1-2013

Acoustic and Perceptual Measurements of Prosody Production on the Profiling Elements of Prosodic Systems in Children by Children with Autism Spectrum Disorders

Joshua John Diehl
University of Notre Dame and Haskins Laboratories

Rhea Paul
Sacred Heart University, paulr4@sacredheart.edu

Recommended Citation

Diehl, J.J.& Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. Applied Psycholinguistics, 34(1), 135-161. doi: 10.1017/S0142716411000646

Applied Psycholinguistics **34** (2013), 135–161
doi:10.1017/S0142716411000646

Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders

JOSHUA JOHN DIEHL
University of Notre Dame and Haskins Laboratories, New Haven, CT

RHEA PAUL
Yale University School of Medicine

Received: May 16, 2010 Accepted for publication: December 29, 2010

ADDRESS FOR CORRESPONDENCE
Joshua Diehl, Department of Psychology, University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556. E-mail: joshua.diehl@nd.edu

ABSTRACT
Prosody production atypicalities are a feature of autism spectrum disorders (ASDs), but behavioral measures of performance have failed to provide detail on the properties of these deficits. We used acoustic measures of prosody to compare children with ASDs to age-matched groups with learning disabilities and typically developing peers. Overall, the group with ASD had longer utterance durations on multiple subtests on a test of prosodic abilities, and both the ASD and learning disabilities groups had higher pitch ranges and pitch variance than the typically developing group on one subtest. Acoustic differences were present even when the prosody was used correctly. These findings represent differences in the fine details of the acoustic output beyond its functional interpretation in *both* clinical groups.

Autism spectrum disorder (ASD) is a disorder of social communication (American Psychiatric Association [APA], 2000) that affects approximately 1 in 110 children (Kogan et al., 2009). The "autism spectrum" refers to several disorders, including autism, Asperger syndrome, and pervasive developmental disorder—not otherwise specified (PDD-NOS). Autistic disorder is defined by primary deficits in social interaction, communication, and repetitive behaviors/stereotyped interests. Asperger syndrome also has primary deficits in social interaction and repetitive behaviors/stereotyped interests, but individuals in this group show no early language delays, although they do exhibit difficulties in the pragmatic use of language. Individuals with PDD-NOS have difficulties with one or more of the areas described above, but do not meet full criteria for autistic disorder or Asperger syndrome. Research suggests that current definitions do not create reliable categorical differentiations between these closely related conditions (Howlin, 2003; Miller & Ozonoff, 2000; Volkmar, Lord, Bailey, Schultz, & Klin, 2004), and therefore they are often grouped together as ASD.

Deficits in the pragmatic use of language are a hallmark characteristic of ASD (Paul, Orlovski, Marcinko, & Volkmar, 2009; Tager-Flusberg, Paul, & Lord, 2005; Young, Diehl, Morris, Hyman, & Bennetto, 2005). Atypicalities in prosody production, including rhythm, rate, and intonation patterns, are some of the most commonly reported social–communicative features of the disorder (McCann & Peppé, 2003; Paul, Augustyn, Klin, & Volkmar, 2005; Shriberg et al., 2001) and also some of the earliest characteristics to appear (Schoen, Paul, & Chawarska, 2010; Werner, Dawson, & Osterling, 2000; Wetherby et al., 2004). Prosodic atypicalities have been reported clinically and in research at all levels of ability in ASD, including Asperger syndrome (Shriberg et al., 2001), and high-functioning[1] autism (Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007). Prosody production has been shown to be related to measures of general social–communicative functioning in ASD (Paul, Shriberg, et al., 2005) and it has been suggested that it is a "bellwether" behavior (i.e., microcosm) that is indicative of the specific cognitive and social functioning profile of an individual with the disorder (Diehl & Berkovits, 2010). The perceived differences in prosodic patterns is considered to be one of the most stigmatizing aspects of the disorder (Shriberg et al., 2001). Nevertheless, research has not provided detailed characterization of these deficits that goes beyond the subjective impressions observed clinically. In this paper, we examine acoustic characteristics of prosody in children with ASD in order to identify these subtle, but clinically relevant characteristics of speech. Furthermore, we attempt to investigate the specificity of these deficits in comparison to individuals with learning disabilities (LDs) and typically developing (TD) peers.

PROSODY PRODUCTION IN AUTISM AND RELATED DISORDERS

Studies of prosody production have had mixed findings, and most studies focus on the functional use of prosody, rather than the "fine details" of the acoustic output that could identify subtle (but perceivable) group differences. For example, prosody serves many functions in speech, including (but not limited to) the communication of affect, the structuring of discourse, or the clarification of syntactic structure. In addition, prosodic patterns could be exaggerated or monotone, fast or slow, and although they are functionally correct, these "fine details" of the acoustic output can provide the signature, or "idiolect," of an individual's speech.

A review of early research on prosody performance found that few studies had sufficient sample sizes for groupwise comparisons, and those that had sample sizes greater than 20 tended to have wide age ranges and/or inconsistent diagnostic criteria (McCann & Peppé, 2003). Paul, Augustyn, and colleagues (2005) recently used an experimental paradigm to address some of these issues and found that children with ASD showed deficits in stress production, particularly in using stress at the lexical level (e.g., RE-call vs. re-CALL) and at the phrasal level (e.g., CHOCOLATE ice cream vs. chocolate ICE CREAM) to communicate meaning. The group with ASD, however, performed similar to that of typical peers on several measures, including the grammatical and pragmatic/affective production of intonation and phrase breaks.

Another series of studies used a test of prosodic abilities, the Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppé & McCann, 2003) to examine prosodic performance in ASD. One study found that all participating children with ASD had difficulty in at least one aspect of prosodic ability (McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007). Another study showed group differences in specific aspects of the PEPS-C (Peppé et al., 2007). Peppé and colleagues (2007) found that children with ASD had more difficulty using prosody to show affect or phrasal level stress (similar to Paul et al., 2007) when compared to typical comparisons matched on verbal mental age. Still, many individuals with ASD did very well on these tasks and the lack of group differences in multiple areas was more surprising than the few subtests that found group differences.

Although some recent studies have found prosody production atypicalities in children with ASD, these studies do not appear to fully capture the extent of prosody production deficits in this population. A surprising finding in many of these studies is that individuals with ASD, especially those considered high functioning, have performed well on many of the tasks. Moreover, there is evidence that speakers with other language-related disabilities also exhibit difficulties in the prosody performance (Catterall, Howard, Stojanovik, Szcerbinski, & Well, 2006; Marshall, Harcourt-Brown, Ramus, & van der Lely, 2009; Stojanovik, Setter, & van Weijk, 2007; Wells & Peppé, 2003), although these are less commonly reported and the investigation of their characteristics is less detailed. Because of these issues, many studies have noted the importance of using acoustic analysis, rather than behavioral observation/judgment, in order to measure more subtle prosodic differences that may be missed by other techniques (Diehl et al., 2009; McCann & Peppé, 2003; Paul, Augustyn, et al., 2005).

ACOUSTIC ANALYSIS OF PROSODY PRODUCTION IN ASD

Several studies have used acoustic measures to create experimental stimuli in studies of prosody comprehension in ASD (Chevallier, Noveck, Happé, & Wilson, 2008; Järvinen-Pasley, Paisley, & Heaton, 2008), but only a handful of peer-reviewed published studies have acoustically examined prosody production in this population. The use of acoustic analysis to understand prosodic deficits in impaired populations is both important and challenging (Green & Tobin, 2009; Peppé, 2009). One issue is that the coding and analysis of prosody is more difficult than other linguistic devices, and many of the available techniques are very time consuming. Still, the availability of automated computer software programs, such as PRAAT (Boersma & Weenink, 2009), has allowed for more complex analyses of speech in disordered populations.

Acoustic analyses are important for several reasons. First, they afford an objective measure of speech performance that does not rely on the subjective judgments of raters. Second, acoustic analysis might allow for a more fine-grained sensitivity to group differences. Third, acoustic analysis allows the researcher to examine the details of elements of production rather than forcing an assignment of each production to a binary category. Fourth and finally, acoustic analysis of prosody has the potential for the automated processing and classification of the speech signal. This would allow for treatment paradigms in which individuals with speech disorders could receive instrumental (and possibly instantaneous) feedback for speech errors.

Several studies have examined acoustic characteristics of prosody in ASD, but only recently have there been large-scale studies of group differences. Early studies found acoustic properties in echolalic speech in ASD (Loca & Wootton, 1995; Paccia & Curcio, 1982) that were important clues to how children with ASD were using echoed speech to communicate. Two early studies used acoustic measures of prosody in nonecholalic speech (Baltaxe, 1984; Fosnot & Jun, 1999), but these studies had very small sample sizes (four to five participants) and only provided an initial indication that there were prosodic differences that could be measured instrumentally.

