J. Child Lang. **38** (2011), 485–503.
doi:10.1017/S0305000910000140

Prelinguistic predictors of language development in children with autism spectrum disorders over four–five years

KAREN D. BOPP* AND PAT MIRENDA

University of British Columbia, Canada

(*Received* 9 *March* 2009 – *Revised* 5 *October* 2009 – *Accepted* 13 *February* 2010 – *First published online* 8 *July* 2010)

ABSTRACT

This study examined relationships between prelinguistic variables from the MacArthur-Bates CDI and the development of language comprehension and production in children with autism. Forty-four children were assessed at baseline and 6, 12, 24, 33 and 53 months later. Growth Curve Modeling was used to examine the extent to which three composite CDI variables and three CDI item groupings predicted language development over 4–5 years. When examined individually, prespeech and early gestures were significant predictors of change for both comprehension and production, but late gestures were not. In addition, initiating joint attention and games and routines predicted comprehension and production over 4–5 years, and conventional gestures also predicted production. When all factors were considered simultaneously, children's ability to participate in games and routines was the only significant predictor of language production over time. The results are discussed with regard to their implications for understanding the complex factors that affect developmental outcomes.

Autism spectrum disorder (ASD) is a term that refers to a group of developmental disorders that includes autism, Pervasive Developmental Disorders-Not Otherwise Specified (PDD-NOS) and Asperger syndrome.

[*] This research was funded through a post-doctoral fellowship from the National Alliance for Autism Research/Autism Speaks and a grant from the Ministry of Children and Family Development in British Columbia, Canada. We are grateful to the children and families who participated and to the many evaluators and research assistants involved in the project. Portions of this study were presented at the 2008 International Meeting for Autism Research in London, UK; and the 2008 Conference of the American Speech-Language-Hearing Association in Chicago, IL. Address for correspondence: The University of British Columbia – Educational Counselling Psychology, and Special Education, 2125 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada. tel: 604 822 9691; fax: 604 822 3302; email: bopp@interchange.ubc.ca

The course of language development in individuals with ASD varies widely and there is also a wide heterogeneity in the outcomes of intervention (Schriebman, 2000). Identifying variables that predict variance may assist in a better understanding of the process of language development in this population. Recently, researchers have identified a number of behaviors that appear to characterize or predict language development in young children with ASD using a parent report measure, the MacArthur-Bates Communicative Development Inventory (CDI; Fenson *et al.*, 1993; Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). The CDI has proven to be both valid and cost-effective for assessing vocabulary size and general language development in samples of both typically and atypically developing young children (Dale, Price, Bishop & Plomin, 2003; Luyster, Kadlec, Carter & Tager-Flusberg, 2008).

Using the Words and Gestures (WG) version of the CDI (CDI:WG), (Charman, Drew, Baird & Baird, 2003) found that, although a sample of 134 preschoolers with ASD demonstrated severe language delays as a whole, there was considerable variability in the size of the children's vocabularies, with some approaching the typical range. In general, word comprehension was more delayed than production when compared to the typical developmental norms provided with the CDI:WG, and early gestures (as measured by first communicative gestures and games-and-routines) were more delayed than late gestures (as measured by actions with objects, pretending to be a parent, and imitating other adult actions). This study was one of the first to document the prelinguistic skills that appear to be associated with language development in young children with ASD.

In a recent study, Luyster, Lopez & Lord (2007) partially replicated the findings of Charman *et al.* (2003), using either the CDI:WG or CDI:WS (Words and Sentences) collected from the parents of 93 young children with autism or PDD-NOS. As in Charman *et al.*, both diagnostic groups showed considerable variability in word production and in phrase and word comprehension, with some children's profiles approximating those of typically developing children. Similarly, Luyster, Lopez & Lord found that word production was more delayed than both word and phrase comprehension, compared to the normative pattern. When the level of comprehension was held constant, the sample with ASD mastered early gestures more slowly than typically developing controls, although this was not the case for late gestures. However, while Charman *et al.* found that mastery of early gestures remained relatively stable once a child understood 20 words, Luyster, Lopez & Lord documented a steady increase in both early and late gestures as expressive vocabulary increased.

Smith, Mirenda & Zaidman-Zait (2007) used the CDI to examine the rate of vocabulary development over time in thirty-five children with autism. This study was unique in that the children were assessed a total of four times over two years. At Time 1 (T1), all of the children spoke fewer than 60 words and were assessed using the CDI:WG, which occurred just prior to the initiation of intervention. Subsequent assessments that occurred six (T2), twelve (T3) and twenty-four months (T4) thereafter involved the CDI:WG for children who produced $\leqslant$50 words and the CDI:WS for those with $>$50 words. There were four distinct patterns (i.e. clusters) of vocabulary development within the sample: (a) Cluster 1: slight incline, with a mean increase of 10 words over 24 months ($n=15$); (b) Cluster 2: slow incline, with a mean increase of 200 words over 24 months, most of which occurred between T3 and T4 ($n=8$); (c) Cluster 3: high, steady incline, with a mean increase of 453 words over 24 months ($n=7$); and (d) Cluster 4: steep incline, with a mean increase of 638 words over 24 months ($n=5$). At T1, children in Cluster 1 (slight incline) had significantly lower verbal imitation scores than did children in the other three clusters. They also had lower scores related to using objects to pretend than did children in both Clusters 2 and 4. At T1, children in Cluster 4 (steep incline) had higher scores for initiating joint attention (IJA) than did the children in either Clusters 1 or 2. This study was the first to examine prelinguistic predictors of vocabulary development over time and suggests that verbal imitation, 'pretending to be a parent' behaviors, and IJA gestures may be important indicators in this regard.

