesting that these classes

were well characterized. Although they represented opposing directions of change in autism severity levels, the most relevant and unifying characteristic of the Worsening and Improving classes may be that they were comprised of children with more mild CSS. Indeed, the LCGM approach to identifying latent classes takes into account not only individual children's rates of growth, but also their baseline levels of CSS. There were more similarities than differences between the Worsening and Improving classes in functional skill trajectories (see discussion below), suggesting that a slight increase or decrease in mild autism severity level has a relatively limited impact on children's development of cognition, language skills, and adaptive behaviors. Because a relatively small number of children comprised the Worsening ( $n = 10$ ) and Improving ( $n = 18$ ) classes, these findings must be interpreted cautiously. Future work is needed to further quantify the implications of increasing or decreasing trajectories of mild to moderate autism severity.

The current results pointed to persistent and stable trajectories at moderate and high levels of autism severity and less stable trajectories at mild to moderate levels, which suggests that the stability of the CSS metric may depend to some extent on the severity of children's autism symptom presentation itself. In other words, CSS showed greater longitudinal stability in children whose severity levels were moderate or high than in children whose autism severity levels were relatively mild. One potential explanation may be that the children whose ASD is less severe show more variable symptom presentation on repeat administrations of the ADOS. As discussed by Hus et al. (in press), for example, restricted behaviors and repetitive interests are marked by the *presence* of atypical behaviors and can be relatively rare and thus difficult to observe in a context as limited as a single ADOS administration (but see Kim and Lord 2010). If some children with mild ASD symptoms show evidence of considerable restricted and repetitive behaviors during one ADOS administration, but not another, the resulting CSS may be less stable than for children who consistently show more marked levels of these behaviors. Presentation of atypical social communication behaviors may also vary in children with milder autism symptoms. This increased complexity of quantifying more mild ASD symptoms may also lead to decreased inter-rater reliability on the ADOS.

Further work is needed to clarify the source of variability in mild trajectories of autism severity. Hus et al. (in press) proposed a calibrated metric that provides separate CSS for the ADOS domains of Social Affect and Restricted and Repetitive Behaviors. Although domain-specific trajectories have not yet been examined using this standardized scoring system, Hus and colleagues provided some evidence that the Social Affect and Restricted and Repetitive Behaviors domains may show very different

trajectories within individual children—patterns that can be obscured by relying on CSS alone. Examining domain trajectories was outside the scope of the current study—one important avenue for future inquiry is to model longitudinal trajectories of autism severity within each domain.

#### Impact of Demographic Variables and Experiential Factors on Trajectory Class

Consistent with prior work (Gotham et al. 2012; Lord et al. 2012b), children's assigned trajectory class of autism severity was not statistically related to gender, race/ethnicity, maternal education, or age. It is in some ways encouraging that these factors—which are, in essence, static and unchangeable—do not appear to play a role in children's presentation of autism symptomatology over early childhood. Because our sample contained relatively limited racial and ethnic diversity, studies with more diverse samples should investigate this issue. Additionally, we found no evidence that early language loss was predictive of a particular trajectory of autism severity, which was also consistent with the findings of Gotham et al. (2012) and Lord et al. (2012b).

Children who had ever received intensive autism intervention services were more likely to be placed in the Persistent High class of autism severity than any other class. Although our study is observational and thus cannot speak directly to the causal direction of this relationship, it is our strong suspicion that this finding is an artifact of eligibility criteria for receiving funding for intensive, in-home autism-specific intervention through a state Medicaid waiver program. At the time of this study, the publically funded program was the only means for most families in the state to obtain intensive, in-home intervention for children with ASD, and eligibility criteria were based on level of functional skill impairment, including cognitive, communication, social, and daily living skill deficits. Contrary to this finding, Gotham et al. (2012) identified no trajectory class differences between children who had received high levels of intervention (specifically, over 20 h of a parent-mediated intervention or over 1,667 h of applied behavior analysis intervention), compared to those children who had less or no intervention. Lord et al. (2012b), however, found that more children in the Severe-Persistent class received applied behavior analysis intervention than children with ASD in the Improving and Worsening classes, though this difference was not significant.