One of the first large-scale studies to acoustic analyses was conducted by Paul and colleagues and examined imitated speech in 44 youth and adults with ASD (ages 7–28) in comparison to 20 individuals with typical development (Paul, Bianchi, Augustyn, Klin, & Volkmar, 2008). Participants listened to and imitated stressed and unstressed nonsense syllables. Paul and colleagues (2008) found that participants with ASD showed significantly less difference in duration between stressed and unstressed syllables than the TD comparison group. They also found a tendency for pitch range to be larger for participants with ASD for both stressed and unstressed syllables, although this difference failed to reach statistical significance.

A second study (Green & Tobin, 2009) with a much smaller sample size examined spontaneous speech from elicited conversations and speech collected from paragraphs that were read aloud. This study compared children 10 Hebrew-speaking children with high-functioning ASD between the ages of 9 and 13 and 10 typical peers matched on chronological age and language level. The authors concluded that children with ASD could produce a wide range of prosodic contours, but chose a limited repertoire that were repeatedly used in conversations.

Diehl and colleagues (2009) examined aspects of spontaneous prosodic production in narratives in two samples of individuals with ASD: one with children and one with adolescents. They found increased fundamental frequency (f0) variance in the groups with ASD when compared to typical comparison groups matched on chronological age, IQ, and measures of verbal abilities. Moreover, acoustic measures correlated with clinical judgments of autism-specific communication impairments. One drawback to this study is that the narratives were entire story retellings, and it was difficult to compare across specific acoustic characteristics of prosody (e.g., duration, f0 range) because story retellings can differ (sometimes dramatically) in their length, content, and degree of expressed emotion, among other important narrative elements.

In sum, there has been a recent resurgence in the use of acoustic measurement to understand prosodic performance in ASD. Studies have acoustically analyzed echolalic, imitative, and spontaneous speech in conversation and narratives. These studies have found important prosodic differences in f0 variance, duration of syllables, and prosodic contours. Still, these studies have examined a limited number of the acoustic features that are involved in prosody production. Moreover, several studies have examined prosodic performance across functions (e.g., to communicate affect, to structure discourse, or to determine syntactic structure), but these have generally not included acoustic measures. Thus, our knowledge of acoustic differences shown by children with ASD across these different functions is limited. Finally, none of these studies have compared individuals with ASD to a clinical comparison group without ASD in order to determine which features are specific to ASD, and which are related to developmental or learning delays.

PURPOSE OF THE STUDY

The purpose of the present study is to investigate prosody in children and adolescents with ASD using acoustic analysis of speech in comparison to a group with typical development and one with LDs. We chose the PEPS-C to elicit speech because (a) it has been used in several studies to examine prosodic performance in ASD as well as in language impairment (Wells & Peppé, 2003), (b) it provides data on perception and production, (c) it provides probes on several functions of prosody, and (d) the production content is standardized across participants, which allows for comparisons of multiple acoustic characteristics of prosody. In this study, we examine several acoustic characteristics of prosody (pitch range, pitch variance, duration, intensity) in PEPS-C tasks measuring four basic functions of prosody (affect, turn-end type, chunking/syntax, stress). Based on previous studies, we predicted acoustic differences in prosody production between children with ASD and the two comparison groups. Based on previous research, we predict that participants with ASD will have larger pitch ranges and greater pitch variance than both control groups, and that there would be differences in duration, although we did not have a prediction on the direction of duration differences. Based on clinical reports, we also predict acoustic differences in the intensity of speech, although these analyses are exploratory based on the paucity of previous research on this characteristic. We predicted that differences would be present across different functions, and present both for correct and incorrect responses, and that this would indicate differences in the fine details of the acoustic output, whether the function was judged as correct by listeners.

METHODS

Participants

ASDs. Participants in this group were 24 individuals with ASD between the ages of 8 and 16 (see Table 1 for descriptive characteristics). Participants with ASD were selected for the study from a database of children who had participated in either clinical or research activities at the Yale Child Study Center during the 5 years prior to data collection for the present study. Diagnostic characterization included the Autism Diagnostic Interview—Revised (Rutter, Le Couteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule—Generic (Lord et al., 2000). Each participant met *DSM-IV-TR* (APA, 2000) criteria for one of the three ASDs (autistic disorder, Asperger syndrome, or PPD-NOS). Clinical diagnoses

Table 1. *Descriptive characteristics of the sample*

	ASD			LD			TD				
	M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range	*F*	*p*
N	24			16			22				
Gender (M/F)	16:8			12:4			15:7				
Chronological age	12.31	2.32	8–16	12.99	2.25	9–16	12.21	2.64	9–17	0.55	.58
CELF-IV											
Receptive Language Index	93.67	19.49	58–121	88.73	17.63	58–119				0.64	.43
Expressive Language Index	100.54	16.22	75–126	90.00	14.95	65–114				4.13	.05
Core language	97.21	18.61	67–132	88.94	16.02	60–117				0.82	.37
Nonverbal IQ[a]	103.61	17.14	75–133	96.85	11.13	67–109				1.54	.22

Note: CELF-IV, Clinical Evaluation of Language Fundamentals, Fourth Edition. ASD, autism spectrum disorder; LD, learning disability; TD, typically developing.

[a]Nonverbal IQ was measured using either the performance IQ scaled score from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) or the Nonverbal IQ Scale from the Differential Ability Scales (Elliott, 1990). Some nonverbal IQ scores were missing for participants, so for these analyses $N = 18$ for the ASD group and $N = 13$ for the LD group.

were confirmed independently by two experienced clinicians. Interrater reliability between these clinicians for diagnostic assignment was high, with kappa values ranging from 0.80 to 0.95 in related research projects (Klin, Lang, Cicchetti, & Volkmar, 2000). Children with ASD were given a language evaluation using the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) in order to determine the current level of language functioning. Participants were excluded from this study if had any uncorrected sensory disorders (vision or hearing) or any known neurological conditions.

Learning disabilities. Participants included 16 children between the ages of 9 and 16 (see Table 1 for descriptive characteristics). Participants in this group were recruited through a community speech–language pathologist. For inclusion in the LD group, children needed to have a clinically diagnosed learning disability and needed to be at the appropriate grade level for their chronological age. Per parent report, 5 participants were identified as having a reading disability, and 11 were identified as having a language-based learning impairment, although the type of impairment was not specified. All were screened via clinical interview and found not to exhibit ASD. Family history showed no evidence of a first degree relative with an ASD diagnosis. Children with LD were also given a brief cognitive and language evaluation. The groups were matched on chronological age, and had similar nonverbal IQs, CELF-4 core language and receptive language scores. Because many participants in the group with LD had a language-based LD, the ASD group tended to have higher scores on both measures, and had significantly higher CELF-4 expressive language scores.

TD comparison group. Participants included 22 children between the ages of 8 and 17 (see Table 1 for descriptive characteristics). TD participants were recruited from the community. All participants in this group had typical development as reported by their parents, had no first-degree relatives with an ASD, no previous history of clinical diagnosis or special educational services, and were reported to be in the appropriate grade for their age in school. The TD group was matched to the ASD and LD groups on chronological age.

Procedures

Setup and equipment. Participants were seated at a table on which there was a Dell Inspiron 3900 computer that contained the PEPS-C program. Participants wore a Shure SM10A professional unidirectional head-mounted dynamic microphone, which was connected to a TASCAM US-122 USB Audio/MIDI Interface. The TASCAM connected directly into the Dell computer. Recordings had a sampling rate of 44.1 kHz.

Stimuli. Participants completed the PEPS-C (Peppé & McCann, 2003), an instrument designed to assess prosody performance in children aged 4–16 (Wells & Peppé, 2004). This measure contains 12 subtests divided into six categories (affect, turn-end type, chunking, focus, prosody, intonation). For each of the categories, there is an "input" (comprehension) and "output" (production) subtest. Four of the categories (affect, turn-end type, chunking, focus) are "function" categories that test a participants' ability to understand and use prosody in a way that communicates a specific function, such as an affective state. Two of the categories (prosody, intonation) are "form" categories, and are comprised of simple pitch/melody discrimination and productive imitation. Although comparisons between form and function can elucidate many questions about the nature of prosodic deficits, these comparisons are beyond the scope of this paper and will be addressed in a separate report. The PEPS-C has been used to gather some normative data from TD children (Peppé & McCann, 2003), and trained listeners' perceptual judgments of PEPS-C responses have been used in several studies to investigate prosody performance in children with a range of disabilities (e.g., Catterall et al., 2006; Marshall et al., 2009; McCann et al., 2007; Peppé et al., 2007; Peppé & McCann, 2003; Stojanovik et al., 2007; Wells & Peppé, 2003). The PEPS-C is perhaps the most widely used standard measure of prosodic function in this literature (Peppé, 2009).