Recently, Charman, Taylor, Drew, Cockerill, Brown & Baird (2005) examined the extent to which CDI total scores for words understood and words said at ages 2 and 3 predicted outcomes at age 7 in a sample of twenty-nine children with ASD. While scores at age 2 were not predictive, those at age 3 were correlated at age 7 with both autism symptom severity, as measured by the Autism Diagnostic Interview-Revised (ADI-R; LeCouteur, Lord & Rutter, 2003), and adaptive behavior, as measured by the Vineland Adaptive Behavior Scale – Survey Edition (VABS; Sparrow, Balla & Cicchetti, 1984). These results were partially replicated by Luyster, Qiu, Lopez & Lord (2007), who provided a more extended examination of the use of the CDI to predict outcomes in a number of domains between ages 2–3 and age 9 in both children with ASD ($n=62$) and those with developmental delay (DD; $n=19$). Like Charman *et al.*, they found that CDI receptive and expressive language scores at age 3 predicted autism severity and adaptive functioning several years later. Similarly, they found that age 3 scores were more accurate outcome predictors than those obtained at age 2. In addition, they found that the number of late gestures at age 2 predicted age 9 verbal IQ, expressive language and adaptive behavior scores in the children with ASD, while the number of early gestures and other 'prespeech' indicators did not. This was surprising, since the behaviors contained in both the prespeech variable (e.g. responds to name) and the early gestures variable (i.e. CDI first communicative gestures and games-and-routines) are reflective of general social engagement, which is almost always impaired in young children with ASD and is thought to be related to later development across domains. This result also appeared to be inconsistent with the findings of Luyster *et al.* (2008), who reported that CDI early gestures robustly predicted concurrent receptive and expressive language in toddlers with autism, aged 1;6–2;9. Luyster, Qiu *et al.* suggested that problems with the accuracy of parent reports could account for this unexpected finding and noted the need for additional explorations of the utility of the CDI to describe and predict language development and other variables over time.

The purpose of the present study was to extend the work of both Charman *et al.* (2005) and of Luyster and her colleagues related to prelinguistic predictors of language and adaptive behavior over time. Specifically, we wanted to further examine the relationship between CDI prespeech, early gestures, and late gestures and language development over time, in a longitudinal sample of young children with ASD who were slightly older than those in the Luyster, Qiu *et al.* (2007) study but who were nonetheless significantly delayed at the time of initial data collection. We also wanted to examine the extent to which individual CDI subsection scores acted as predictors of language development. For both of these, we followed children over a 4·5 year period.

METHOD

Participants

Participants included forty-four children (84% males, mean age = 3;11; range = 1;9 to 5;11). Thirty-six were diagnosed with autism and eight were diagnosed with pervasive developmental disorder-not otherwise specified (PDD-NOS) by experienced community-based clinicians who were not involved in the study. Specifically, twenty-one children (48%) received a diagnosis from an autism team that included, at a minimum, a psychologist or psychiatrist, a pediatrician and a speech-language pathologist; twelve (27%) received a diagnosis from a developmental pediatrician with experience in autism; eight (18%) received a diagnosis from a psychiatrist; and three (7%) received a diagnosis from a registered psychologist or another qualified professional with autism experience.

In addition, the Childhood Autism Rating Scale (CARS; Schopler, Reichler & Renner, 1988) was administered at the outset of the study by psychologists who were naive to the original diagnosis. The mean CARS score at T1 was 37·8 (range = 25·0–50·5). Scores above 30 ($n = 40$) on the CARS indicate autism, and scores below 30 ($n = 4$) are indicative of PDD-NOS, based on the results of Perry, Condillac, Freeman, Dunn-Geier & Belair (2005), who examined the concordance rate between the CARS and

DSM-IV clinical diagnoses in 274 preschool children (aged 2 to 6). The CARS was used in the present investigation because when the children were diagnosed, neither of the current 'gold standard' diagnostic instruments for autism – the ADI-R and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore & Risi, 2002) – were commonly used in the children's home province, because of a lack of trained diagnosticians. The CARS is a 15-item behavior rating scale that provides a total raw score and a raw score for each of 15 rating categories. Saemundsen, Magnússon, Smári & Sigurdardóttir (2003) examined both the CARS and the ADI-R with fifty-four children (ages 1;10 to 9;6) who were referred for possible autism. They found a significant correlation ($r=0{\cdot}81$; $p<0{\cdot}001$) between the ADI-R total score and the total score on the CARS. Similarly, Pilowsky, Yirmiya, Shulman & Dover (1998) examined the agreement between these two instruments with seventy participants who were suspected of having autism. They found an 85·7% agreement between the CARS and ADI-R diagnoses. Together, these results provide support for the concurrent validity of the two measures.

