J Autism Dev Disord (2016) 46:899–909
DOI 10.1007/s10803-015-2632-1

ORIGINAL PAPER

Language Differences at 12 Months in Infants Who Develop Autism Spectrum Disorder

DeWayne C. Lazenby[1,9] · Georgios D. Sideridis[1] · Noelle Huntington[1] · Matthew Prante[2,11] · Philip S. Dale[3] · Suzanne Curtin[4] · Lisa Henkel[4] · Jana M. Iverson[5] · Leslie Carver[6] · Karen Dobkins[6] · Natacha Akshoomoff[6] · Daina Tagavi[7,10] · Charles A. Nelson III[1,8] · Helen Tager-Flusberg[7]

Published online: 17 October 2015

Abstract Little is known about early language development in infants who later develop autism spectrum disorder (ASD). We analyzed prospective data from 346 infants, some of whom were at high risk for developing ASD, to determine if language differences could be detected at 12 months of age in the infants who later were diagnosed with ASD. Analyses revealed lower receptive and expressive language scores in infants who later were diagnosed with ASD. Controlling for overall ability to understand and produce single words, a Rasch analysis indicated that infants who later developed ASD had a higher degree of statistically unexpected word understanding and production. At 12 months of age, quantitative and qualitative language patterns distinguished infants who later developed ASD from those who did not.

Keywords Autism spectrum disorder (ASD) · Expressive/receptive vocabulary · Item response theory · Infant-sibling · MacArthur-Bates Communicative Developmental Inventories (CDI)

Introduction

Language development is vital to learning, communication and social interaction. Infants often recognize single words as early as 6 months of age (Bergelson and Swingley 2012) and produce words as early as 8 months of age (Fenson et al. 2007). For infants 12–18 months of age, single word use characterizes the "first words" phase of spoken language acquisition (Tager-Flusberg et al. 2009). In typical language development, there is wide individual variability. Based on parent report using the CDI: Words and Gestures, males progress at a slightly slower rate than females (though single word production is not significantly different until 16 months) and there is a pattern of receptive

Electronic supplementary material The online version of this article (doi:10.1007/s10803-015-2632-1) contains supplementary material, which is available to authorized users.

✉ DeWayne C. Lazenby
dlazenby@gmail.com

1 Division of Developmental Medicine, Boston Children's Hospital, Harvard Medical School, 300 Longwood Ave, Boston, MA 02115, USA

2 Utah State University, Logan, UT, USA

3 University of New Mexico, Albuquerque, NM, USA

4 University of Calgary, Calgary, AB, Canada

5 University of Pittsburgh, Pittsburgh, PA, USA

6 University of California, San Diego, CA, USA

7 Boston University, Boston, MA, USA

8 Harvard Graduate School of Education, Cambridge, MA, USA

9 Present Address: Joint Base Elmendorf Richardson Hospital, 1234 Juneau Ave, Jber, AK 99505, USA

10 Present Address: Harvard University, Cambridge, MA, USA

11 Present Address: Charlotte-Mecklenburg School System, Charlotte, NC, USA

language advancing ahead of expressive language (Fenson et al. 2007).

ASD is a common, yet heterogeneous neurodevelopmental disorder often accompanied by impairment in some aspects of language. Language acquisition is an important focus in ASD research and an important variable for predicting outcomes (Tager-Flusberg et al. 2009). Despite its common co-occurrence, language delay is neither necessary nor specific to ASD. While many children with ASD have significant language impairments, others have age-appropriate or advanced language abilities on standardized measures (Tek et al. 2014). However, they may have other language differences that are not easily detected on standardized language measures.

Infant-Sibling Studies

ASD affects 1 in 68 children (Center for Disease Control and Prevention 2014) and the recurrence risk in later-born siblings of children with ASD is close to 1 in 5, or 18.7 % (Ozonoff et al. 2011). Prospective studies of later-born, high-risk siblings offer a unique opportunity to study early language before ASD is diagnosed through comparisons between three groups of infants: high-risk infant siblings who later develop ASD (HR ASD), high-risk infant siblings who do not develop ASD (HR Non-ASD), and low risk infants who have only typically developing siblings and do not develop ASD (LR Non-ASD). This prospective, longitudinal approach, beginning in infancy, reduces the recall and sampling biases present in prior retrospective studies. The use of both HR Non-ASD and LR Non-ASD comparison groups is particularly important to studies of early communication and language development as the developmental and linguistic environment of high-risk infants differs from low-risk infants in several ways. Families members of a child with ASD are more likely to have social and communication deficits (Ruser et al. 2007), as well as higher rates of the broader autism phenotype in fathers and siblings (Schwichtenberg et al. 2010). Although not specific to ASD, other family factors that may differ include quality of sibling role model, differential parental treatment, level of family stress, and resource dilution (McHale et al. 2012). In addition, 20 % of HR Non-ASD infants have higher ASD severity and lower levels of developmental functioning (Messinger et al. 2013), further differentiating HR Non-ASD and LR Non-ASD comparison groups.

Early Language Development in Infants Who Later Developed ASD

Some infant-sibling studies investigated language development using both direct assessment and parent report (Hudry et al. 2014; Mitchell et al. 2006; Zwaigenbaum et al. 2005) while others used direct assessment only (Landa and Garrett-Mayer 2006; Ozonoff et al. 2014) prior to 18 months in the context of known ASD developmental outcomes by 24 or 36 months of age. The HR ASD sample sizes were small (7–17) in the mixed method studies and varied (24 and 51) in the direct assessment studies. In all of the above except Ozonoff et al. (2014) the age range was wide (11–18 months) at the time of their early language assessment. All of the studies directly assessed infants with the Receptive and Expressive Language subscales of the Mullen Scales of Early Learning (Mullen; Mullen 1995), which measure comprehension of phrases and questions as well as gesture, jargon and word production. All studies that included parent report data used the MacArthur-Bates Communicative Developmental Inventories (CDI: Words and Gestures; Fenson et al. 2007), which includes checklists for 28 early phrases and 396 commonly used words.