Data collection. For each of the subtests, the experimenter demonstrated the correct answer on two training trials in order to explain the task to the participant. The participant then completed two additional practice items, during which the experimenter would correct the participant for incorrect responses in order to ensure that the participant understood the task. Each subtest contained 16 experimental trials. Behavioral data was collected for the input subtests. Participants indicated their response by clicking on one of two possible response choices, and the computer program automatically coded the response as correct or incorrect. For the output subtests reported here (see Table 2), participants' responses were judged as correct or incorrect by a trained examiner. For all subtests that involved examiner judgment, responses from a randomly selected 10% sample of participants were independently scored by a second examiner. Average point to point reliability for correct/incorrect judgments on each subtest across participants ranged from 0.84 to 0.96, with an average agreement of 0.88 across subtests.

Acoustic analyses. Output responses were audio-recorded, and vocalizations from the critical trials on all of the function output subtests were examined for the following acoustic characteristics of prosody: average f0 across the entire utterance, the difference between maximum and minimum f0, or f0 range (sometimes called *accent range*; Paul et al., 2008), standard deviation (*SD*) of f0 (as measured in Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008), utterance duration, and utterance intensity. We used PRAAT, a program for speech analysis and synthesis (Boersma & Weenink, 2009), for acoustic analyses. The computer automatically divided sound files into individual trials. In PRAAT, text grids were created to delineate the beginning and end of each vocalization. PRAAT scripts were created to extract automatically the desired acoustic characteristics for each trial.

RESULTS

Data analysis plan

All analyses were conducted using a one-way analysis of variance (ANOVA) with group membership (ASD, LD, TD) as the independent variable and either

Table 2. *Description of PEPS-C output tasks*

	Child Task	Examiner Score
Affect	Child sees picture of a food (e.g., cabbage); say name of food with prosody that expresses like or dislike; also indicates like/dislike by pressing button	The examiner, who could not see the participant's response, pressed a button to indicate whether the participant liked or disliked the food based on prosody. The prosody was considered to be correct if the examiner's button press matched the participant's.
Turn end	Child sees picture of a person speaking about a food (e.g., cabbage) and is instructed to say word as if they were asking a question (e.g., *cabbage?*) or reading the word in a book (e.g., *cabbage.*) based on a picture cue accompanying food; Half the items were intended to elicit a question (i.e., final rise intonation), half a statement	Examiner, who could not see the participant's screen, pressed a button to indicate whether child utterance was perceived as a question or statement. The statement was considered correct if the examiner's response matched the response that the item was intended to elicit and incorrect if the two did not match or the child's response was ambiguous.
Chunking	Short phrases (e.g., chicken fingers and fries) accompanied by two (chicken fingers and fries) or three chicken, fingers, and fries) pictures; Child describes what is seen	Judge whether the phrase contained two or three elements, based on pause placement
Focus	Participants saw a picture (black sheep with ball) and heard a sentence that did not match the picture (e.g., "The black cow has the ball"); participants asked to correct the speaker ("No, the black SHEEP has the ball")	The examiner judged whether the stress was on the adjective (BLACK sheep) or noun (black SHEEP).

Note: PEPS-C, Profiling Elements of Prosodic Systems in Children.

behavioral response accuracy or one of the five acoustic measures as the dependent variables. An omnibus one-way ANOVA was conducted including all three groups. Because of the large number of measurements, we only conducted paired group comparisons (i.e., ASD vs. LD, ASD vs. TD, LD vs. TD) if the omnibus F was significant ($p < .05$) in order to reduce the likelihood of Type I error. Effect sizes were calculated as partial eta squared ($\eta^2_{partial}$), which refers to the proportion of variance attributable to a given effect, after partialing out other nonerror sources of variance (Cohen, 1973). Behavioral response data appear in Figure 1 and Figure 2. Statistical test results on acoustic measures for all subtests appear in

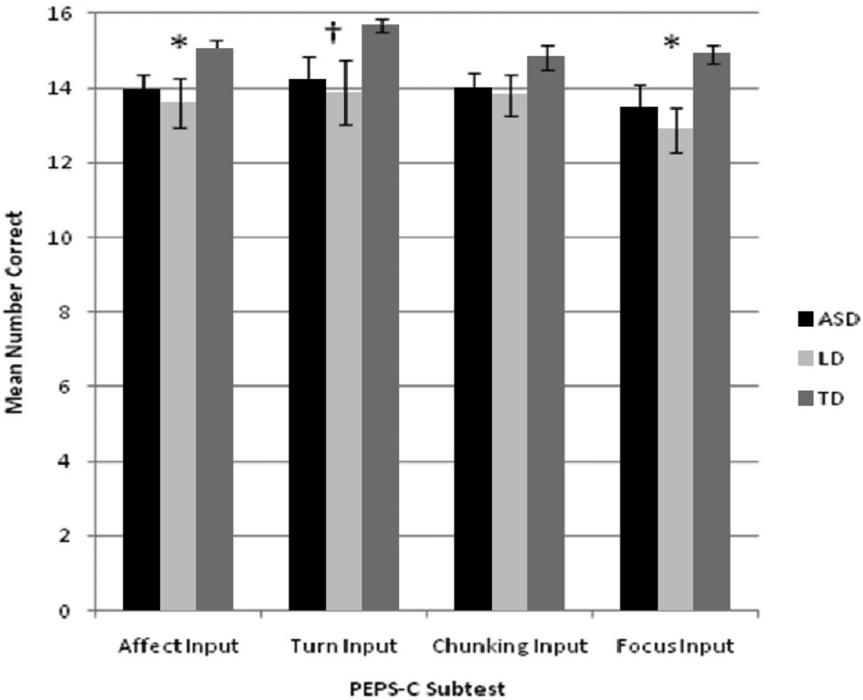

Figure 1. The mean number correct on behavioral responses from Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppé & McCann, 2003) input (comprehension) tasks. The statistical tests represent an omnibus analysis of variance for each subtest. †$p < .10$. *$p < .05$.

Table 3 and Table 4, and means and standard errors appear in Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7.

Our original plan was to separately investigate acoustic differences on correct and incorrect responses for each subtest. We were unable to conduct analyses on incorrect responses for any item, however, because a substantial minority of participants in each group did not miss any items on any one particular subtest. This greatly reduced the sample sizes for incorrect responses, and limited our ability to detect even large effect sizes. We instead conducted analyses on all responses (correct and incorrect), and then on only correct responses.

Affect

Affect behavioral findings. In this subtest, we examined the ability of participants to recognize types of affective prosody, and their ability to use prosody to indicate whether they liked or disliked a food item. One-way ANOVAs revealed that the groups differed on the affect input subtest, F (2, 59) = 3.45, $p < .05$, $\eta^2_{partial} = 0.11$, but not on the affect output subtest, F (2, 59) = 1.85, $p = .17$,

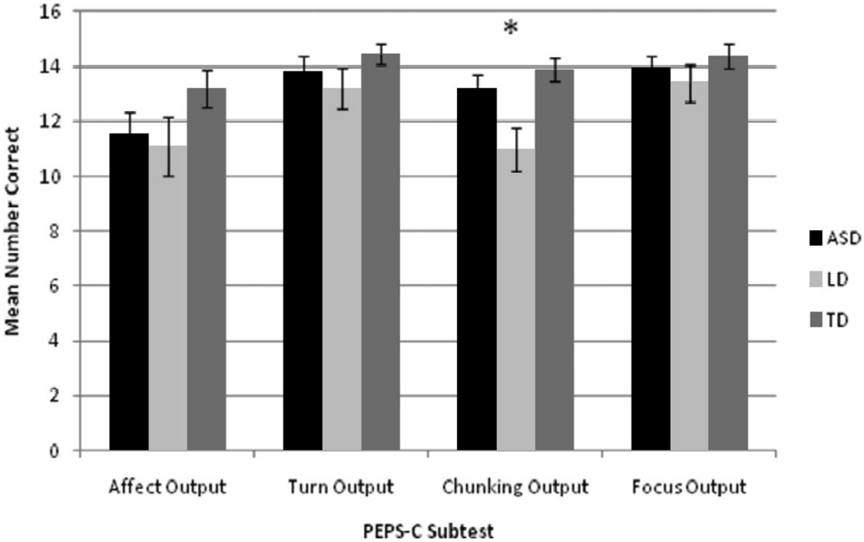

Figure 2. The mean number correct on behavioral responses from Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppé & McCann, 2003) output (production) tasks. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. $^{*}p < .05$.

$\eta^2_{partial} = 0.06$. On the affect input subtest, the TD group performed significantly better than the group with ASD, F (1, 44) = 5.97, $p < .05$, $\eta^2_{partial} = 0.12$, and the group with LD, F (1, 36) = 5.90, $p < .05$, $\eta^2_{partial} = 0.14$, but the groups with ASD and LD did not differ in their performance, F (1, 38) = 0.27, $p = .61$, $\eta^2_{partial} = 0.007$.