The participants' ethnic backgrounds included European-Canadian ($n=25$), European ($n=2$), Asian ($n=7$), Asian-Canadian ($n=5$), Hispanic ($n=2$), Caribbean/African ($n=1$), Middle Eastern ($n=1$) and multiple ethnicities ($n=1$). At baseline, mothers, on average, had completed some college or university courses and were considered semi-skilled workers (e.g. machine operator, grocery store clerk) (Hollingshead, 1975). Fathers, on average, had completed some university training and were considered skilled workers (e.g. department manager, administrative assistant) (Hollingshead, 1975). The families included thirty-six two-parent families, one separated family, four divorced families and three other arrangements (e.g. common-law). All of the children received approximately fifteen to twenty hours per week of year-round early intervention services for two years. The majority of children (87%) also attended preschool or school during this time.

Measurement

Data were collected at baseline (T1), and (on average) 6, 12, 24, 33 and 53 months later (T2–T6). At T6, the children's mean age was 9;4 (range = 6;2 to 10;4). Data collection for all measures occurred in each child's home or early intervention center by assessment teams that included registered psychologists, certified speech-language pathologists (SLPs) and trained graduate students who acted as family interviewers. At each time-point, data collection was typically conducted by the same assessment team members and the data collection location was consistent for each child, except at T6 when all data were collected in the children's homes. None of

https://doi.org/10.1017/S0305000910000140 Published online by Cambridge University Press

the assessors were involved in service provision, either to the children or to their families. Only measures of overall development, prelanguage skills and communication/language development were included in this investigation.

General development. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is a multidomain standardized measure that can be used with children aged between 0;3 and 5;8. The MSEL yields both mental age equivalence scores and an Early Learning Composite (ELC) score. Typically developing children achieve an ELC of 100 (+ or − 15), which is derived from their performance in four domains of development: visual reception, receptive language, expressive language and fine motor skills. The MSEL that was administered at T1 was used in this study.

Prelanguage skills. The CDI Words and Gestures version (CDI:WG) is a parent report measure that assesses, among other things, prelanguage skills such as communicative gestures and symbolic behaviors. The CDI:WG was completed at T1 by the caregivers (mostly mothers) of all forty-four participants.

Two sets of predictor variables from the CDI:WG were used in this analysis. The first set included three broad prelanguage variables that were identical to those used by Luyster, Qiu *et al.* (2007). They included (1) PRESPEECH, compiled from items reported as mastered from Part I, sections A and C (i.e. a total of five items: responds when name is called; responds to 'no, no'; responds to 'there's mommy/daddy' by looking around for them; imitates words; and names or labels objects with social intent); (2) EARLY GESTURES, compiled from items reported as mastered from Part II, sections A (i.e. first communicative gestures) and B (i.e. games-and-routines), for a total of eighteen items; and (3) LATE GESTURES, compiled from items reported mastered from Part II, sections C (i.e. actions with objects), D (i.e. pretending to be a parent) and E (i.e. imitating other adult actions), for a total of forty-five items. Wherever ordinal scales were used, 'never' and 'not yet' were coded as 'no' (0) and 'sometimes' and 'often' were coded as 'yes' (1).

The second set of predictor variables included six CDI:WG groupings that were more specific than those used by Luyster, Qiu *et al.* (2007) for early and late gestures and that were also used in a previous study examining the predictors of expressive vocabulary growth over two years in young children with ASD (Smith *et al.*, 2007). Several categories of behaviors that were grouped together on the CDI:WG were constructed: (a) the number of behaviors used to INITIATE JOINT ATTENTION, from a list of 3 (Part II, A1–3); (b) the number of CONVENTIONAL GESTURES, from a list of nine (Part II, A4–12); (c) the number of GAMES-AND-ROUTINES in which the child was reported to participate, from a list of six (Part II, section B); (d) the number of ACTIONS WITH OBJECTS, from a list of seventeen (Part II, section C); (e) the number of actions used to PRETEND TO BE A PARENT from a list of thirteen (Part II, section D); and (f) the number of IMITATIONS OF OTHER ADULT ACTIONS, from a list of fifteen (Part II, section E).

Measures of language development. Composite scores for language comprehension and production were created for this analysis by combining each child's raw scores from vocabulary and language measures that were administered by a certified speech-language pathologist at each time-point. The comprehension measures included the Peabody Picture Vocabulary Test–IIIA/B (PPVT-III; Dunn & Dunn, 1997) and the auditory comprehension (AC) subscale of the Preschool Language Scale–3 (PLS-3-AC; Zimmerman, Steiner & Pond, 1992). The PPVT has been shown to be a reliable and valid measure of receptive vocabulary for persons aged between 2;6 and over 90;0, with standard and raw scores well correlated to chronological age and intellectual functioning (Dunn & Dunn, 1997). The PLS-3 is also a reliable and valid measure of auditory comprehension, language expression and overall language, with both standard and raw scores that are highly correlated with chronological age and other measures of language (Zimmerman *et al.*, 1992). Composite scores for language production were created from raw scores on the Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000) and the PLS-3 expressive communication subscale (PLS-EC). The EOWPVT is a reliable and valid measure of expressive vocabulary whose standard and raw scores are highly correlated with chronological age and measures of cognitive ability, language, academic achievement and receptive vocabulary (Brownell, 2000).