In two studies, receptive language was lower in HR ASD infants relative to comparison groups on both the Receptive Language subscale of the Mullen and understanding of phrases on the CDI (Mitchell et al. 2006; Zwaigenbaum et al. 2005). Landa and Garrett-Mayer (2006) found lower Receptive Language on the Mullen in 24 HR ASD infants relative to Non-ASD infants, but not relative to a language delayed comparison group without ASD. Conversely, Hudry et al. (2014) reported no differences on the Mullen in Receptive Language for HR ASD infants relative to comparison groups. None of the studies reported a significant difference in single word understanding on the CDI.

Expressive language scores on the Mullen were lower in HR ASD infants relative to comparison groups for some studies (Landa and Garrett-Mayer 2006; Mitchell et al. 2006) but not for others (Zwaigenbaum et al. 2005; Hudry et al. 2014). Similar to receptive single word understanding, none of the studies reported a difference in single word production. The inconsistencies in the results of these infant-sibling studies may be due to their small sample sizes as well as the range of actual ages of participants at the time of their visit.

Only one infant-sibling study to date examined relevant language data at 12 months of age in a large sample. Ozonoff et al. (2014) reported on 420 high- and low-risk infants as part of an investigation of the early emergence of the broader autism phenotype (BAP). They found that the Receptive and Expressive Language subscales scores of the Mullen were lower beginning at 12 months of age in HR ASD infants relative to low- and high-risk typically developing infants as well as to high-risk infants with other developmental concerns, such as Global Developmental Delay or BAP.

After 18 months of age, evidence converges to a greater degree, pointing to lower scores on expressive and receptive language in HR ASD infants relative to a variety of comparison groups, Barbaro and Dissanayake (2012; language/developmental delay), Landa and Garrett-Mayer (2006; LR Non-ASD), Mitchell et al. (2006; LR and HR Non-ASD), and Ozonoff et al. (2014; LR and HR Non-ASD, Non-typically developing). The earliest age at which lower single word comprehension scores has been reported was 18 months (Mitchell et al. 2006). Thus it has been hypothesized that word acquisition slows for children with ASD in the second year of life when advanced learning strategies are required (Jones et al. 2014). However, some studies have reported no difference at 18 months (Hudry et al. 2014; Talbott et al. 2013), perhaps due to the wide variation of language abilities within the HR ASD group, and small sample sizes. In summary, prospective studies have reported inconsistent findings with regard to language delays on standard measures in high-risk infants who later develop ASD. Furthermore, no prospective studies have investigated *qualitative* language differences in a high-risk population.

Studies Conducted After ASD Diagnosis

Several studies using cross-sectional or retrospective methodologies have investigated language development in young children with ASD relative to typically developing infants or developmentally delayed toddlers. These studies find several elements of language development that are similar for children with or without ASD, such as a wide variation in individual vocabulary size that increases with age, noun predominance, and similar vocabulary composition by semantic category and word class (Charman et al. 2003; Weismer et al. 2010;Weismer et al. 2011; Luyster et al. 2007). These studies have also identified variations from typical development, including a higher proportion of severe language delay, greater variation in vocabulary growth rates and weaker associations between vocabulary size and grammatical complexity (Charman et al. 2003; Weismer et al. 2010; Luyster et al. 2007). Among preschoolers with ASD, studies show greater impairment in their receptive language than their expressive language. (Barbaro and Dissanayake 2012; Weismer et al. 2010; Hudry et al. 2010, 2014).

Only one study has investigated lexical composition at the word level in children with ASD (Rescorla and Safyer 2013) using the Language Development Survey (LDS; Rescorla 1989), a parent report measure of expressive language. The authors concluded that the children with ASD were acquiring essentially the same words as children with typical development. However, the typically developing children were younger (age 2;0–2;4 than the ASD group (age 1;6–5;11). In addition, ~20 % of the children in the ASD group were rated as non-autistic on the Childhood Autism Rating Scale and it is unknown on what basis they were given an ASD diagnosis. While this study had the strength of a large sample size and replicated prior findings with regard to noun predominance and vocabulary composition by semantic category and word class, comparison between the ASD and typically developing groups is limited by the characteristics of the groups.

Current Study

This paper presents a prospective investigation of early language development in infants who were later diagnosed with ASD compared with infants who were not. Due to the inconsistent language findings, as assessed by the Mullen and CDI, in prior ASD studies with small samples of infants 11–18 months of age, we first analyzed standardized measures in a large sample of infant-siblings with a narrowly defined age range. Based on the findings by Ozonoff et al. (2014), we hypothesized that the magnitude and prevalence of receptive and expressive language delay in infants who later develop ASD would be sufficient to distinguish the HR ASD group at 12 months of age from comparison groups. The wide variation in linguistic ability found in individuals with ASD complicates the interpretation, but may provide clarity to the question of whether language development slows in the second year of life or is present from the beginning, particularly for infants with more language impairment. Second, we attempted to identify qualitative lexical differences at 12 months of age specific to the HR ASD group, irrespective of language ability, by means of an item-level Rasch analysis. We hypothesized that specific word use at the group level could distinguish the HR ASD group from comparison groups.

Methods

Participants

Data were pooled from four independently funded research sites from the Baby Siblings Research Consortium. Each site employed similar recruitment strategies, sampling methodologies, inclusion/exclusion criteria and standardized diagnostic assessment procedures. Families were recruited from clinics, community events, media announcements, mailings, and word-of mouth.