Affect acoustic findings.

LIKE RESPONSES.[2] For all correctly interpreted items on which participants indicated that they liked the food, one-way ANOVAs revealed no significant group differences in average f0, f0 variability or range, utterance intensity, or utterance duration. We then examined responses that were intended to convey liking by the participant and were correctly interpreted as “like” by raters, and these findings were similar across groups.

DISLIKE RESPONSES.[3] For all items on which participants indicated that they did not like the food, one-way ANOVAs revealed no significant group differences on any of our acoustic measures. When only utterances that were intended to be and judged to be dislike responses (correct dislike) were analyzed, there was a significant group difference on utterance duration, but no significant group differences on the other acoustic measures. The group with ASD had significantly longer utterance duration on correct dislike responses than did the group with LD,

Table 3. *Analyses of variance on acoustic measures of fundamental frequency*

	Sig. Pairwise Diff.	Average F_0				F_0 Standard Deviation				F_0 Range			
		F	df	p	$\eta^2_{partial}$	F	df	p	$\eta^2_{partial}$	F	df	p	$\eta^2_{partial}$
Affect													
Like-all		0.96	2, 59	.39	0.03	2.65	2, 59	.08	0.08	1.01	2, 59	.37	0.03
Like-correct		0.85	2, 58	.44	0.03	0.77	2, 58	.18	0.06	0.90	2, 58	.41	0.03
Dislike-all		0.65	2, 58	.53	0.02	1.40	2, 58	.26	0.05	1.38	2, 58	.26	0.05
Dislike-correct		1.91	2, 57	.16	0.06	0.66	2, 57	.52	0.02	1.67	2, 57	.20	0.06
Turn-end													
Question-all	**ASD = TD > LD**	**4.43**	**2, 59**	**.05**	**0.13**	0.18	2, 58	.84	0.006	2.24	2, 59	.12	0.07
Question-correct	**ASD = TD > LD**	**5.67**	**2, 59**	**.01**	**0.17**	0.10	2, 59	.91	0.003	2.66	2, 59	.08	0.09
Statement-all		1.91	2, 59	.16	0.06	1.17	2, 59	.32	0.04	2.47	2, 59	.09	0.08
Statement-correct	**ASD = TD > LD**	**3.92**	**2, 59**	**.05**	**0.12**	0.36	2, 59	.70	0.01	1.81	2, 59	.17	0.06
Chunking													
All	**ASD = TD > LD**	**4.02**	**2, 59**	**.05**	**0.12**	1.99	2, 59	.15	0.06	1.69	2, 59	.19	0.05
Correct	**ASD = TD > LD**	**4.22**	**2, 59**	**.05**	**0.13**	1.50	2, 59	.23	0.05	1.37	2, 59	.26	0.05
Focus													
All	**F_0: ASD = TD > LD** ***SD* of F_0:** **ASD = LD > TD** **F_0 range: ASD > TD**	**4.96**	**2, 59**	**.01**	**0.14**	**3.40**	**2, 59**	**.05**	**0.10**	**4.54**	**2, 59**	**.05**	**0.13**
Correct	**F_0: ASD = TD > LD** ***SD* of F_0: ASD > TD** **F_0 range: ASD > TD**	**5.54**	**2, 59**	**.01**	**0.16**	**3.08**	**2, 59**	**.05**	**0.10**	**3.91**	**2, 58**	**.05**	**0.12**

Note: Bold groups are significantly different at $p < .05$. Pairwise difference between diagnostic groups for significant omnibus F is significant at $p < .05$. ASD, autism spectrum disorder; TD, typically developing; LD, learning disability.

Table 4. *Analyses of variance on acoustic measures of utterance duration and intensity*

		Utterance Duration				Utterance Intensity			
	Sig. Pairwise Diff.	*F*	*df*	*p*	$\eta^2_{partial}$	*F*	*df*	*p*	$\eta^2_{partial}$
Affect									
Like-all		2.86	2, 59	.07	0.09	0.85	2, 59	.43	0.03
Like-correct		0.47	2, 58	.63	0.02	1.11	2, 58	.34	0.04
Dislike-all		2.44	2, 58	.10	0.08	0.34	2, 58	.71	0.01
Dislike-correct	**ASD > LD = TD**	**4.77**	**2, 57**	**.01**	**0.14**	0.63	2, 57	.54	0.02
Turn-end									
Question-all	**ASD > TD**	**3.66**	**2, 59**	**.05**	**0.11**	2.14	2, 59	.13	0.07
Question-correct		2.92	2, 59	.06	0.09	2.49	2, 59	.09	0.08
Statement-all	**ASD > LD > TD**	**4.39**	**2, 59**	**.05**	**0.13**	1.65	2, 59	.20	0.05
Statement-correct	**ASD > LD > TD**	**5.00**	**2, 59**	**.01**	**0.14**	2.50	2, 59	.09	0.08
Chunking									
All		1.62	2, 59	.21	0.05	1.99	2, 59	.15	0.06
Correct		0.85	2, 59	.43	0.03	2.45	2, 59	.10	0.08
Focus									
All		1.98	2, 59	.15	0.06	2.82	2, 59	.07	0.09
Correct	**ASD > LD**	2.13	2, 59	.13	0.07	**3.33**	**2, 59**	**.05**	**0.10**

Note: Boldface groups are significantly different at $p < .05$. Pairwise difference between diagnostic groups for significant omnibus F is significant at $p < .05$. ASD, autism spectrum disorder; LD, learning disability; TD, typically developing.

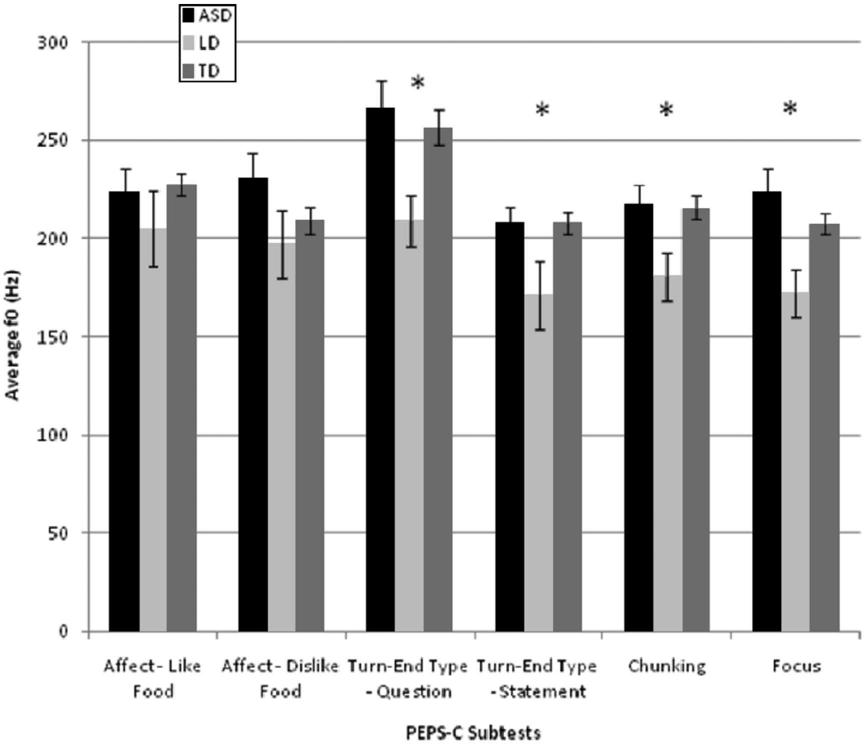

Figure 3. The average fundamental frequency, organized by task, for each participant group. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. PEPS-C, Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). *$p < .05$.

F (1, 36) = 4.00, $p < .05$, $\eta^2_{\text{partial}} = 0.10$, and the TD group, F (1, 42) = 7.56, $p < .01$, $\eta^2_{\text{partial}} = 0.15$. The LD and TD groups did not differ on this measure, F (1, 36) = 0.29, $p = .60$, $\eta^2_{\text{partial}} = 0.008$.

AFFECT SUMMARY. The TD group scored higher than the groups with LD and ASD on the affect input task, but there were no group differences in behavioral responses on the output task. The participants with ASD tended to produce longer utterances than other groups on items for which the participant intended to express dislike for the food and the rater interpreted the response as an expression of dislike (correct dislike responses). No other acoustic differences were observed among groups.