Coefficient alpha was computed to gauge the reliability of the composite scores based on the average correlation among items (Cronbach, 2004). Typically, a coefficient alpha $\geqslant 0{\cdot}70$ indicates that items used to calculate a composite score are internally consistent with little measurement error (Streiner & Norman, 1989). At all six time-points, the coefficient alpha for language comprehension was $\geqslant 0{\cdot}80$ and the alpha for language production was $\geqslant 0{\cdot}75$.

Data analysis

Individual Growth Curve Modeling (IGCM) using SAS Proc Mixed computer software was used to determine if any of the prelanguage skills measured with the CDI:WG at T1 predicted changes in the developmental trajectories of communication skills in young children with autism over four to five years. Individual growth trajectories are a rich and flexible alternative to traditional methods for analyzing longitudinal data (e.g. regression). IGCM was used in order to take advantage of the multi-wave data that were available for investigation. IGCM enables the examination of change as a continuous process, such that the amount of change between time periods for a given participant is a result of that participant's underlying growth trajectory (Francis, Fletcher, Stuebing, Davidson & Thompson, 1991). In addition, IGCM includes path analysis and regression to explore relationships among changes in individual variables over time (Schumacker & Lomax, 2004).

Data analysis was completed using a four-step process. In all steps and models, an unstructured correlation structure was used (Singer, 1998). Age was used as the measure of time in order to take into account the wide age range at T1; and age was centered (i.e. age minus the mean age at T1) to facilitate interpretation of the output (Singer & Willet, 2003).

In Step 1, unconditional models were run separately for each composite language variable (i.e. comprehension and production) over fifty-three months. In an unconditional model, no factors (e.g. gender, IQ score, etc.) are used to predict individual growth. Thus, during this step, only the shape of the growth curve and the variance in the intercept and the slope were examined. Because neither the cubic nor the quadratic model was significant, all subsequent steps (i.e. Steps 2–4) were conducted using the slope and intercept of the linear model.

Step 2 of the analysis examined whether the intercept and slope varied as a function of individual differences between participants (Chen & Cohen, 2006). In Step 2, separate conditional models were run for each composite variable to determine the influence of T1 autism severity and T1 non-verbal IQ (NVIQ) scores over fifty-three months. Autism severity was measured using the CARS total scores. NVIQ was estimated for thirty-nine children by combining the *t*-scores from the visual reception and fine motor subscales from the MSEL (see Luyster, Qiu *et al.*, 2007); general developmental growth scores were not available for the remaining five children. Autism severity and NVIQ were examined because past research has found a relationship between these two variables themselves (Eaves & Ho, 1996) and between NVIQ and other developmental outcomes (Bopp, Mirenda & Zumbo, 2009; Luyster, Qiu *et al.*, 2007).

In Step 3, relationships between the slope of both composite variables and the variables of interest were examined. Conditional models were run with each of the Luyster, Qiu *et al.* (2007) variables (i.e. prespeech, early gestures and late gestures) as predictors of both the intercept and slope. In addition, conditional models were run using each of the specific CDI:WG groupings (described previously) as predictors of the intercept and slope of the composite language measures. All models included T1 CARS and/or T1 NVIQ as predictors of the intercept and/or slope, when warranted from the results of Step 2.

In Step 4, final conditional models were run for both composite variables that included only the variables that were found in Steps 2 and 3 to be significant predictors of the intercept and/or slope. These variables included, when warranted, T1 CARS and/or NVIQ; prespeech from Luyster, Qiu *et al.* (2007); and the six specific CDI:WG groupings. The Luyster, Qiu *et al.* early and late gestures variables were not included because the six specific CDI:WG groupings are subsets of these variables.

RESULTS

Child language development (T1 to T6)

Table 1 summarizes the mean raw scores, ranges and standard deviations for the composite measures of comprehension and production at T1–T6 for all forty-four children.

Unconditional model

In Step 1, unconditional models were examined to determine whether there was significant variance in the initial value (i.e. intercept) and the rate of change (ROC; i.e. slope) for each of the composite language measures over fifty-three months. Table 2 presents these findings and the Akaike Information Criterion (AIC), the model's goodness-of-fit statistic, for each measure. There is no statistical test available to assess the differences in AIC between models; however, models with lower AIC values are preferred and are considered a better fit (Tabachnick & Fidell, 2007). Overall, the results indicated that there was significant variance in the intercept and that participants changed significantly over fifty-three months on both composite measures (i.e. the variance in the ROC over fifty-three months was significant).

Conditional model

In Step 2, multilevel conditional models were used to examine whether autism severity (i.e. CARS scores) and/or NVIQ measured at T1 predicted the individual difference variance in the intercept and/or the ROC over fifty-three months for either composite language measure. A variable that predicts the intercept indicates significant differences in INITIAL SCORES on a measure. A variable that predicts the slope indicates significant differences in the rate of change on a measure OVER TIME. Table 3 presents these findings.