All participants in the high-risk group were identified as being the full biological younger sibling of a proband with an ASD diagnosis confirmed by either an ADOS and clinical best estimate (for 321 probands) or parent report (for 25 probands). In addition, inclusion required the absence of genetic or neurological conditions known to cause or increase risk for ASD. Infants in the low-risk group had an older typically developing sibling and no first-degree relatives with an ASD diagnosis.

For inclusion in the sample for our analyses, younger siblings had to be enrolled prior to 12 months of age, complete the CDI at the 12-month visit, and have diagnostic outcome data at 36 months of age. ASD outcome was made by Clinical Best Estimate. Of the initial sample of 348 infants, 213 were high-risk (HR) and 135 were low-risk (LR) and their outcomes fell into one of four categories: (1) high-risk infant siblings who later developed ASD (HR ASD; n = 43), (2) high-risk infant siblings who did not develop ASD (HR Non-ASD; n = 170), (3) low-risk infants without ASD (LR Non-ASD; n = 133) and (4) low-risk infants who developed ASD (n = 2; not included in analysis).

Mullen data from the 12-month visit were available for 287 of the infants who had concurrent MCDI data. One site did not administer the Mullen to low risk participants. Analysis showed no differences in race, ethnicity, gender, maternal education, language ability or diagnostic outcome between infants who did versus did not have Mullen data. Informed consent was obtained from all participants and each institutions' Institutional Review Boards approved the study. Participant characteristics are presented in Table 1.

Measures

Early Language

The MacArthur-Bates Communicative Developmental Inventories (CDI; Fenson et al. 2007) Words and Gestures form is a standardized parent-report measure of early communication abilities, assessing first signs of understanding, understanding of 28 everyday phrases, starting to talk and a vocabulary checklist. The vocabulary inventory of 396 commonly acquired words are divided into 19 semantic categories (e.g., animals, actions, body parts, etc.), and parents indicate whether each word is "understood" or "understood and produced" by their infants. The CDI is a reliable, validated instrument standardized for use in typically developing infants 6–16 months of age. It is the most widely used parent-report instrument of early language development in children with ASD (Tager-Flusberg et al. 2009). As reported in the CDI Technical Manual (Fenson et al. 2007), total CDI expressive vocabulary is strongly correlated with direct measures of production, including language samples and tests, with correlations in the range of 0.6–0.8. There are fewer studies of the validity of the comprehension measure, in part due to the lack of good comprehension measures to serve as a 'gold standard' at this early age. However, total CDI receptive vocabulary does show substantial correlation with other measures of early receptive language, with correlations in the range of 0.5–0.6.

The present study focused on the vocabulary checklist, for which the data was processed with two-pass verification. The American version of the CDI was used at all four sites; there were no significant dialect differences and no word substitutions were used. Raw scores for comprehension and production were used; comprehension is computed by adding what the parent reports the infant 'understands and says' and 'understands', which assumes that words produced are also understood.

Mullen Scales of Early Learning (Mullen; Mullen 1995)

The Mullen is a direct assessment tool measuring five areas of functioning: visual reception, expressive language, receptive language, fine motor and gross motor. The two language subscale T-scores were dependent variables in our analysis and the non-verbal ability T-score (average of the fine motor and visual reception) was a covariate. The

Table 1 Sample characteristics

Variable	High-risk		Low-risk	
	ASD (n = 43)	Non-ASD (n = 170)	Non-ASD (n = 133)	Chi square
Sex (% male)	73.5 %	53.6 %	50.6 %	5.49
Age in months (mean, SD)	11.97 (0.64)	11.98 (0.44)	11.84 (0.43)	1.109
Age range	11–13	11–13	11–13	
Hispanic (%)	21.7 %	10.4 %	15.8 %	2.295
Non-Caucasian (%)	17.1 %	10.7 %	22.4 %	5.999
Maternal education (% H.S.)[†]	27.6 %	8.5 %	6.8 %	10.782*
Mullen nonverbal ability (mean, SD)	50.7 (9.5)	55.6 (8.2)	56.8 (7.6)	16.334*
Mullen early learning composite (mean, SD)	90.8 (15.2)	101.4 (14.0)	106.2 (11.4)	7.091*

[†] H.S.—High School: schooling terminated at or before high school completion

* Significant at $p \leq 0.05$

receptive language subscale tests for understanding of phrases like "give it to me" (with gesture), "give it to mommy" (with no gesture). It also tests for comprehension of questions (about a chair and a door) and following directions (with a block and a car). The expressive language subscale tests for the ability to play gesture/language games, first words, jabbering with inflection, combining jargon and gestures, and words and gestures. The reliability and validity of the Mullen is well established as well as its use in evaluating infants with ASD. Relevant to the present study, a high degree of agreement has been found between the Mullen and the CDI in measuring receptive and expressive language (Luyster et al. 2008).

Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000)

The ADOS is a standardized assessment of ASD symptoms across the following domains: social interaction, communication, repetitive and stereotyped behaviors, and play. The ADOS yields a total score and clinical cut-off scores for use in the diagnosis of ASD.

Clinical Best Estimate (CBE)

CBE diagnoses were made or verified by licensed clinicians when infants reached 36 months of age and were informed by ADOS scores, DSM-IV criteria, and cognitive and behavioral assessments. Clinical diagnoses were dichotomized into either ASD (including pervasive developmental disorder—not otherwise specified and autistic disorder) or Non-ASD.

Data Analysis

Firstly, in order to test the hypothesis of quantitative language differences, our data were analyzed using Analyses of Covariance (ANCOVA). CDI "Vocabulary Comprehension" and "Vocabulary Production" as well as Mullen "Expressive Language" and "Receptive Language" T-score point estimates were compared across groups, after first controlling for site and variables known to be related to a child's language development (i.e. maternal education, race, gender, and nonverbal ability).