Turn-end type

Turn-end type behavioral findings. In this subtest, we measured participants' ability understand when prosody was being used to ask a question, and their

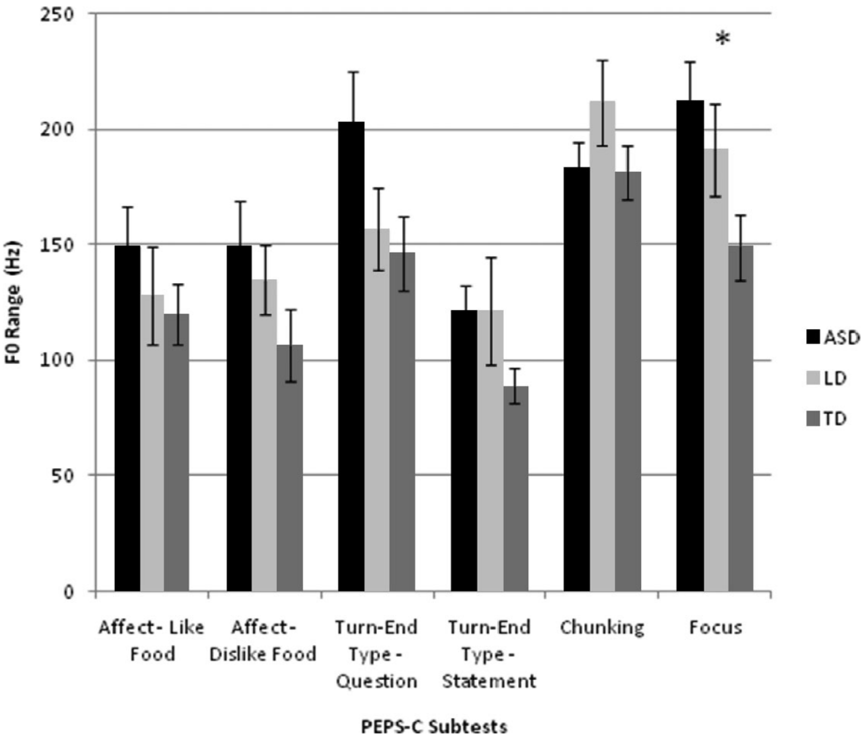

Figure 4. The fundamental frequency range for correct responses, organized by task, for each participant group. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. PEPS-C, Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). $^{*}p < .05$.

ability to use prosody to indicate a question or a statement. There was a marginal omnibus effect of group on the input task, F (2, 59) = 2.68, $p = .08$, $\eta^2_{\text{partial}} = 0.08$. The TD group scored significantly higher than the groups with ASD, F (1, 44) = 4.75, $p < .05$, $\eta^2_{\text{partial}} = 0.10$, and LD, F (1, 36) = 5.54, $p < .05$, $\eta^2_{\text{partial}} = 0.13$, but there was no difference in performance between the groups with ASD and LD, F (1, 38) = 0.10, $p = .75$, $\eta^2_{\text{partial}} = 0.003$. The three groups were not significantly different in their performance on the turn-end type output subtests, F (2, 59) = 1.18, $p = .31$, $\eta^2_{\text{partial}} = 0.02$.

Turn-end type acoustic findings.

QUESTIONS. For items that were designed to elicit rising intonation contours (questions), one-way ANOVAs revealed significant group differences in average f0 and utterance duration, but no significant differences in f0 range, *SD* of f0, or intensity. The group with LD had a significantly lower average f0 than the group with ASD, F (1, 38) = 6.52, $p < .05$, $\eta^2_{\text{partial}} = 0.15$, and TD, F (1, 36) = 8.90,

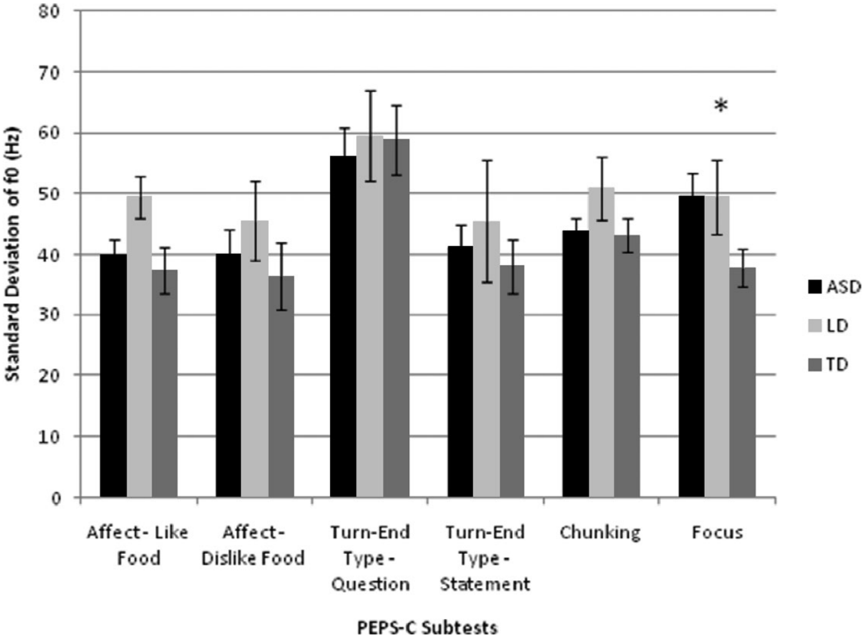

Figure 5. The standard deviation of the fundamental frequency for correct responses, organized by task, for each participant group. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. PEPS-C, Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). $^{*}p < .05$.

$p < .01$, $\eta^2_{partial} = 0.20$. The ASD and TD groups did not significantly differ on average f0, $F(1, 44) = 0.11$, $p = .74$, $\eta^2_{partial} = 0.003$. The ASD group produced utterances that were significantly longer in duration than the TD comparison group, $F(1, 44) = 5.58$, $p < .05$, $\eta^2_{partial} = 0.12$, and marginally longer than the group with LD, $F(1, 38) = 2.84$, $p < .10$, $\eta^2_{partial} = 0.07$. The LD and TD comparison groups were not significantly different on duration, $F(1, 36) = 0.73$, $p = .40$, $\eta^2_{partial} = 0.02$. We also examined vocalizations that were designed to elicit questions and the responses were interpreted as questions by the examiner (correct questions). Here, one-way ANOVAs revealed significant group differences in average f0, but not on any other acoustic measures. The group with LD had a significantly lower average f0 on correct questions than the group with ASD, $F(1, 38) = 8.56$, $p < .01$, $\eta^2_{partial} = 0.19$, as well as the TD comparison group, $F(1, 36) = 9.63$, $p < .01$, $\eta^2_{partial} = 0.22$. The ASD and TD groups did not differ on average f0, $F(1, 44) = 0.34$, $p = .56$, $\eta^2_{partial} = 0.008$.

STATEMENTS. For items that were designed to elicit responses with falling intonation contours (statements), one-way ANOVAs revealed a significant group difference in utterance duration, but not for any other acoustic variable. The group

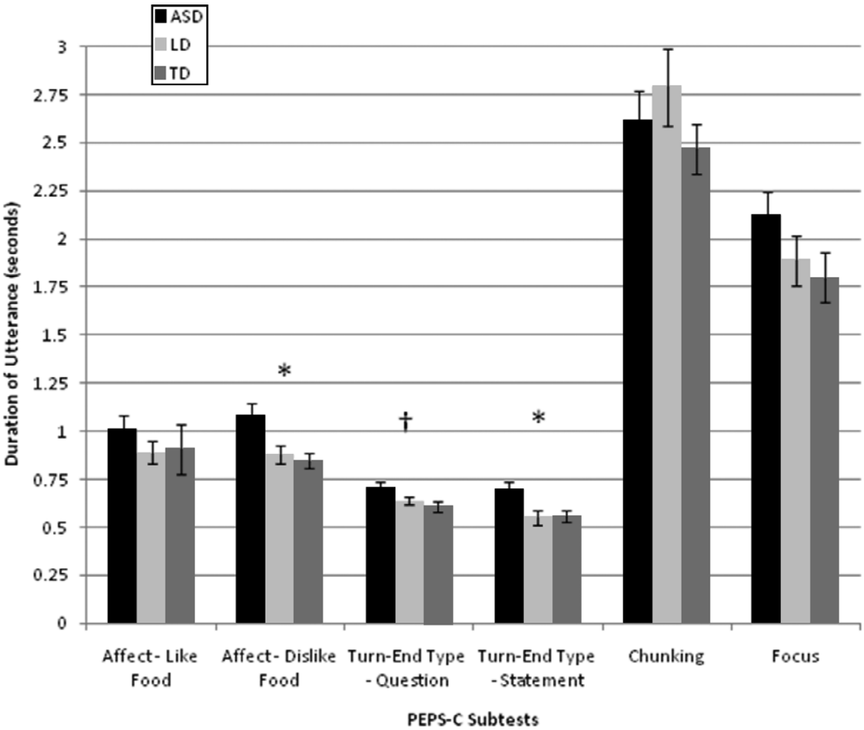

Figure 6. The average utterance duration for correct responses, organized by task, for each participant group. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. PEPS-C, Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). †$p < .10$. *$p < .05$.