T1 NVIQ alone but not T1 CARS alone was a significant predictor of the intercept for comprehension, and both NVIQ and CARS alone were predictors of the intercept for production. However, when NVIQ and CARS were entered simultaneously into the model, only NVIQ continued to be a significant intercept predictor. These results are evident by comparing the AIC values in Table 3 to those in Table 2. Based on the results of Step 3, all subsequent conditional model analyses included T1 NVIQ as a predictor of the intercept for both of the composite language measures.

TABLE 1. *Mean raw scores, ranges and standard deviations for composite measures of comprehension and production at T1–T6*

	T1		T2		T3		T4		T5		T6	
Composite measure	MRS[a] (range)	*SD*	MRS (range)	*SD*	MRS (range)	*SD*	MRS (range)	*SD*	MRS (range)	*SD*	MRS (range)	*SD*
Comprehension	12·66 (3–60)	12·60	24·77 (6–92)	22·08	33·41 (5–107)	28·76	49·50 (8–112)	33·87	56·22 (9–127)	39·41	85·56 (17–138)	33·32
Production	13·70 (4–52)	11·04	23·98 (6–71)	18·04	29·73 (5–75)	22·25	42·84 (7–111)	30·76	50·11 (9–114)	34·73	72·06 (19–123)	33·10

[a] Mean raw score.

TABLE 2. *Unconditional models of change for composite measures of comprehension and production at T1–T6 (Step 1)*

Value	Comprehension	Production
Intercept estimate	17·70	17·28
t-value	5·75*	6·42*
Slope estimate	1·16	0·97
t-value	9·46*	8·45*
AIC	1754·4	1645·7

* $p < 0·0001$.

TABLE 3. *T1 CARS and NVIQ as predictors of comprehension and production from T1 to T6*[a] *(Step 2)*

Value	Comprehension	Production
	T1 CARS alone	
Intercept estimate	−1·22	−0·80
t-value	−2·24*	−1·64
Slope estimate	−0·02	−0·04
t-value	−0·81	−1·70
AIC	1753·6	1645·6
	T1 NVIQ alone	
Intercept estimate	1·24	1·08
t-value	3·75**	3·84**
Slope estimate	0·009	0·18
t-value	0·55	1·22
AIC	1562·9	1456·9
	T1 CARS plus T1 NVIQ	
T1 CARS intercept estimate	−0·80	−0·16
t-value	−1·52	−0·37
T1 NVIQ intercept estimate	1·12	0·95
t-value	3·61**	3·59**
AIC	1560·9	1458·2

* $p < 0·05$, ** $p < 0·001$.
[a] Centered age as calculated by age minus mean age at T1 was used as the measure of time.

Table 4 summarizes the means and standard deviations of all of the predictor variables that were used in Steps 3–4.

In Step 3, relationships between the ROC of both of the composite language measures and the three Luyster, Qiu *et al.* (2007) composite variables were examined, using a Bonferroni adjustment of 0·017 (0·05 ÷ 3) for each analysis (Abdi, 2007). For both comprehension ($\beta = 7·69$, $t(44) = 3·28$, $p < 0·01$) and production ($\beta = 7·27$, $t(44) = 3·81$, $p < 0·001$), prespeech was a significant predictor of the intercept and was thus included in all subsequent analyses. For the slope, both prespeech and early gestures

https://doi.org/10.1017/S0305000910000140 Published online by Cambridge University Press

TABLE 4. *Means and standard deviations of the predictor variables used in Steps 3–4*

Predictor variable	Mean	*SD*
Prespeech (5 items)	3·4	1·09
Early gestures (18 items)	8·77	4·11
Late gestures (45 items)	20·82	9·05
Initiates joint attention (3 items)	1·73	1·11
Conventional gestures (9 items)	3·98	2·26
Games and routines (6 items)	3·07	1·47
Actions with objects (17 items)	9·57	3·74
Pretending to be a parent (13 items)	2·98	3·08
Imitating an adult (15 items)	8·27	3·47

TABLE 5. *Predictors of the individual difference variance in the rate of change of composite language measures over fifty-three months (Step 3)*

	Composite language measure			
	Comprehension		Production	
Predictor variable	β	t	β	t
Prespeech (Luyster, Qiu *et al.*, 2007)	0·39	3·35*	0·38	3·57*
Early gestures (Luyster, Qiu *et al.*, 2007)	0·96	3·33*	0·10	3·87*
Late gestures (Luyster, Qiu *et al.*, 2007)	0·02	1·77	0·025	1·85
Initiates joint attention	0·35	3·27*	0·35	3·66*
Conventional gestures	0·12	2·14	0·13	2·48*
Games and routines	0·29	3·54*	0·31	4·17*

* $p < 0·017$.

significantly predicted the individual difference variance in the ROC over fifty-three months for both comprehension and production, but late gestures did not. Thus, none of the CDI:WG groupings that made up late gestures (i.e. actions with objects, pretending to be a parent and imitating other adult actions) was included in subsequent analyses.

Next, relationships between the ROC of both composite language measures and the three CDI:WG early gestures groupings (i.e. initiating joint attention, conventional gestures and games-and-routines) were examined, again using an adjusted alpha of 0·017 for each analysis (Abdi, 2007). For comprehension, two specific CDI:WG grouping variables (initiating joint attention and games-and-routines) predicted the ROC when placed into the conditional model in isolation. For production, all three CDI:WG grouping variables (initiating joint attention, conventional gestures and games-and-routines) predicted the ROC. Table 5 summarizes the results of Step 3 for both composite language measures.

