JSLHR

Review Article

Word Processing in Children With Autism Spectrum Disorders: Evidence From Event-Related Potentials

Micheal Sandbank,[a] Paul Yoder,[b] and Alexandra P. Key[c]

Purpose: This investigation was conducted to determine whether young children with autism spectrum disorders exhibited a canonical neural response to word stimuli and whether putative event-related potential (ERP) measures of word processing were correlated with a concurrent measure of receptive language. Additional exploratory analyses were used to examine whether the magnitude of the association between ERP measures of word processing and receptive language varied as a function of the number of word stimuli the participants reportedly understood.
Method: Auditory ERPs were recorded in response to spoken words and nonwords presented with equal probability in 34 children aged 2–5 years with a diagnosis of autism spectrum disorder who were in the early stages of language acquisition. Average amplitudes and amplitude differences between word and nonword stimuli within 200–500 ms were examined at left temporal (T3) and parietal (P3) electrode clusters. Receptive vocabulary size and the number of experimental stimuli understood were concurrently measured using the MacArthur–Bates Communicative Development Inventories.
Results: Across the entire participant group, word–nonword amplitude differences were diminished. The average word–nonword amplitude difference at T3 was related to receptive vocabulary only if 5 or more word stimuli were understood.
Conclusions: If ERPs are to ever have clinical utility, their construct validity must be established by investigations that confirm their associations with predictably related constructs. These results contribute to accruing evidence, suggesting that a valid measure of auditory word processing can be derived from the left temporal response to words and nonwords. In addition, this measure can be useful even for participants who do not reportedly understand all of the words presented as experimental stimuli, though it will be important for researchers to track familiarity with word stimuli in future investigations.
Supplemental Material: https://doi.org/10.23641/asha.5614840

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in communication and social function (American Psychiatric Association, 2013). Though no longer a core symptom of the disorder, impairments in language are common in children with ASD (Anderson et al., 2007; Hus, Pickles, Cook, Risi, & Lord, 2007; Pickett, Pullara, O'Grady, & Gordon, 2009), with language comprehension being relatively more impaired than language production (Charman, Drew, Baird, & Baird, 2003; Hudry et al., 2010; Tager-Flusberg, 2000). In fact, early substantial deficits in language comprehension relative to production may serve as an early sign of ASD (Barbaro & Dissanayake, 2012), which may help to differentiate young children with ASD from late talkers with similar expressive abilities (Paul, Chawarska, Cicchetti, & Volkmar, 2008) and children with nonspectrum developmental delay (Weismer, Lord, & Esler, 2010). This is important because language comprehension is likely integral to the development of social understanding. Longitudinal studies of individuals with ASD suggest that early language comprehension scores are strongly predictive of social functioning in adulthood (Rutter, Mawhood, & Howlin, 1992). Moreover, receptive language ability may govern the extent to which a child with ASD can benefit from a naturalistic developmental language intervention (Paul, Campbell, Gilbert, & Tsiouri, 2013). Thus, a greater understanding of the underlying processes of language comprehension development in children with ASD is critical.

[a]Department of Special Education, University of Texas at Austin
[b]Special Education Department, Vanderbilt University, Nashville, TN
[c]Vanderbilt Kennedy Center for Research on Human Development and Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN

Correspondence to Micheal Sandbank: msandbank@austin.utexas.edu

Editor-in-Chief: Julie Liss
Editor: Bharath Chandrasekaran

Received January 9, 2017
Revision received May 1, 2017
Accepted July 10, 2017
https://doi.org/10.1044/2017_JSLHR-S-17-0011

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

The Potential Utility of Event-Related Potentials for Explaining Individual Differences in Receptive Language

Researchers have called for the increased use of novel technologies, such as event-related potentials (ERPs), to address questions surrounding the development of language comprehension and related skills in young children with ASD (Tager-Flusberg & Kasari, 2013). ERP is a useful technology for measuring language processing in young children with ASD for a variety of reasons. First, the high temporal resolution of ERP measurement procedures allows investigators to measure cognitive processes as they happen at the millisecond level. Second, passive ERP measurement procedures do not require task comprehension or motivation to cooperate with the instructions. ERPs also allow more movement than some alternative neural measures (e.g., functional magnetic resonance imaging). Thus, passive ERP measurement procedures can be used to examine the early development of language processing, particularly in populations that may have difficulty with other types of neuroimaging, such as very young children and individuals with disabilities (Molfese & Molfese, 1985; Molfese, Molfese, & Espy, 1999; Tsao, Liu, & Kuhl, 2004).