Secondly, to test the hypothesis of qualitative language differences, we used a more contemporary approach on word-level data from the CDI. The Rasch model was employed, specifically the Differential Item Functioning (DIF) procedure, in order to evaluate the likelihood that a word "behaves" in a significantly different manner in non-ASD outcome (LR and HR) versus ASD outcome (HR-A) groups (see Appendix A of Supplementary Material for analytical details). In order to avoid the potential problem of multiple comparisons, we relied on effect size metrics along with significant findings. The decisive factor to determine an interpretable effect was the effect size (>1.0 logit in difference scores) accompanied with a statistic that exceeded conventional levels of significance.

Differential Item Functioning Analysis

The differential probability of success between populations is termed Differential Item Functioning (DIF) analysis in the Item Response Theory (IRT) context (Holland and Thayer 1988). Based on Dorans and Holland (1993): "… …DIF is an unexpected difference among groups of examinees who are supposed to be comparable with respect to attribute measured by the item and the test on which it appears" (p. 37) and its major advantage lies in the fact that differential ability between groups is controlled for. It suggests that the rank-ordering of the words (from easy to difficult) is different in one group compared to another group independent of underlying language ability per group. In other words, a word that is easy or difficult for one group does not hold a similar level of difficulty for another group.

For illustration purposes, Fig. 1 displays hypothetical Item Characteristic Curves (ICCs) for HR ASD infants and LR Non-ASD infants with regard to their ability to produce a word. Each curve has two values of importance. The first is the Ability value (x axis) at the point that the curve crosses 0.5 Probability of Success (y axis) which indicates the language ability required for infants to have a 50 % probability of producing or understanding the word. An Ability value of zero indicates average ability, positive values indicate above average ability and negative values below average. As Fig. 1 shows (stars), the level of ability required for the HR ASD infants to be successful 50 % of the time with this word was equal to +1.00 logits (above average ability) which was significantly higher compared to zero logits (average ability) for the LR Non-ASD infants. Stated in terms of the Ability Contrast, an ability one logit higher was required by the HR ASD group to be equally successful on this word compared with the LR Non-ASD group.

The second value of importance is the Probability of Success at the point that the curve crosses an Ability value of zero (average ability) which indicates the probability of the word being used for infants of average language ability. As Fig. 1 shows (triangle), HR ASD infants of average ability have a 15 % probability of using the word successfully, compared to 50 % for the LR Non-ASD infants. For more information about the estimation method of DIF and the respective effect size employed see Appendix A of Supplementary Material.

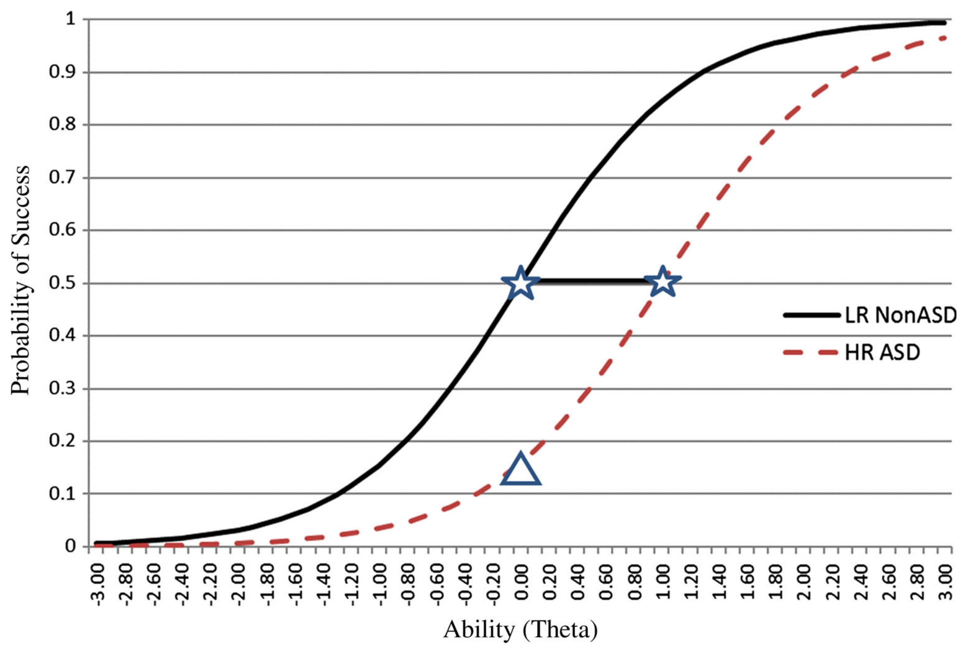

Fig. 1 Hypothetical Item Characteristic Curves (ICCs) for a word in which performance in the two populations showed significant Differential Item Functioning (DIF, equivalent to 1 logit or greater). Specifically, DIF favored the LR Non-ASD group in that the ability levels required by them to be successful on that word 50 % of the time (*stars*) were equal to zero logit (average ability) compared to +1 logit (higher than average ability) for the HR ASD group. In addition, children of average language ability (zero logit) in the HR ASD group have only a 15 % probability of using this word (*triangle*)

Power Analysis for Rasch Model

Power was simulated using Mplus version 7.1 (Muthén and Muthén 2006) for a Rasch model using the full sample size and 200 items per subscale for which there were full data with ample variance (non-constant). The program file is in Appendix B of Supplementary Material. Results indicated that the mean square error for the estimates of item difficulties ranged between 0.0241 and 0.347, which is negligible. Power levels ranged between 92.6 and 97.2 %, and thresholds were significantly different from zero 100 % of the time.