#### Skill Trajectories Across Autism Trajectory Classes

##### *Nonverbal Cognition*

Autism severity trajectory classes differed on baseline levels of nonverbal cognition, but not in rates of cognitive

growth over time. Nonverbal cognition at Time 1 was significantly lower in the Persistent High class than in the Improving class, and marginally lower in the Persistent High class than in the Worsening class; the Persistent High and Persistent Moderate classes did not differ. In other words, children with more marked deficits in nonverbal cognition during toddlerhood were more likely to show persistent, severe autism symptomatology than more mild autism symptoms that improved or worsened over time. The similarity in growth rates of nonverbal cognition across the four classes demonstrates that the children with ASD in this study generally maintained the extent of delay in nonverbal cognition that they demonstrated early in life, regardless of the trajectory and severity of their autism symptoms. Gotham et al. (2012) found no relationship between baseline nonverbal IQ and severity trajectory class, which may be partially explained by the fact that on average, children in the current study had slightly higher levels of nonverbal cognition than participants in Gotham et al. (2012) and Lord et al. (2012b) found that initial nonverbal IQ did not predict class membership in their sample of toddlers at risk for ASD, but children with ASD in the Improving class showed higher rates of growth in nonverbal mental age than those with Severe Persistent trajectories of autism severity. This finding is interesting because it points to the possibility that examining children's absolute nonverbal ability may reveal developmental differences related to autism severity, even when standard scores of nonverbal ability do not.

#### *Daily Living Skills*

Similar to findings for nonverbal cognition, significant class differences were identified in baseline levels of daily living skills, but not in rates of growth over time. Children in the Persistent High class had significantly lower levels of daily living skills at Time 1 than children in all other classes, meaning that many toddlers who have considerable deficits in skills such as personal care (e.g., toilet training, teeth brushing, and dressing); domestic skills (e.g., helping with chores, cleaning, and cooking); and community living (e.g., talking on the telephone, using the radio or TV, and showing awareness of safety guidelines) also show persistently high levels of autism severity. As Fig. 3b illustrates, mean daily living skills standard scores were generally stable over time across all classes, meaning that on average, children did not gain or lose ground from toddlerhood to school age. Gotham et al. (2012) found no class differences in daily living skills at age 2, but children in the Improving class had significantly better daily living skills than children in the other classes at age 6.

#### *Receptive and Expressive Language*

Patterns of language development contrasted with patterns of nonverbal cognition and daily living skill development, such that the trajectory classes differed in rates of receptive and expressive language growth over time but not in baseline language levels. As Fig. 3c, d illustrates, children demonstrated considerable receptive and expressive language delays at Time 1, regardless of the autism severity trajectory class to which they were assigned. An initial deficit in language skills, then, should not be taken as a definite indication that a child will show a particular trajectory of autism severity. We find it particularly encouraging that there was no systematic relationship between autism severity trajectory class and Time 1 expressive or receptive language—despite the fact that the ADOS explicitly accounts only for differences in spoken language (i.e., through selection of the appropriate module).

As Fig. 3c indicates, rates of receptive language growth differed drastically across the severity trajectory classes. The Worsening and Improving classes had significantly higher rates of receptive language growth than the Persistent High class, meaning that children with persistent, severe levels of autism symptomatology are also at risk for persistent, severe deficits in language comprehension. Despite the slowed rate of growth in the Persistent High class, all classes demonstrated higher mean receptive language standard scores at Time 4 than at Time 1. This indicates that children not only gained absolute receptive language skills over time, but also gained ground in comparison to their typically developing peers.

In terms of expressive language, all classes had significantly higher rates of expressive language growth than the Persistent High class. Although mean expressive language standard scores increased in most classes from Time 1 to Time 4, mean expressive language standard scores in the Persistent High class decreased, meaning that on average, children in this class became more delayed relative to age expectations over time. In other words, the negative slope for expressive language standard scores indicated not that children in the Persistent High class lost language skills they had previously acquired, but that they fell further behind their typically developing peers over time. On average, expressive language scores for children in the Persistent High class decreased by 2 standard score points per year. One potential interpretation of this finding is that the subset of children with ASD who do not go on to develop functional spoken language are most likely to be those who demonstrate persistently severe symptoms of autism throughout development. Expressive language is a particularly important intervention target for these children, perhaps along with some form of alternative or augmentative communication to help them express their wants and needs through an alternate modality.

Regardless of trajectory class membership, the children with ASD in this study demonstrated severe receptive and expressive language delays at age 2½. Significant class differences in rates of language growth suggest, however, that some children possess learning abilities that allow them to acquire language skills more quickly than others. In fact, by age 5½, children in the Improving and Worsening groups performed within age expectations for receptive and expressive language, whereas children in the Persistent Moderate and Persistent High classes demonstrated continued delays.