TABLE 6. *Final conditional models for composite language measures over fifty-three months (Step 4)*

Value	Production
T1 NVIQ intercept estimate	0·905
t-value	3·79**
T1 prespeech intercept estimate	7·16
t-value	3·04*
T1 games and routines slope estimate	0·314
t-value	4·17**
AIC	1455·7

* $p < 0{\cdot}01$, ** $p < 0{\cdot}001$.

Finally, in Step 4, separate conditional models were run for each composite language measure, using only the variables that were found to be significant in Steps 2 and 3. For comprehension, these included both T1 NVIQ and prespeech as predictors of the intercept, and prespeech, initiating joint attention, and games-and-routines as predictors of the slope. For production, the same variables were used for the intercept, and conventional gestures was added for the slope. Results indicated that, for comprehension, none of the individual variables was a significant predictor of the slope when combined in the model. For production, only games-and-routines remained a significant predictor of the slope at $p < 0{\cdot}01$ when combined with prespeech, initiating joint attention and conventional gestures. The final model for production is presented in Table 6.

Overall, the results revealed that the inclusion of T1 NVIQ and prespeech as predictors of the intercept and games-and-routines as a predictor of the slope or ROC resulted in a model for production only that was a better fit, as reflected in the lower AIC values in Table 6, compared to Table 3.

Individual Growth Curve Modeling (IGCM) software used in this analysis (i.e. SAS Proc Mixed) does not yield conventional effect size estimates such as those available in other statistical packages (e.g. SPSS). However, the statistical effects can be illustrated by examining the average rate of change in language production growth for a prototypical child with low and high baseline games-and-routines and prespeech scores ± 1 standard deviation above and below the mean, as shown in Figure 1 (Brooks & Meltzoff, 2008; Singer & Willet, 2003). Figure 1 suggests that a child with more games-and-routines behaviors at T1 (+1 *SD*) would have a rapidly accelerating growth curve, while a child with fewer games-and-routines behaviors at T1 (–1 *SD*) would have a much slower acceleration in growth language production over four to five years.

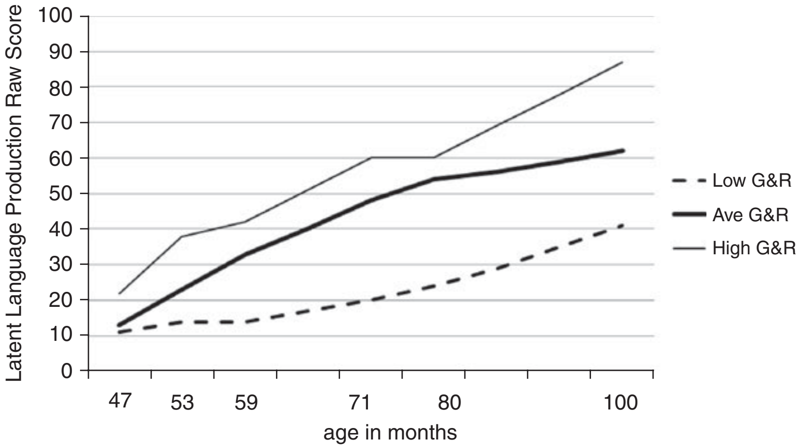

Fig. 1. Average growth curve of language production scores over four to five years for a prototypical child with T1 games-and-routines (G&R) scores that were at the mean and ±1 standard deviation above and below the mean.

DISCUSSION

The results of this study add to the growing body of research on factors that predict language development over time in young children with ASD. In addition, it is one of only a few studies to date to examine the impact of predictors from the CDI over longer than two years. The first aim of this study was to further examine the predictive value of three composite variables on the CDI:WG: prespeech, early gestures and late gestures on language development. In our Step 3 analysis, which considered each of the potential predictor variables in isolation, we found that only prespeech and early gestures predicted expressive language development over four to five years. This is not consistent with the findings of Luyster, Qiu *et al.* (2007), who found that only late gestures predicted language outcomes in children with autism over six to seven years. Our results are more similar to the findings of Luyster *et al.* (2008), who reported that early gestures on the CDI were good concurrent predictors of both expressive and receptive language.

The most likely explanation for the differences between our results and those of Luyster, Qiu *et al.* (2007) pertains to how language development was measured at around age 9;0 in the two studies. Luyster, Qiu *et al.* used one of several different measures (i.e. the Clinical Evaluation of Language Fundamentals (Semel, Wiig & Secord, 1995; Wiig, Secord & Semel, 1992); the Sequenced Inventory of Communication Development – Revised (Hedrick, Prather & Tobin, 1984); the Reynell Developmental Language

Scales (Reynell & Gruber, 1990); and the Vineland Adaptive Behavior Scales, 2nd edition (Sparrow, Cicchetti & Balla, 2005)) to measure expressive and receptive language. Because of this wide variability in language measures, they used age 9;0 age equivalents (AEs) as language scores. Luyster, Qiu *et al.* acknowledged the problems inherent in the use of AEs to measure language development, which is not nearly as linear as AE scores imply; but they used them nonetheless because no other option was available. In contrast, because we used the same four measures (i.e. the PPVT-III, PLS-AC, EOWPVT and PLS-EC) with the children in our study at all time-points, including T6 at age 9;4, we were able to use raw scores to reflect language development. The problems with the use of AE scores, especially the fact that they are not measured on an interval scale, were discussed at length by Mervis & Klein-Tasman (2004). It is likely that we were able to identify different predictors because our raw scores were more precise reflections of children's abilities in language development over time.