Sample Size Considerations

Following a mean square error analysis, valid vocabulary data were available for 333 infants for words understood. The 13 eliminated cases were severely violating the Guttman pattern and the main theoretical premises of the Rasch model with outfit mean square values greater than Z = 2.0. These cases were essentially degrading measurement and were distorting the meaning of each subscale, thus, only contributing measurement error to the measures of words produced and words understood (Wu and Adams 2007). For that reason, they were eliminated as recommended by pioneers of the Rasch model methodology, following three iterations (Wu and Adams 2007). Dropped participants did not differ significantly from those included in the study on gender [$\chi^2(1) = 2.265$, $p = 0.139$], infant age [$t(336) = 0.658$, $p = 0.511$], ethnicity [$\chi^2(5) = 1.710$, $p = 0.888$], maternal education [$\chi^2(4) = 4.983$, $p = 0.289$], nonverbal ability [$t(284) = 0.250$, $p = 0.803$], words produced [$F(1, 343) = 1.229$, $p = 0.268$], group distribution [$\chi^2(2) = 3.863$, $p = 0.145$], or site distribution [$\chi^2(3) = 4.383$, $p = 0.223$]. Significant differences were observed with regard to words understood [$F(1, 344) = 5.441$, $p = 0.020$], dropped participants had higher scores on number of words understood.

Results

Mean Comparison Between Groups on Language Ability

As shown in Table 2, the HR ASD group had significantly lower Mullen T-scores relative to HR Non-ASD and LR Non-ASD comparison groups on both the Receptive Language and Expressive Language subscales. Both differences remained significant controlling for site and factors potentially influencing language development, such as gender, race, ethnicity, and maternal education. On the CDI, a statistically significant difference was detected in single word understanding with the ASD group having significantly lower scores from both the HR non-ASD group and the LR group; there were no significant differences between the HR non-ASD and LR groups. In contrast, there were no significant differences in single-word production, possibly due to a floor effect. After controlling for gender, race, ethnicity, maternal education, nonverbal ability, and site, differences between groups remained. Single word understanding remained significant and single word production continued to be non-significant. The mean numbers of words produced and understood by group are shown in Table 2.

Table 2 Comparisons between groups on language measures

Score	High-risk		Low-risk	Unadjusted	Adjusted
	ASD (N = 43) Mean (SD) Range N	Non-ASD (N = 170) Mean (SD) Range N	Non-ASD (N = 133) Mean (SD) Range N		
Mullen receptive language T-score	$37.61_{b,c}$ (8.66) 20–55 39	43.57_{a} (10.23) 20–79 148	46.78_{a} (9.99) 31–80 99	$F(2, 281) = 11.86$ $p < 0.001$ (E.S†. = 0.078)	$F(2, 277) = 9.74$ $p < 0.001$
Mullen expressive language T-score	$41.26_{b,c}$ (11.20) 20–78 39	46.817_{a} (11.44) 20–78 148	51.13_{a} (9.02) 33–76 98	$F(2, 282) = 12.82$ $p < 0.001$ (E.S†. = 0.083)	$F(2, 278) = 11.79$ $p < 0.001$
CDI single word understanding raw	$35.88_{b,c}$ N = 32	50.54_{a} N = 165	69.19_{a} N = 126	$F(2, 330) = 8.30$ $p < 0.05$ (E.S†. = 0.048)	$F(2, 250) = 8.24$ $p < 0.01$
CDI single word production raw	2.05 N = 37	5.84 N = 155	6.06 N = 120	$F(2, 309) = 2.36$ $p = 0.096$ (E.S†. = 0.015)	$F(2, 241) = 1.09$ $p = 0.34$

Mullen–Mullen Scales of Early Learning, CDI-MacArthur-Bates Communicative Developmental Inventories. Subscripts a, b, and c represent the 3 groups HR ASD, HR non ASD and Low Risk, respectively, and denote with which groups a particular value has a significant difference. The presence of any subscripts after a mean estimate suggest significant differences between the respective groups

† Estimates of effect size involve the eta square statistic, which is analogous to the R^2 statistic. Values of 0.01, 0.06 and 0.14 are indicative of small, medium, and large effects, respectively

Differential Single Word Production and Understanding Between Groups

Tables 3 and 4 present the results from the DIF analysis for words produced and understood, respectively, with only significant findings shown. A combination of DIF contrast values greater than |1.0| (Wang and Chen 2004) and the presence of a statistically significant finding were used to determine the presence of meaningful DIF between two populations. Given that the ASD group is the reference group, a positive 'DIF Ability Contrast' (comparison group Ability minus reference group Ability) indicates that the specific word was more likely to be produced/understood for the ASD group, compared to LR or HR Non-ASD

Table 3 Differential item functioning analysis for words produced

Word	HR ASD		Non-ASD comparison			Ability contrast††	More likely for…	Rasch-Welch t statistic
	Ability†	Ability standard error	Group	Ability	Ability standard error			
Baabaa	−5.31	.49	LR	−3.60	.28	1.71	HR ASD	3.02**
Block	−2.54	1.06	LR	1.45	1.68	3.99	HR ASD	2.01*
Bye	−5.31	.49	LR	−4.10	.25	1.21	HR ASD	2.19*
Baabaa	−5.31	.49	HR	−3.89	.25	1.42	HR ASD	2.55**
Uncle	−2.54	1.08	HR	1.08	1.14	3.61	HR ASD	2.30*
Hello	−3.32	.77	HR	−1.17	.53	2.15	HR ASD	2.31*
Bite	−2.54	1.06	HR	1.08	1.14	3.62	HR ASD	2.32*

† Ability level measured in logits, (*DIF* differential item functioning). The more negative an ability estimate for a group the more likely it was produced by the specified group. Only meaningful differences are shown DIF > 1 and $p < 0.05$

†† Ability contrasts calculated as HR ASD Ability minus comparison group Ability. Absolute values greater than one indicate large effect size (see Appendix A of Supplementary Material)

* Significant at $p \leq 0.05$, ** $p < 0.01$

Table 4 Rasch-based analysis of words understood using the differential item functioning procedure