What developmental processes underlie the trajectory class differences in rates of language growth? It is possible that children with lower levels of autism severity can better generalize their language abilities to the higher-order tasks that comprise many of the later items on the PLS-4 (e.g., making grammatical judgments, using language to describe quantitative and qualitative concepts, constructing narratives). Other learning abilities that may contribute to superior language skills include statistical learning (i.e., detection of patterns in language; Romberg and Saffran 2010), increased accuracy and efficiency of spoken language processing (Venker et al. 2013), and better integration of and access to semantic and syntactic representations. Language learning in children with ASD may also be supported by the ability to extend novel words to appropriate categories (McGregor and Bean 2012) or make effective use of adult feedback during word learning (Bedford et al. 2012).

Prior work has shown that decreases in restricted and repetitive behaviors are associated with increases in receptive and expressive language abilities in young children with ASD (Ray-Subramanian and Ellis Weismer 2012), indicating yet another reason that autism severity and language may be linked. It is also possible that higher levels of social interest and engagement lead to increased language-learning opportunities and that better general attentional abilities (i.e., sustained, selective, or flexible attention) lead to better language outcomes. Future studies are needed to more precisely identify the mechanisms that underlie optimal language outcomes in this population.

Gotham et al. (2012) found that children in the Improving and Worsening classes tended to have higher verbal IQ at age 2, with the Improving class showing the highest rate of growth. At age 6, verbal IQ was significantly higher in the Improving class and significantly lower in the Persistent High class than in all other classes. Although it is worthwhile to interpret our findings regarding class differences in language trajectories in reference to the findings of Gotham et al., it should be noted that our findings may contrast due to a number of factors. As we have acknowledged, participants in the current study and that by Gotham et al. differed in age and language

levels. Additionally, Gotham et al. did not separately examine receptive and expressive language skills. The fact that we identified qualitatively different patterns of development in receptive and expressive language—particularly the decline in expressive language standard scores for the Persistent High class—underscores the importance of separately examining these aspects of language.

Finally, Gotham et al. (2012) used verbal IQ—most commonly measured by the Mullen Scales of Early Learning, as reported in Gotham et al. (2009)—as a measure of language ability. Although there are similarities between verbal IQ and language skills as measured by the PLS-4 in the current study, these two constructs are not identical (also see Shumway et al. 2012). The Auditory Comprehension and Expressive Communication subscales of the PLS-4 were designed to assess a broader range of language skills than the Mullen, ranging from basic vocabulary and vocal development to making inferences and demonstrating phonological awareness (Zimmerman et al. 2002). The Mullen manual reports correlations ranging from 0.72 to 0.85 with the subtests on an earlier version of the PLS (Mullen 1995), which provides evidence of overlapping but non-identical measures.

## Conclusion and Limitations

In summary, this study identified four discrete trajectory classes of autism severity in early childhood, based on ADOS CSS: Persistent High, Persistent Moderate, Worsening, and Improving. These classes are strikingly similar to the four primary classes identified by Gotham et al. (2012). Important differences in functional skill trajectories by class emerged, including different rates of growth in receptive and expressive language skills. Our findings also indicate that early deficits in nonverbal cognition and daily living skills may be predictive of a persistent and severe trajectory of autism severity. The robustness of these autism severity trajectories across independent samples contributes to our understanding of ASD as a developmental disorder and may offer clinicians empirical information to inform a child's short-term prognosis.

One strength of this study is that it examined an independent sample of young children with ASD diagnosed no earlier than 2006. One related limitation, however, is that this participant sample was relatively small ( $n = 129$ ) compared to the sample in Gotham et al. (2012;  $n = 345$ ). Although a sample size of 129 is adequate for many statistical analyses (e.g., linear regression), when using LCGM one runs the risk of identifying latent classes that contain small subsets of the original sample. For example, the Worsening class in the current study contained only 8 children at the final time point. Despite their relatively

small size, the fact that the Worsening and Improving classes emerged statistically in this study and that by Gotham et al. suggests that they should be acknowledged, though replication is critical. Although the current study included considerably fewer participants than Gotham et al. (2012), the posterior probabilities of assigned class membership were similar, ranging from 73 to 90 ( $M = 79.5$ ) in the current study, and from 68 to 82 ( $M = 77.5$ ) in the study by Gotham et al.—which suggests that trajectory class assignment was comparably robust across both studies.