Our second aim was to examine a number of more specific CDI:WG item groupings that were subsets of the Luyster, Qui *et al.* (2007) early gestures variable, to determine if any of them predicted language development trajectories over four to five years. In the Step 3 analysis, several items did emerge as potential predictors of both receptive language (as measured by both the PPVT and PLS-AC) and/or expressive language (as measured by both the EOWPVT and PLS-EC). For receptive language, these included the CDI item groups for initiating joint attention and engaging in games-and-routines. For expressive language, the same predictors emerged and conventional gestures was added. However, when all of the relevant items were entered together into the conditional models (Step 4), none of these variables continued to predict comprehension, and only the number of games-and-routines in which the children participated remained a predictor of language production over time.

Our inability to identify a viable predictor for comprehension is most likely due to the fact that some of the items that emerged as potential predictors in Step 3 were more closely related to production rather than comprehension, and that they shared a significant amount of variance. Thus, there was considerable overlap among the potential predictors, to the degree that none of them was sufficiently robust to emerge as a predictor of comprehension when all were entered into the model simultaneously. In part, this was because we chose to apply a Bonferroni adjustment to the alpha level used to determine significance, in consideration of both the small sample size and the number of statistical tests required for the analysis. Additional research with larger sample sizes might enable detection of more stable predictors over time and result in different conclusions for comprehension.

The finding that social games-and-routines emerged as the sole predictor of the development of language production over four to five years is consistent with the results reported by Charman *et al.* (2005), who found that the rate per minute of non-verbal communicative acts that occurred during a play-based assessment predicted outcomes in several domains (including language development) over seven years. In that study, the acts that were counted consisted primarily of brief interactive episodes in which a child either initiated or responded to an adult's bid for joint attention. In our study, games-and-routines scores were assigned by asking parents to identify which of the following their child was able to do: play peek-a-boo, play patty cake, play 'so big', play chase, sing and dance – all of which also involve joint attention, in addition to other skills. For example, playing peek-a-boo requires a child to take turns, be aware that people and things exist even when they are not visible (i.e. object permanence), imitate actions with objects (e.g. remove a blanket put over the head to hide) and respond to social bids by a partner. Playing patty cake and 'so big' also require turn-taking and imitation skills as well as joint attention and social engagement. Chase games very much depend on establishing and maintaining joint attention; and when young children 'sing' and 'dance', they do so by imitating a range of conventional gestures in the context of a joint social routine. It is most likely the collection of these underlying skills, rather than the children's ability to play specific games and routines themselves, that are the important predictors.

Perhaps surprisingly, neither T1 NVIQ or autism severity (i.e. CARS scores) emerged as predictors of the ROC over time. In fact, it appears that the specific skills that underlie the ability to engage in games and routines are more important than either of these factors in laying the foundation for later language development. This finding is contrary to the results of other longitudinal research indicating a relationship between initial IQ score (e.g. Sallows & Graupner, 2005) and/or initial autism severity score (Eaves & Ho, 1996) and outcomes over time. It is likely that a combination of factors, including differences in subject samples, the procedures used to measure or estimate IQ and autism severity, whether or not the children were provided with treatment, and the outcomes that were measured, can account for these discrepant results.

Limitations

There are several limitations to this study that may impact the generalizability of the results. First, a relatively small sample size ($N=44$) was used. However, small sample sizes are not uncommon in ASD research (e.g. Charman *et al.*, 2005; Stone & Yoder, 2001). Furthermore, in IGCM analyses such as this one, the use of multiple measurement occasions for

https://doi.org/10.1017/S0305000910000140 Published online by Cambridge University Press

each participant increases the power considerably (Zhang & Wang, 2009). Nonetheless, follow-up studies are needed to examine the predictive relationships found here in a larger group of participants. Second, information was not available about the extent to which each child's early intervention program specifically targeted the language outcomes included in this study. In an examination of the predictors of spoken language level in children with autism, Stone & Yoder (2001) found that the number of hours of speech-language therapy was a strong predictor of later language outcomes, more so than other types of therapy or factors such as object play and joint attention skills. Additional research is needed to replicate this finding and to examine the relative importance of prelinguistic behaviors and intervention factors. Finally, all of the predictor variables originated from a parent report measure, which might have affected their accuracy.

Summary

This study adds to the growing body of research examining factors that are related to long-term language development in children with ASD. It is the only study to date that has examined specific items from the CDI as predictors of expressive and/or receptive language. Although we were unable to replicate the results of previous research using three composite variables on the CDI (Luyster, Qiu *et al.*, 2007), the results suggest that social games and routines lay the foundation for and thus predict later language production. Additional research with larger samples is needed to confirm these results, and to explore the relationships between other prelinguistic variables and related areas of development.