Word	HR ASD		Non-ASD comparison			Ability contrast††	More likely for…	Rasch-Welch t
	Ability†	Ability SE	Group	Ability	Ability SE			
Baabaa	−2.84	.42	LR	−1.69	.22	1.15	HR ASD	2.42**
Pig	−1.84	.49	LR	.00	.32	1.84	HR ASD	3.13**
Bubbles	−2.48	.44	LR	−1.43	.23	1.05	HR ASD	2.11*
Coat	−.96	.61	LR	1.50	.54	2.46	HR ASD	3.01**
Playpen	.02	.82	LR	2.29	.74	2.27	HR ASD	2.05*
Blanket	−3.18	.41	LR	−1.94	.22	1.25	HR ASD	2.70*
Watch	−.54	.69	LR	1.50	.54	2.04	HR ASD	2.32*
Brother	−4.28	.39	LR	−2.26	.21	2.02	HR ASD	4.53**
Child	.02	.82	LR	3.01	.98	2.99	HR ASD	2.3*
Out	−1.30	.56	LR	.35	.36	1.65	HR ASD	2.49**
Open	−2.07	.47	LR	−.83	.26	1.24	HR ASD	2.31*
t-Brush	−1.38	.23	LR	.92	1.11	−2.29	LR	−2.02*
Cheerio	−3.51	1.00	HR	−2.51	.20	.45	HR ASD	2.24*
Boots	.02	.82	HR	4.15	1.72	4.13	HR ASD	2.16*
Pants	−1.30	.56	HR	.03	.33	1.32	HR ASD	2.04*
Hand	−2.07	.47	HR	−.90	.26	1.16	HR ASD	2.18*
S' name	−1.84	.49	HR	−.37	.30	1.47	HR ASD	2.56**
Child	.02	.82	HR	4.17	1.75	4.15	HR ASD	2.15*
Open	−2.07	.47	HR	−.97	.25	1.10	HR ASD	2.06*
Hungry	−2.48	.44	HR	−1.40	.23	1.07	HR ASD	2.16*

† Ability level measured in logits, (*DIF* differential item functioning). The more negative an ability estimate for a group the more likely it was for the specified group to understand the word. Only meaningful differences are shown DIF > 1 and $p < 0.05$

†† Ability contrasts calculated as HR ASD Ability minus comparison group Ability. Absolute values greater than one indicate large effect size (see Appendix A of Supplementary Material)

* Significant at $p \leq 0.05$, ** $p < 0.01$

groups, and a negative DIF indicates the reverse. For example, it was more likely for HR ASD infants to produce the word "block" (−2.54 logits—below average ability) compared to LR Non-ASD infants (1.45 logits—above average ability; Fig. 2). The DIF Ability Contrast between the two groups equaled 3.99 logits, in favor of the ASD group [$t(312) = 2.01$, $p < 0.05$]. Viewed from the probability perspective, HR ASD infants were significantly more likely to produce "block" (>90 % chance of production for infants of average ability), compared to the LR Non-ASD group (<10 % chance of production for infants of average ability). All seven words produced with meaningful DIF Ability Contrast showed bias in favor of the HR ASD group (i.e. they were statistically more likely to produce the words) as did all but one of the 18 words understood.

Discussion

Our study has two key findings. First, we detected significant quantitative language differences on two measures in HR ASD infants at 12 months. Second, we had the apparently contradictory finding that the HR ASD group, despite lower overall language ability, was 'more likely' to produce and understand certain words in a statistically unexpected way. There were several words, either produced or understood, that were used significantly more frequently by HR ASD infants than by either comparison group. In contrast, only one word was used statistically less frequently by HR ASD infants.

Our quantitative Mullen findings replicated those of Ozonoff et al. (2014) as well as the smaller prior studies showing significant differences at 12 months of age (Landa and Garrett-Mayer 2006; Mitchell et al. 2006; Zwaigenbaum et al. 2005). Our findings also were consistent with research demonstrating no single word production difference on the CDI (Hudry et al. 2014; Mitchell et al. 2006; Zwaigenbaum et al. 2005). However, unlike prior infant sibling studies, our study did detect significant differences in single word understanding, with a significantly lower count of words understood by HR ASD infants on the CDI. Our study used a larger sample size with more power and a more narrowly defined age range which may have improved our chances of detecting existing differences at 12 months.