with ASD had significantly longer utterance duration than the group with LD, F (1, 38) = 4.50, $p < .05$, $\eta^2_{\text{partial}} = 0.11$, and the TD group, F (1, 44) = 6.19, $p < .05$, $\eta^2_{\text{partial}} = 0.13$. The LD group had a significantly longer duration than the TD group, F (1, 35) = 9.63, $p < .01$, $\eta^2_{\text{partial}} = 0.22$. For items designed to elicit statements that were interpreted as statements by the raters (correct statements), one-way ANOVAs revealed significant overall group differences in average f0 and utterance duration, but no group differences on the other acoustic measures. On this comparison, the group with LD had a significantly lower average pitch than the group with ASD, F (1, 38) = 4.71, $p < .05$, $\eta^2_{\text{partial}} = 0.11$, or the TD group, F (1, 36) = 5.15, $p < .05$, $\eta^2_{\text{partial}} = 0.13$, while the ASD and TD groups did not differ significantly, F (1, 44) = 0.00, $p = .97$, $\eta^2_{\text{partial}} = 0.001$. For duration of utterance, however, the ASD group had significantly longer utterances than both the group with LD, F (1, 38) = 6.89, $p < .01$, $\eta^2_{\text{partial}} = 0.15$, and the TD group, F (1, 44) = 6.52, $p < .01$, $\eta^2_{\text{partial}} = 0.13$, but the LD and TD groups did not significantly differ on this variable, F (1, 35) = 0.15, $p = .70$, $\eta^2_{\text{partial}} = 0.004$.

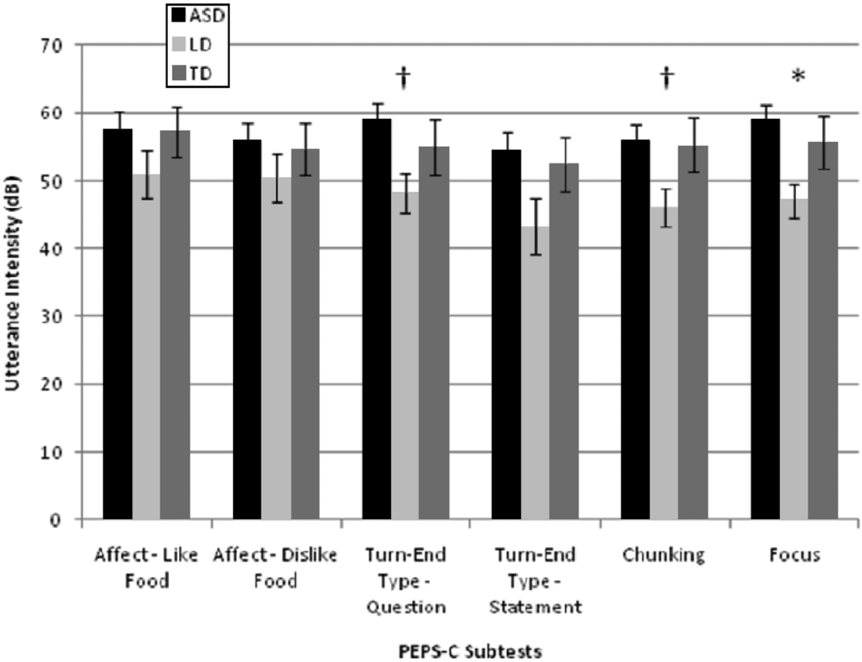

Figure 7. The average utterance intensity for correct responses, organized by task, for each participant group. The statistical tests represent an omnibus analysis of variance for each subtest. Error bars represent the standard error of the mean. PEPS-C, Profiling Elements of Prosodic Systems in Children (Peppé & McCann, 2003). †$p < .10$. *$p < .05$.

Turn-end type summary. Similar to the affect subtests, the groups with ASD and LD had more difficulty on the turn-end type input subtest than the TD group, but there were no group differences on the output subtest. In contrast to the affect subtests, we found fairly consistent group differences in two acoustic measures. Participants with ASD had consistently longer utterances than either group, both for questions and statements. Of interest, their responses were longer even for statements that were judged to be statements with correct prosody. The group with LD had a significantly lower average f0 than the other groups for questions and statements, even when they were judged to be correct in their usage of prosody.

Chunking

Chunking behavioral findings. In this subtest, we examined participants' ability understand prosodic phrase breaks, and their ability to use them to delineate the syntactic structure and meaning of the utterance. We tested group performance on the PEPS-C chunking input and output subtests using one-way ANOVAs. For chunking input subtests, there were no significant overall group differences, F (2, 59) $= 1.61, p = .21, \eta^2_{partial} = 0.05$. For output subtests, there were significant group differences, F (2, 59) = 6.12, $p < .01$, $\eta^2_{\text{partial}} = 0.17$. The LD group performed significantly worse than the group with ASD, F (1, 38) = 5.70, $p < .05$, $\eta^2_{\text{partial}} = 0.13$, and the TD comparison group, F (1, 36) = 11.73, $p < .01$, $\eta^2_{\text{partial}} = 0.25$, although there was no significant difference between the ASD and TD comparison group, F (1, 44) = 1.01, $p = .32$, $\eta^2_{\text{partial}} = 0.02$.

Chunking acoustic findings. Overall, one-way ANOVAs revealed significant group differences in average f0, but not on other acoustic measures. For average f0, the group with LD had a significantly lower average f0 than the group with ASD, F (1, 38) = 5.23, $p < .05$, $\eta^2_{\text{partial}} = 0.12$, and the TD group, F (1, 36) = 7.68, $p < .01$, $\eta^2_{\text{partial}} = 0.18$. The ASD and TD groups did not differ on average f0, F (1, 44) = 0.01, $p = .91$, $\eta^2_{\text{partial}} = 0.001$.

We then analyzed responses on this subtest that were rated as correct by our trained examiners (correct chunking). For these responses, the only significant group difference was in average f0. There were no significant group differences on other acoustic measures. The LD group had a significantly lower average f0 than the groups with TD, F (1, 36) = 7.99, $p < .01$, $\eta^2_{\text{partial}} = 0.19$, and with ASD, F (1, 38) = 5.54, $p < .05$, $\eta^2_{\text{partial}} = 0.13$. The ASD and TD groups did not differ significantly in average f0 for correct responses, F (1, 44) = 0.04, $p = .84$, $\eta^2_{\text{partial}} = 0.001$.

Chunking summary. Overall, the groups performed similarly on the chunking input subtest, but the group with LD had greater difficulty than the other groups on the output subtest. An acoustic analysis of vocalizations on the output subtest revealed that the LD group had a lower average f0 than both groups, even when their responses were correct. There were no significant group differences on other acoustic variables.

Stress (PEPS-C focus subtests)

Focus behavioral findings. In this subtest, we examined participants' ability to understand sentence stress, and their ability to use it to highlight new information in a statement. We tested group performance on the PEPS-C focus input and output subtests using one-way ANOVAs. There were significant group differences on focus input, F (2, 59) = 4.02, $p < .05$, $\eta^2_{\text{partial}} = 0.12$, but not focus output, F (2, 59) = 0.83 $p = .43$, $\eta^2_{\text{partial}} = 0.03$. On the input task, the TD group scored significantly higher than the groups with ASD, F (1, 44) = 4.59, $p < .05$, $\eta^2_{\text{partial}} = 0.09$, and LD, F (1, 36) = 12.37, $p < .001$, $\eta^2_{\text{partial}} = 0.26$, but the groups with ASD and LD did not differ significantly from each other, F (1, 38) = 0.42, $p = .52$, $\eta^2_{\text{partial}} = 0.01$.

Focus acoustic findings. For focus output, one-way ANOVAs revealed significant group differences in average f0, f0 range, and *SD* of f0, but no differences in utterance intensity or duration. The group with LD had a significantly lower average f0 than the group with ASD, F (1, 38) = 7.04, $p < .01$, $\eta^2_{\text{partial}} = 0.16$, and TD group, F (1, 36) = 6.64, $p < .01$, $\eta^2_{partial} = 0.16$, but the ASD and TD groups did not differ, F (1, 44) = 1.57, $p = .22$, $\eta^2_{partial} = 0.03$. For the f0 range, the group with ASD had a significantly higher f0 range than the TD group, F (1, 44) = 9.07, $p < .01$, $\eta^2_{partial} = 0.17$, the LD group had a marginally higher f0 range than the TD group, F (1, 36) = 2.99, $p = .09$, $\eta^2_{partial} = 0.08$, but the groups with ASD and LD did not differ, F (1, 38) = 1.10, $p = .30$, $\eta^2_{partial} = 0.03$. For *SD* of f0 on the focus output subtest, the groups with ASD and LD had a significantly higher *SD* of f0 than the TD group, F (1, 44) = 5.95, $p < .05$, $\eta^2_{partial} = 0.12$, and F (1, 36) = 4.80, $p < .05$, $\eta^2_{partial} = 0.12$, respectively, but again the groups with ASD and LD did not differ significantly from each other, F (1, 38) = 0.10, $p = .75$, $\eta^2_{partial} = 0.003$.