REFERENCES

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (ed.), *Encyclopedia of measurement and statistics*, 103–107. Thousand Oaks, CA: Sage.

Bopp, K. D., Mirenda, P. & Zumbo, B. D. (2009). Behavior predictors of language development over two years in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* **52**, 1106–120.

Brooks, R. & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language* **35**, 207–220.

Brownell, R. (2000). *Expressive One Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.

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**, 213–36.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry* **46**, 500–513.

Chen, H. & Cohen, P. (2006). Using individual growth model to analyze the change in quality of life from adolescence to adulthood. *Health and Quality of Life Outcomes* **4**(10) [Retrieved from www.hqlo.com/content/4/1/10].

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement* **64**, 391–418.

Dale, P. S., Price, T. S., Bishop, D. V. M. & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research* **46**, 544–60.

Dunn, L. M. & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition* (*PPVT-III*) *Technical Manual*. Circle Pines, MN: American Guidance Service.

Eaves, L. C. & Ho, H. H. (1996). Brief report: Stability and change in cognitive and behavioral characteristics of autism through childhood. *Journal of Autism and Developmental Disabilities* **26**, 557–69.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *MacArthur-Bates Communicative Development Inventories: Users guide and technical manual*. San Diego, CA: Singular Publishing Company.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**, 1–173.

Francis, D. F., Fletcher, J. M., Stuebing, K. K., Davidson, K. C. & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. *Journal of Consulting and Clinical Psychology* **59**, 27–37.

Hedrick, D., Prather, E. & Tobin, A. (1984). *Sequenced Inventory of Communication Development-Revised*. Los Angeles, CA: Western Psychological Services.

Hollingshead, A. (1975). Four factor index of social status. Unpublished manuscript, Yale University.

LeCouteur, A., Lord, C. & Rutter, M. (2003). *The Autism Diagnostic Interview – Revised* (*ADI-R*). Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. & Risi, S. (2002). *Autism Diagnostic Observation Schedule* (*ADOS*). Los Angeles, CA: Western Psychological Services.

Luyster, R., Kadlec, M. B., Carter, A. and Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* **38**, 1426–38.

Luyster, R., Lopez, K. & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language* **34**, 623–54.

Luyster, R., Qiu, S., Lopez, K. & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur-Bates communicative developmental inventory. *Journal of Speech, Language, and Hearing Research* **50**, 667–81.

Mervis, C. & Klein-Tasman, B. (2004). Methodological issues in group-matching designs: Levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders* **34**, 7–17.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

Perry, A., Condillac, R. A., Freeman, N. L., Dunn-Geier, J. & Belair, J. (2005). Multi-site study of the Childhood Autism Rating Scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders* **35**, 625–34.

Pilowsky, T., Yirmiya, N., Shulman, C. & Dover, R. (1998). The Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: Differences between diagnostic systems and comparison between genders. *Journal of Autism and Developmental Disorders* **28**, 143–51.

Reynell, J. K. & Gruber, C. P. (1990). *Reynell Developmental Language Scales* (*U.S. edition*). Los Angeles: Western Psychological Services.

Saemundsen, E., Magnússon, P., Smári, J. & Sigurdardóttir, S. (2003). Autism Diagnostic Interview-Revised and the Childhood Autism Rating Scale: Convergence and discrepancy in diagnosing autism. *Journal of Autism and Developmental Disorders* **33**, 319–28.

Sallows, G. & Graupner, T. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation* **110**, 417–38.

Schopler, E., Reichler, R. J. & Renner, B. (1988). *Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.

Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders* **30**, 373–78.

Schumacker, R. E. & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Semel, E., Wiig, E. H. & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals, Third Edition* (*CELF–3*). San Antonio, TX: Psychological Corporation.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* **24**, 323–55.

Singer, J. D. & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event outcome*. New York: Oxford University Press.

Smith, V., Mirenda, P. & Zaidman-Zait, A. (2007). Predictors of expressive vocabulary growth in children with autism. *Journal of Speech, Language, and Hearing Research* **50**, 149–60.

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

Sparrow, S., Cicchetti, D. & Balla, D. (2005). *Vineland Adaptive Behavior Scales*, 2nd edn. Circle Pines, MN: American Guidance Service.

Stone, W. L. & Yoder, P. (2001). Predicting spoken language level in children with autism spectrum disorders. *Autism: The International Journal of Research & Practice* **5**, 341–62.

Streiner, D. L. & Norman, G. R. (1989). *Health measurement scales: A practical guide to their development and use*. New York: Oxford University Press.

Tabachnick, B. G. and Fidell, L. S. (2007). *Using multivariate statistics*, 5th edn. Boston, MA: Allyn and Bacon.

Wiig, E. H., Secord, W. A. & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals—Preschool*. San Antonio, TX: Psychological Corporation.

Zhang, Z. & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods* **41**, 1083–94.

Zimmerman, I., Steiner, V., & Pond, R. (1992). *PLS-3: Preschool Language Scale-3*. San Antonio, TX: The Psychological Corporation.