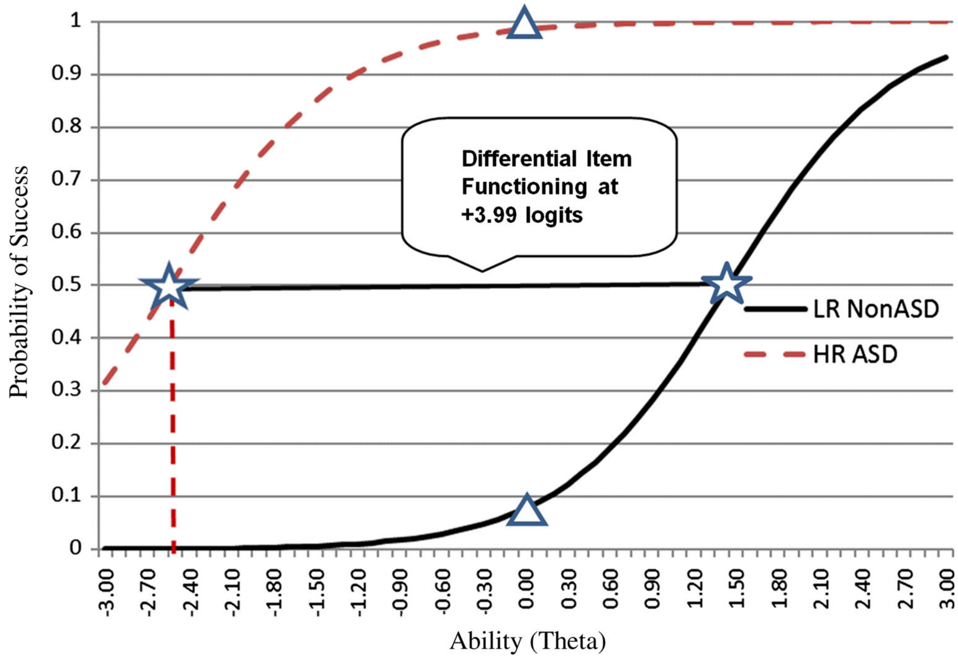

Fig. 2 Actual Item Characteristic Curves (ICCs) for the word "block." The probability of success was significantly higher for the HR ASD group compared to the LR Non-ASD group. In other words, the word "block" required a much lower level of ability (−2.54 logits—below average) for the HR ASD group to be produced compared to the LR Non-ASD group (+1.45—above average), with a differential item function of +3.99 (*stars*). Infants of average ability in the LR Non-ASD group had a <10 % probability of success, relative to a >90 % probability of success for the HR ASD group (*triangles*)

Our second finding, and its analytical approach, is novel among infant sibling studies. However, its findings have some similarities to a large retrospective cross-sectional comparison of Differential Item Functioning in a sample of typically developing young toddlers (<18 months of age) with older children who had received an ASD diagnosis (Bruckner et al. 2007). This analysis investigated only words *understood* and found that some words were more likely to be understood by children with ASD and some words by typically developing infants. The reasons proposed by Bruckner for the differences between groups were age differences, orienting defects, social communication defects, and restricted object use in children with ASD. The latter three reasons may also underlie our findings, though study sample and design differences limit direct comparison.

Only one prior study has investigated lexical differences at the word level (Rescorla and Safyer 2013) and concluded, based on the large degree of overlap among the most frequently used words, that lexical composition was delayed but not atypical in children with ASD. Our study does not dispute the large degree of similarity in lexicons or lexical development, but aims to focus the discussion on the degree of lexical differences present in infants with ASD. Indeed, DSM 5 removed a delay in spoken language as a diagnostic criterion for ASD, but expanded the number of criteria that can be met by atypical language use, such as poorly integrated verbal communication, deficits in conversation, stereotyped, repetitive speech, idiosyncratic phrases, ritualized verbal behavior, and repetitive questioning (APA 2013). In line with this diagnostic redirection, our study focused on the atypical quality of early lexical differences that may be present as early as 12 months of age.

Our study is significant in that it suggests that lexical differences can be detected at 12 months of age, 6 months earlier than previously reported. This suggests that distinguishing language differences are present at 12 months in contrast to the proposal that the rate of word learning slows for toddlers with ASD in the second year of life (Jones et al. 2014).

Strengths of the present study include a more narrowly defined age, a large sample size, multiple sites, clearly defined low- and high-risk groups, prospective assessments on multiple measures, and valid ASD outcome determination. One potential limitation is our focus on younger siblings of children with ASD. The younger siblings who receive their own diagnosis of autism may not be representative of the larger population of children with ASD. The use of the HR Non-ASD comparison group helps to control for some of the potential home environment differences between the HR ASD children and the larger population of children with ASD, particularly differences due to the presence of an older sibling with ASD. However, the HR ASD children also are all, by definition, from multiplex families which should remain a consideration when interpreting our results due to the higher presence of autistic-like traits among unaffected family members in multiplex families (Bernier et al. 2012; Schwichtenberg et al. 2010).

A second potential limitation of our study was reliance on a vocabulary checklist, which does not reveal the context, frequency, accuracy, or range of extension of a child's language. For example, the word "bye" was more likely to be produced by the HR ASD infants. Though "bye" is a word most often used socially at the end of an interaction, infants with ASD may begin saying "bye" shortly after an interaction has begun and repeatedly throughout, possibly indicating they would like to terminate the interaction. More information about word usage is needed to appropriately interpret the meaning of an individual word's differential function. In addition, the CDI was designed to capture the most common words spoken by typically developing children. Atypical words, which may occur more frequently in children with ASD, are not part of the instrument. Therefore, reliance on the CDI may not capture all potential language differences at 12 months of age, possibly underestimating true differences. Future studies could use a bottom up approach, such as language diaries or recordings, to identify atypical words used by infants who later develop ASD. Such studies could also serve as a first step in developing a more sensitive language measure that could accurately capture individual differences or perhaps even a language screening tool in order to apply these findings clinically. Finally, examining individual words on a large instrument increases risk for a series of Type-I errors. To limit this risk we used the more conservative criteria of large effect size plus statistical significance for establishing DIF.

In conclusion, we have extended prior research by demonstrating both qualitative and quantitative elements of infants' expressive and receptive language that distinguish infants who later develop ASD. Early language, like many other aspects of development, may reveal differences and/or deficits consistent with an autism diagnosis as early as 12 months of age.

Acknowledgments We would like to thank the children and their families for their participation and the talented staff at the four study sites. Many thanks to the Scholarly Oversight Committee of the Division of Developmental Medicine for their guidance. We are grateful to the anonymous reviewers of the initial version of this manuscript. This research was supported by the National Institute of Health (NIH) Grant 1R01DC010290-02 and a Simon's foundation Grant 137186 to C.A. Nelson and H. Tager-Flusberg, NIH Grant R01 HD052804-01A2 to K. Dobkins and R01 HD054979 and R01 HD41607 to J. Iverson, MCHB training grant T77 MC 00011-10 to L. Rappaport, as well as a grant from the Alberta Centre for Child, Family, and Community Research to S. Curtin. A poster presentation of a pilot of this study from one site was given IMFAR 2014 and PAS 2014. A poster with a portion of the data in this manuscript was presented at IMFAR 2015 and PAS 2015.