We also analyzed focus output responses that were judged by raters to be correct (correct focus). For correct focus responses, there were significant group differences in average f0, f0 range, *SD* of f0, and utterance intensity, but the difference in utterance duration was not statistically significant. As with overall focus output findings, the group with LD had significantly lower average pitch than the group with ASD, F (1, 38) = 7.76, $p < .01$, $\eta^2_{partial} = 0.17$, and TD group, F (1, 36) = 9.26, $p < .01$, $\eta^2_{partial} = 0.21$, whereas the ASD and TD groups did not differ significantly, F (1, 44) = 1.27, $p = .27$, $\eta^2_{partial} = 0.03$. The findings for f0 range were also nearly identical. The group with ASD had a significantly higher f0 range than the TD group, F (1, 44) = 7.66, $p < .01$, $\eta^2_{partial} = 0.15$, the LD group had a marginally higher f0 range than the TD group, F (1, 36) = 3.17, $p = .08$, $\eta^2_{partial} = 0.08$, but the groups with ASD and LD did not differ, F (1, 38) = 0.58, $p = .45$, $\eta^2_{partial} = 0.02$. For *SD* of f0, the group with ASD had a significantly higher *SD* of f0 than the TD group, F (1, 44) = 6.54, $p < .01$, $\eta^2_{partial} = 0.13$, the LD group had a marginally higher f0 range than the TD group, F (1, 36) = 3.45, $p = .07$, $\eta^2_{partial} = 0.09$, but the groups with ASD and LD did not differ, F (1, 38) = 0.01, $p = .94$, $\eta^2_{partial} = 0.001$. For intensity, the group with ASD had a significantly higher intensity than the LD group, F (1, 38) = 11.00, $p < .001$, $\eta^2_{partial} = 0.23$, but the TD group did not differ from either the group with ASD, F (1, 44) = 0.56, $p = .46$, $\eta^2_{partial} = 0.01$, or the group with LD, F (1, 36) = 2.66, $p = .11$, $\eta^2_{partial} = 0.07$.

Focus summary. Overall, the groups with ASD and LD had more difficulty with the input subtest than the TD groups, but there were no differences in performance on the output subtest. On the acoustic measures, the groups with ASD and LD had significantly higher f0 ranges and *SD* of f0 than the TD group, even for correct responses. Similar to other tasks, the LD group had a lower average f0 than the other two groups. This was the only task that found group differences in utterance intensity. The group with ASD had higher utterance intensity than the group with LD.

DISCUSSION

The purpose of this study was to use acoustic measures in addition to raters' perception of behavioral responses to investigate prosodic performance in children with ASD in comparison to participants with TD and those with LD. Study stimuli were drawn from the PEPS-C, a measure of prosodic abilities that has been used in several studies to examine prosodic performance in ASD and related disorders. The PEPS-C provided structured samples of speech output in response to prosodic tasks involving several different functions of prosody, which were matched by analogous tasks measuring receptive abilities. Thus, production content was standardized across participants, which allowed for acoustic comparisons of multiple aspects of prosody on similar tokens for all participants. We predicted that there would be differences among diagnostic groups in both raters' perception of prosodic output in response to these tasks, as well as in the acoustic parameters of their prosodic production.

The findings on the input tasks are consistent with previous research that shows both children with ASD and those with language-based learning difficulties have difficulty accurately interpreting some (but not all) prosodic cues. Children with LD and ASD made more errors than TD children in interpreting prosody marking distinctions between questions and statements. Children with LD and ASD also shared the difficulties in interpreting affect and emphatic stress. These shared difficulties suggest a more pervasive difficulty in making sense of the suprasegmental layer of meaning in discourse. The findings from the LD group should be interpreted with caution, however, because our group with LD was heterogeneous. Although most had language-based deficits, there was a small minority who did not present with language-based disabilities, but had reading problems only.

The only significant behavioral difference in the output tasks was that children with LD were more likely to be misinterpreted on their production of sentence phrasing in the chunking task. These findings could be interpreted several ways. One interpretation might stem from the divergence in findings on input versus output items. That is, for children with both LD and ASD interpreting others' prosodic productions may require the integration of several levels of information processing (segmental, semantic, syntactic, suprasegmental), and processing prosody in conjunction with the processing of material at other levels that present challenges may be especially difficult. If this were the case, production may be relatively spared because much of the production to be generated was cued by the pictured or written stimuli on the task, leaving processing resources relatively freed up to focus on prosody. If this were the case, it would suggest that, perhaps, it is not prosody per se that presents difficulty, but rather the processing load inherent in the need to interpret multiple levels of linguistic input. This speculation could be tested by further research. These findings do, however, emphasize that other groups with communication disorders are also vulnerable to prosodic problems. Although prosodic deficits have long been associated with ASD, numerous studies (e.g., Catterall et al., 2006; Marshal et al., 2009; Stojanovik et al., 2007; Wells & Peppé, 2004) have shown them to be present in other populations with communication difficulties, and the present findings lend support to the notion that children with ASD show prosodic difficulties that are not entirely unlike those of other communicatively impaired children.

An additional aim of the present study was to expand the present understanding of findings based on listeners' perception of prosody by examining the acoustic properties of prosodic output in order to better characterize the prosodic production of children with ASD and LD. Our findings do reveal some acoustic differences between the three comparison groups. In general, the differences for children with ASD tended to be seen in longer durations, particularly on the two tasks that involved only single word productions (affect and turn-end). This finding echoes that of Paul et al. (2008), who showed that the timing of stressed versus unstressed syllables was less differentiated in speakers with ASD than in those with TD. Taken together, these findings suggest difficulties in the controlling precise temporal aspects of word production. Shriberg, Paul, Black, and van Santen (2011) have suggested that one explanation for speech difficulties in speakers with ASD is a difficulty in attunement to ambient conventions for community-acceptable production. That is, speakers with ASD have difficulty "tuning up" their productions to emulate models provided by other speakers, so that subtle differences that are detectable without necessarily affecting category boundaries, are present. Paul et al. (2008) speculated that these attunement difficulties may be attributable to a dearth of social motivation to "talk just like" other speakers in the community, which can result in a range of subtle differences, including failure to acquire community-appropriate accent or dialect (Baron-Cohen & Staunton, 1995) and persistence of distortions in speech sounds that are "outgrown" by typical speakers (Shriberg et al., 2001).

For children with LD, the major differences were seen in terms of average fundamental frequency, but not on all tasks, as would be the case if their voices were simply pitched lower. In addition, both ASD and LD children showed more variable pitch than those with TD when attempting to express emphatic stress (again replicating Paul et al., 2008), and the group with ASD showed significantly increased intensity in this context as well. These acoustic differences were present despite the fact that listeners perceived the children with ASD and LD to be producing acceptable prosody in all the tasks; the only output task to show significant deficit was the production of sentential phrasing on the chunking task by the LD group. Thus, despite difficulties in some aspects of prosody perception, these children are producing more or less acceptable prosodic tokens in this structured assessment, despite subtle differences in the fine details of the acoustic output.

Thus, a second potential explanation for these present findings showing more between group differences in production than in perception is that this IS a structured task. These high-functioning children were given a relatively large number of training items prior to each task. The administration of the test has practice items for which the correct prosody usage is modeled for the participant (Diehl & Paul, 2009). Second, the large number of trials, and the similar nature of trials within a subtest that are said one after another, limits the ecological validity of the utterances that are produced.

Therefore, participants' high level of performance on output tasks, despite some difficulties with perception, could be due to their learning the basic rules of each PEPS-C subtest and being able to implement them in this relatively simple task in which they interacted with a computerized interlocutor, rather than a person. Thus, as attractive as a standard task may be for studying prosodic ability, it may have less validity when compared to the kinds of varying, fast-paced, on-line production planning that goes on in spontaneous conversation. Of interest, previous acoustic studies found significant groups differences with more naturalistic tasks such as narratives (Diehl et al., 2009), but those using prosodic imitation tasks did not find similar differences (Paul et al., 2008).

Limitations and future directions

We were unable to examine incorrect responses because several participants in each group did not miss a single item on each of the subtests. This occurrence resulted in significantly reduced sample sizes in the incorrect conditions that limited our ability to detect very large effect sizes. Future studies should consider using younger children for their sample in order to reduce these ceiling effects.