Selection of any analytical approach inherently involves both strengths and limitations. One advantage of the LCGM approach (Muthén and Muthén 2000) used in this study is that it does not assume a Gaussian (normal) distribution of growth trajectory parameters and thus can theoretically accommodate any distribution. LCGM also allows for variability between but not within trajectory classes, leading to more straightforward interpretation of classes than approaches that introduce variability at both levels. As mentioned, however, one potential disadvantage of this approach is that it is sensitive to latent classes that include a relatively small number of participants, and solutions can therefore be unstable when several such classes are present in the data. Population-based studies are required to determine the prevalence rates of autism severity trajectories in the broader population of children with ASD.

The current study included a maximum of four time points per child, which led to convergence problems when attempting to fit LCGMs with effects above linear effects (e.g., quadratic effects). Although the majority of individuals in Gotham et al. (2012) contributed data at two or three time points, one-fourth of the sample contributed between four and eight assessments, which likely helped in their being able to consider trajectories with a quadratic component. Relatedly, the current study used a fixed occasion design with time as a predictor, whereas Gotham et al. used a variable occasion design with age as a predictor. Despite these differences, it is important to note that inclusion of the quadratic term in the Gotham et al. study did not produce a better fitting model, meaning that the final latent class model in both studies included only intercept and linear effects. In addition, the inclusion of a quadratic component in our analyses was viewed as less critical given our focus on a narrower window of time. Future studies including more frequent assessments during early childhood (e.g., every 3–4 months) may determine whether trajectories of early autism severity measured by CSS are best modeled with both linear and quadratic effects.

A limitation of all observational studies is that definitively determining causation is not generally possible.

Although we identified significant relationships between trajectories of autism severity and trajectories of functional skills, it is not possible to say with certainty whether increased autism severity leads to decreased functional skill levels (in our opinion, the more likely interpretation), or whether lower functional skill levels lead to more severe autism symptomatology. In actuality, the relationship between autism severity and foundational developmental skills likely involves complex, bidirectional influences that shift over the course of development. Finally, this study explored only one measure of autism severity: the ADOS CSS. Although the justification for selecting the CSS is clear, future work is may determine whether trajectories of autism severity using other measures align with the current findings.

**Acknowledgments** We would first and foremost like to thank the children and families who participated in this study. Without them, this study would not have been possible. We would also like to thank the members of Language Processes Lab for their assistance in data collection and data management. This work was supported by NIH R01DC007223-05 (Ellis Weismer, PI; Gernsbacher, co-PI); T32DC005359-10 (Ellis Weismer, PI); P30HD003352-46 (Seltzer, PI).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Bedford, R., Gliga, T., Frame, K., Hudry, K., Chandler, S., Johnson, M. H., et al. (2012). Failure to learn from feedback underlies word learning. *Journal of Child Language*, *40*(1), 29–46.
- Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*, *116*(5), 331–343.
- Centers for Disease Control and Prevention (CDC). (2012). *Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States, 2008*. Morbidity Weekly Report (MMWR); Vol. 61(3).
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (infant form). *Journal of Child Language*, *30*, 213–236.
- Constantino, J. N., & Gruber, C. P. (2005). *Social responsiveness scale (SRS)*. Los Angeles, CA, USA: Western Psychological Services.
- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model. *Pediatrics*, *125*, e17–e23.
- de Bildt, A., Oosterling, I. J., van Lang, N. D., Sytema, S., Minderaa, R. B., van Engeland, H., et al. (2011). Standardized ADOS scores: Measuring severity of autism spectrum disorders in a Dutch sample. *Journal of Autism and Developmental Disorders*, *41*(3), 311–319.
- Ellis Weismer, S., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers



- with developmental delay. *Journal of Autism and Developmental Disorders*, 40(10), 1259–1273.
- Gilliam, J. E. (2006). *Gilliam autism rating scale* (2nd ed.). Austin: ProEd.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705.
- Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, 130(5), e1278–e1284.
- Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 37(4), 613–627.
- Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., & Howlin, P. (2010). The PACT consortium. *Lancet*, 375, 2152–2160.
- Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, 45(2), 212–229.
- Hus, V., Gotham, K., & Lord, C. (in press). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-012-1719-1.
- IBM Corp. Released. (2012). *IBM SPSS Statistics for Windows, Version 21.0*. Armonk, NY, USA: IBM Corp.
- Jones, L. A., & Campbell, J. M. (2010). Clinical characteristics associated with language regression for children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 40(1), 54–62.
- Kanne, S. M., Gerber, A. J., Quirnbach, L. M., Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2011). The role of adaptive behavior in autism spectrum disorders: Implications for functional outcome. *Journal of Autism and Developmental Disorders*, 41, 1007–1018.
- Kim, S. H., & Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research*, 3(4), 162–173.
- Kover, S. T., McDuffie, A. S., Hagerman, R. J., & Abbeduto, L. (2013). Receptive vocabulary in boys with autism spectrum disorder: Cross-sectional developmental trajectories. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1823-x.
- Lord, C., Luyster, R. J., Gotham, K., & Guthrie, W. (2012a). *Autism diagnostic observation schedule, second edition (ADOS-2) manual (Part II): Toddler module*. Torrance, CA, USA: Western Psychological Services.
- Lord, C., Luyster, R., Guthrie, W., & Pickles, A. (2012b). Patterns of developmental trajectories in toddlers with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, 80(3), 477–489.
- Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism diagnostic observation schedule manual*. Los Angeles: Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012c). *Autism diagnostic observation schedule, second edition (ADOS-2) manual (Part I): Modules 1-4*. Torrance, CA, USA: Western Psychological Services.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3–9.
- Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., et al. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(9), 1305–1320.
- McGregor, K. K., & Bean, A. (2012). How children with autism extend new words. *Journal of Speech, Language, and Hearing Research*, 55, 70–83.
- Meilleur, A. A., & Fombonne, E. (2009). Regression of language and non-language skills in pervasive developmental disorders. *Journal of Intellectual Disability Research*, 53(2), 115–124.
- Mullen, E. M. (1995). *Mullen scales of early learning* (AGS ed.). Circle Pines, MN, USA: American Guidance Service Inc.
- Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA, USA: Muthén & Muthén.
- Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism, Clinical and Experimental Research*, 24(6), 882–891.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2010). *HLM* (Version 7.0). [Computer software]. Chicago, IL: Scientific Software International, Inc.
- Ray-Subramanian, C. E., & Ellis Weismer, S. (2012). Receptive and expressive language as predictors of restricted and repetitive behaviors in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42(10), 2113–2120.
- Ray-Subramanian, C. E., Huai, N., & Weismer, S. E. (2011). Brief report: Adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*, 41(5), 679–684.
- Rogers, S. J. (2004). Developmental regression in autism spectrum disorders. *Mental Retardation and Developmental Disabilities Research Reviews*, 10(2), 139–143.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914.
- Rutter, M., Bailey, A., & Lord, C. (2003a). *The social communication questionnaire: Manual*. Los Angeles, CA, USA: Western Psychological Services.
- Rutter, M., LeCouteur, A., & Lord, C. (2003b). *Autism diagnostic interview-revised*. Los Angeles, CA, USA: Western Psychological Services.
- Schopler, E., Reichler, R., & Renner, B. (1986). *The childhood autism rating scale (CARS) for diagnostic screening and classification of autism. Diagnosis and teaching curricula for autism and developmental disabilities series*. New York, NY, USA: Irvington.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood autism rating scale* (2nd ed.). Los Angeles, CA, USA: Western Psychological Services.
- Shumway, S., Farmer, C., Thurm, A., Joseph, L., Black, D., & Golden, C. (2012). The ADOS calibrated severity score: Relationship to phenotypic variables and stability over time. *Autism Research*, 5(4), 267–276.
- Shumway, S., Thurm, A., Swedo, S. E., Deprey, L., Barnett, L. A., Amaral, D. G., et al. (2011). Brief report: Symptom onset patterns and functional outcomes in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 41(12), 1727–1732.
- Sparrow, S., Cicchetti, D. V., & Balla, D. (2005). *Vineland-II adaptive behavior scales*. Circle Pines, MN, USA: American Guidance Service.
- Venker, C. E., Eernisse, E. R., Saffran, J. R., & Ellis Weismer, S. (2013). Individual differences in the real-time comprehension of children with ASD. *Autism Research*. doi:10.1002/aur.1304.
- Volden, J., Smith, I. M., Szatmari, P., Bryson, S., Fombonne, E., Miranda, P., et al. (2011). Using the Preschool Language Scale

- to characterize language in preschoolers with autism spectrum disorders. *American Journal of Speech-Language Pathology*, 20(3), 200–208.
- Yoder, P., & Stone, W. L. (2006). Randomized comparison of two communication interventions for preschoolers with autism spectrum disorders. *Journal of Consulting and Clinical Psychology*, 74, 426–435.
- Zimmerman, I. L., Steiner, B. S., & Pond, R. E. (2002). *Preschool language scale* (4th ed.). San Antonio, TX, USA: The Psychological Corporation.