Author Contributions DL, NH, MP, PD, CAN and HTF conceived of the study, participated in its design and coordination, analysis and interpretation of data, drafted the manuscript, and revised it for important intellectual content. DT participated in study coordination as well as drafted and critically revised the manuscript. GS made substantial contributions to study conception and design, drafted and critically revised the manuscript, and performed the statistical analysis. SC, LH, JI, LC, KD, NA made substantial contributions to study design, acquisition of data, analysis and interpretation of data, and critically reviewed the manuscript for important intellectual content. All authors read and approved the final manuscript.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington: American Psychiatric Association.

Barbaro, J., & Dissanayake, C. (2012). Developmental profiles of infants and toddlers with autism spectrum disorders identified prospectively in a community-based setting. *Journal of Autism and Developmental Disorders, 42*(9), 1939–1948. doi:10.1007/s10803-012-1441-z.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America, 109*(9), 3253–3258. doi:10.1073/pnas.1113380109.

Bernier, R., Gerdts, J., Munson, J., Dawson, G., & Estes, A. (2012). Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. *Autism Research, 5*(1), 13–20.

Bruckner, C., Yoder, P., Stone, W., & Saylor, M. (2007). Construct validity of the MCD-I receptive vocabulary scale can be improved: Differential item functioning between toddlers with autism spectrum disorders and typically developing infants. *Journal of Speech, Language, and Hearing Research., 50*, 1631–1638.

Centers for Disease Control and Prevention (2014) Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 Sites, United States, 2010. MMWR March 28, 2014/63(SS02);1–21.

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*(1), 213–236.

Cohen, J. (1995). A power primer. *Psychological Bulletin, 112*, 155–159.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Harcourt.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenzel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Paul H. Brookes Publishing Company.

Hambleton, R., Swaminathan, H., & Robers, H. (1991). *Fundamentals of item response theory*. NY: Sage.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Inc.

Hudry, K., Chandler, S., Bedford, R., Pasco, G., Gliga, T., Elsabbagh, M., et al. (2014). Early language profiles in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(1), 154–167. doi:10.1007/s10803-013-1861-4.

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., & Charman, T. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language and Communication Disorders, 45*(6), 681–690. doi:10.3109/13682820903461493.

Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience and Biobehavioral Reviews, 39*, 1–33. doi:10.1016/j.neubiorev.2013.12.001.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*(6), 629–638. doi:10.1111/j.1469-7610.2006.01531.x.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., & Rutter, M. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(8), 1426–1438. doi:10.1007/s10803-007-0510-1.

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*(3), 623–654.

McHale, S. M., Updegraff, K. A., & Whiteman, S. D. (2012). Sibling relationships and influences in childhood and adolescence. *Journal of Marriage and the Family, 74*(5), 913–930.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127–143.

Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., et al. (2013). Beyond autism: A baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(3), 300.e1–308.e1. doi:10.1016/j.jaac.2012.12.011.

Mitchell, S., Brian, J., Zwaigenbaum, L., Roberts, W., Szatmari, P., Smith, I., & Bryson, S. (2006). Early language and communication development of infants later diagnosed with autism spectrum disorder. *Journal of Developmental and Behavioral Pediatrics: JDBP, 27*(2 Suppl), S69–S78.

Mullen, E. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service Inc.

Muthén, L. K., & Muthén, B. O. (2006). *MPLUS: Statistical analysis with latent variables: User's guide* (4th ed.). Los Angeles, CA: Muthén and Muthén.

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., et al. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry, 53*(4), 398.e2–407.e2. doi:10.1016/j.jaac.2013.12.020.

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics, 128*(3), e488–e495. doi:10.1542/peds.2010-2825.

Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement, 29*, 150–151.

Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43*, 295–312.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*, 353–368.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods, 8*, 164–184.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *The Journal of Speech and Hearing Disorders, 54*(4), 587–599.

Rescorla, L., & Safyer, P. (2013). Lexical composition in children with autism spectrum disorder (ASD). *Journal of Child Language, 40*(1), 47–68. doi:10.1017/S0305000912000232.

Ruser, T. F., Arin, D., Dowd, M., Putnam, S., Winklosky, B., Rosen-Sheidley, B., et al. (2007). Communicative competence in parents of children with autism and parents of children with specific language impairment. *Journal of Autism and Developmental Disorders, 37*(7), 1323–1336. doi:10.1007/s10803-006-0274-z.

Schwichtenberg, A. J., Young, G. S., Sigman, M., Hutman, T., & Ozonoff, S. (2010). Can family affectedness inform infant sibling outcomes of autism spectrum disorders? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 51*(9), 1021–30. doi:10.1111/j.1469-7610.2010.02267.x.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., & Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research: JSLHR, 52*(3), 643–652. doi:10.1044/1092-4388(2009/08-0136).

Talbott, M. R., Nelson, C. A., & Tager-Flusberg, H. (2013). Maternal gesture use and language development in infant siblings of children with autism spectrum disorder. *Journal of Autism and Developmental Disorders,*. doi:10.1007/s10803-013-1820-0.

Tek, S., Mesite, L., Fein, D., & Naigles, L. (2014). Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(1), 75–89. doi:10.1007/s10803-013-1853-4.

Wang, W., & Chen, H. (2004). The standardized mean difference within the framework of item response theory. *Educational and Psychological Measurement, 64*, 201–223.

Weismer, S. E., Gernsbacher, M. A., Stronach, S., Karasinski, C., Eernisse, E. R., Venker, C. E., & Sindberg, H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders, 41*(8), 1065–1075. doi:10.1007/s10803-010-1134-4.

Weismer, S. E., 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. doi:10.1007/s10803-010-0983-1.

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2–3), 143–152. doi:10.1016/j.ijdevneu.2004.05.001.