Although the ASD and LD groups were not significantly different on nonverbal IQ, they were individually matched, and the group with LD, consistent with their diagnosis of language-based learning disability, also had lower CELF scores than the ASD group. Moreover, the group with LD was heterogeneous and included individuals with a range of profiles including both language-based learning disabilities and more specific reading deficits. Nonetheless, the fact that the group with LD performed worse (although not always significantly worse) on the behavioral measures of perception and production of prosody was surprising given prevalence of reports on prosodic deficits in ASD. Although this could be attributed to differences in general verbal ability, Wells and Peppé (2003) have argued from data on children with specific language impairments that prosodic ability appears to be relatively discrete from other levels of language functioning in this population, but could be a factor of the lower general language abilities of this group. For this study, we aimed to include a sample of children with high-functioning ASD that was representative of this population, with its generally high levels of formal language skill (Tager-Flusberg et al., 2005), rather than one that was specifically selected to match the LD group with its lower verbal abilities. It would be useful in future studies to match contrast groups more closely on verbal and nonverbal IQ in order to determine what prosodic deficits are present beyond general language deficits.

Another important question is that, even though we found acoustic differences, how can we tell whether or not the acoustic differences are within the acceptable range of human speech? One way to examine this question would be to collect subjective ratings of perceived differences in prosody production beyond its functional use. These data could be collected from trained professionals, naïve listeners, or both. Future studies should examine whether the subtle acoustic differences we have observed are related to subjective perception ratings of others.

Finally, the acoustic measures we used were broad and administered across the entire utterance, and did not look at specific functions in areas within the word or utterance. Green and Tobin (2009) suggested that individuals with ASD use a restricted range of prosodic contours, and McCann and colleagues (2007) found that all of their participants with ASD were impaired in one area of prosodic performance, but the area of impairment differed between individuals. Shriberg et al. (2001) reported that only half the adults with high-functioning autism they studied could be said to have any prosodic difficulties. Our measures were not sensitive to these types of individual differences.

Clinical implications

This study has several important clinical implications. First, acoustic analysis of prosody has the potential for providing automatic feedback for children who struggle with prosodic production (Kim, Newland, Paul, Scassellati, & Diehl, 2008). This study suggests that the acoustic differences in utterance duration, pitch range, and perhaps intensity are areas of particular difficulty for individuals with ASD. Several commercial software applications are available that provide visual feedback on these parameters, and these may prove useful in helping to normalize prosodic production in individuals with ASD. Second, the study supports findings of others that individuals with language-based learning disorders are also challenged in prosody performance. As more effective prosodic treatments emerge, LD as well as ASD populations may be considered as candidates for these treatments. Third, our findings emphasize, as earlier studies (e.g., Paul, Augustyn, et al., 2005) have suggested, that it is not only production of prosody that requires intervention. For whatever reason, the participants in this study in both clinical groups experience difficulty in selected aspects of both interpretation and production of prosody. As treatments are developed to address this area of communication disorder, they will need to be focused on both correct interpretation of prosodic input as well as acceptable production. Fourth, prosodic interventions should consider factors beyond the correct linguistic application of prosody and also focus on differences in stylistic presentation of speech (pitch range, intensity, etc.) that go beyond the content of the communication. Fifth and finally, prosodic interventions may benefit from an understanding not only of the degree to which speakers with ASD get prosody "right" or "wrong," but also of the ways in which their production differ subtly from that of typical speakers even when they are interpreted correctly. This understanding can help to inform interventions aimed at decreasing the perceived "oddity" so often reported for speakers with ASD.

ACKNOWLEDGMENTS

This paper was supported by the NIDCD Grant K24 HD045576 (to R.P.), NICHD Grant P01-HD03008 (Project 3), and the James Hudson Brown and Alexander Brown–Coxe Postdoctoral Fellowship through the Yale School of Medicine. We thank the children and families who made this work possible. We also thank Lauren Berkovits, Daria Diakonova, Casey Dolezal, Kate Elliott, Tracey Gemmell, Allison Lee, Joshua Noffsinger, Beck Roan, Lauren Schmitt, Elizabeth Schoen, Nicole Shea, Kristin Uhland, and Megan Van Ness for their contributions to this project, which included participant recruitment, data collection, and data management.

NOTES

1. The use of the term *high functioning* is generally used to refer to individuals with autism who are in the average to above average range of cognitive functioning, although some studies have used general language measures to make this distinction. As such, there is no accepted or recognized definition of "high functioning."

2. Note that one LD participant missed every *Like* response and therefore had no *Like Correct* data.
3. Note that one TD participant liked every food, and one TD participant did not get a single *Dislike* response correct. Therefore, this participant has no data for those categories.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, text revision* (4th ed.). Washington, DC: Author.

Baltaxe, C. A. M. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27*, 97–105.

Baron-Cohen, S., & Staunton, R. (1995). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language, 14*, 241–248.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (version 5.1.20) [Computer program]. Retrieved November 11, 2009, from http://www.praat.org/

Catterall, C., Howard, S., Stojanovik, V., Szczerbinski, M., & Wells, B. (2006). Investigating prosodic ability in Williams syndrome. *Clinical Linguistics and Phonetics, 20*, 531–538.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2008). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with asperger syndrome. *Research in Autism Spectrum Disorders, 3*, 502–516.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement, 33*, 107–112.

Diehl, J. J., & Berkovits, L. (2010). Prosody as a diagnostic and cognitive bellwether of autism spectrum disorders. In A. E. Harrison (Ed.), *Speech disorders: Causes, treatment and social effects*. Hauppage, NY: Nova Science Publishing.

Diehl, J. J., & Paul, R. (2009). The assessment of prosodic disorders and neurological theories of prosody. *International Journal of Speech–Language Pathology, 11*, 287–292.

Diehl, J. J., Watson, D. G., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics, 30*, 385–404.

Elliott, C. D. (1990). *Differential ability scales*. San Diego, CA: Harcourt Brace Jovanovich.

Fosnot, S. M., & Jun, S. (1999). *Prosodic characteristics in children with stuttering or autism during reading and imitation.* Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult . . . but worth it: A study in high functioning autism. *International Journal of Speech–Language Pathology, 11*, 308–315.

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 33*, 3–13.

Järvinen-Pasley, A., Paisley, J., & Heaton, P. (2008). Is the linguistic content of speech more salient than its perceptual features in autism? *Journal of Autism and Developmental Disorders, 38*, 239–248.

Kim, D., Newland, E., Paul, R., Scassellati, B., & Diehl, J. J. (2008). *A robotic therapist for affective prosody in high-functioning autistic children.* Paper presented at the International Meeting for Autism Research, London.

Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and *DSM-IV* criteria for autistic disorder: Results of *DSM-IV* autism field trial. *Journal of Autism and Developmental Disorders, 30*, 163–167.

Kogan, M. D., Blumberg, S. J., Schieve, L. A., Boyle, C. A., Perrin, J. M., Ghandour, R. M., et al. (2009). Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics, 124*, 1395–1403.

Loca, J., & Wootton, T. (1995). Interactional and phonetic aspects of immediate echolalia in autism—A case study. *Clinical Linguistics & Phonetics, 9*, 155–184.

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

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. J. K. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language and Communication Disorders, 44*, 466–488.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders, 38*, 325–350.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. D. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *Journal of Language and Communication Disorders, 42*, 682–702.

Miller, J. N., & Ozonoff, S. (2000). The external validity of Asperger disorder: Lack of evidence from the domain of neuropsychology. *Journal of Abnormal Psychology, 109*, 227–238.

Paccia, J. M., & Curcio, F. (1982). Language processing and forms of immediate echolalia in autistic children. *Journal of Speech and Language Research, 25*, 42–47.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*, 205–220.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders, 2*, 110–124.

Paul, R., Orlovski, S., Marcinko, H. C., & Volkmar, F. R. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*, 115–125.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*, 861–869.

Peppé, S. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech–Language Pathology, 11*, 258–271.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics, 17*, 345–354.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 50*, 1015–1028.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism diagnostic interview—Revised*. Los Angeles: Western Psychological Services.

Schoen, E., Paul, R., & Chawarska, K. (2010). Vocal development in toddlers with ASD. In R. Paul & P. Flipsen (Eds.), *Child speech sound disorders: In honor of Lawrence D. Shriberg*. San Diego, CA: Plural Publishers.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: Psychological Corporation.

Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P. H. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*, 405–426.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and asperger syndrome.*Journal of Speech, Language, and Hearing Research, 44*, 1097–1115.

Stojanovik, V., Setter, J., & van Ewijk, L. (2007). Intonation abilities of children with Williams syndrome: A preliminary investigation. *Journal of Speech, Language, and Hearing Research, 50*, 1606–1617.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior* (3rd ed., pp. 335–364). Hoboken, NJ: Wiley.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45*, 135–170.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: Pearson.

Wells, B., & Peppé, S. (2003). Intonation abilities in children with speech and language impairments. *Journal of Speech, Language, and Hearing Research, 46*, 5–20.

Wells, B., & Peppé, S. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*, 749–778.

Werner, E., Dawson, G., & Osterling, J. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders, 30*, 157–162.

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34*, 473–493.

Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, Speech, and Hearing Services in Schools, 36*, 62–72.