In the effort to better understand the development of language comprehension, researchers have examined toddlers' neural responses to the aural presentation of known words and contrast stimuli (e.g., nonwords, unknown words, or backward words; Mills, Coffey-Corina, & Neville, 1993, 1994, 1997; Mills et al., 2004; Mills, Plunkett, Prat, & Schafer, 2005; Molfese, 1990). The goal of this experimental paradigm is to elicit and isolate neural activity associated with word processing, a prerequisite skill to language comprehension. Previous ERP studies of word processing in typically developing children have shown differential patterns of neural activity emerging in response to known words compared to contrast stimuli as early as 13 months of age, with known words eliciting more negative amplitudes than contrast stimuli, between 200 and 500 ms after stimulus onset (Mills et al., 1997, 2004). In young children (13–17 months), this response was bilateral and broadly distributed across anterior and posterior regions (Mills et al., 1997, 2004). In older children (20 months), this response was limited to left temporal and parietal regions (Mills et al., 1993, 2004). Subsequent studies of word processing, which compared younger and older late-talking toddlers, typically developing toddlers with high and low productive and understood vocabularies, and 20-month-olds before and after vocabulary training, have demonstrated that the apparent left-hemisphere specialization for word processing occurs as a function of language experience, rather than age (Mills et al., 1997; Mills, Conboy, & Paton, 2005; Mills, Plunkett, et al., 2005). That is, as children initially familiarize themselves with new vocabulary, they exhibit a broad and bilateral neural response to known words, marked by negative amplitudes 200–500 ms after stimulus onset. With development and increased language proficiency, word processing activity shifts to the left hemisphere and is focused in the temporal and parietal regions (Mills et al., 2004).

Although documentation of the timing and location of differential neural responses to certain stimuli within specific populations can be useful for refining theories about cognitive processes, more is needed for electrophysiological evidence to have value for clinical populations. One way to improve the basis for inferring that ERP responses to word versus contrast stimuli reflects word processing is to examine the association of the ERP measure with variables thought to be associated with word processing (Cronbach & Meehl, 1955). This type of validity is called "nomological" validity, and it is useful when there is no gold standard measure of a construct (Yoder & Symons, 2010). The variables that are predicted to correlate with the ERP measure are called "nodes" on the nomological net. For ERP measures used in clinical practice, among the most useful nodes are those that are measured outside the ERP procedure, because both measures would have to reflect trait-like characteristics to be associated, instead of simply being reflections of within-procedure states. In addition, correlations between ERP measures of word processing and nodes on the nomological net provide a more precise test of the theory that generated the prediction than between–intact groups mean differences, because such correlations require quantification of word processing at the individual level, not just the group mean level. Such associations are also most relevant for testing whether ERP measures of word processing might have clinical value because clinical decisions are made on an individual, not group, level.

Prior Findings for ERPs as a Predictor of Receptive Language in Children With ASD

Only one ERP investigation has examined word processing in young children with ASD and tested the nomological validity of the ERP measure outside the ERP procedure. Kuhl and colleagues (2013) compared the ERP response to known versus unknown word stimuli in 2-year-olds with ASD and age-matched typical participants. Results suggested that brain responses of children with ASD did differentiate between known and unknown words. However, only children with more adaptive scores on a measure of social functioning exhibited word processing activation patterns that were focused and left-lateralized, similar to the typically developing group. In contrast, children with ASD with less adaptive social scores exhibited word processing activation patterns that were diffuse and right-lateralized. Furthermore, average amplitudes to word stimuli at a left parietal (P3) location strongly and significantly predicted receptive language in children with ASD at 2 and 4 years later ($r = -.671$ and $r = -.785$, respectively). Thus, neural markers of word processing have strong potential to be a useful predictor of language outcomes in children with ASD.

Need for Replication and Extension of Prior Work

The derivation of an ERP measure of word processing that has a strong association with receptive language could enhance our ability to identify the subgroup of preverbal children with ASD who show evidence of poor word processing through atypical ERP responses and who may be at risk for experiencing poor speech outcomes in the long term. Although the findings of Kuhl and colleagues (2013) suggest that ERP measures show great promise for predicting outcomes in clinical populations, further evidence is needed for these measures to have clinical utility. Significant associations between behavioral outcomes and ERP measures that were derived using the timing and locations at which between-conditions significant differences were observed within a sample are likely to be sample specific. In order to advance the science, researchers should make a priori determinations about measure derivation based on prior evidence and document theoretically predictable associations for those measures.

Purpose and Research Questions

In the current investigation, which featured young language-learning children with ASD, we sought to extend the utility of previous ERP findings by examining the validity of word processing measures derived using the timing and locations documented in prior ERP investigations of this construct. To this end, we addressed two primary research questions. First, we examined whether young children with ASD differentiated words from nonwords. On the basis of the findings from prior studies, a more negative left temporal or parietal response to words than nonwords within 200–500 ms would indicate typical and proficient word processing (Kuhl et al., 2013; Mills, Conboy, et al., 2005; Mills et al., 1997; Mills, Plunkett, et al., 2005). Second, we evaluated whether neural indices of word processing were associated with concurrent parent report measures of receptive vocabulary. Greater negative amplitudes to word or greater differences between average amplitude responses to words and nonwords at left temporal and parietal regions should be positively associated with higher scores of concurrent receptive vocabulary. Finally, in a post hoc exploratory analysis, we sought to determine whether the magnitude of the associations between word processing and receptive vocabulary varied based on whether participants reportedly understood the words featured in the experimental stimuli list. Many of the previous ERP word processing investigations presented a varied word stimuli list based on parent-reported child understanding. In our investigation, a static list of word stimuli was presented to all participants so that variability in ERPs could not be attributable to stimulus differences, but word understanding for each word on the list was concurrently tracked using the MacArthur–Bates Communicative Development Inventories (MCDI; Fenson et al., 2007). If participants varied in their understanding of the words on the stimuli list and if the ERP measures derived based on extant literature truly indexed word processing, then the association between these measures and concurrent scores of receptive vocabulary should vary based on the number of word stimuli the child understood and that association would be strongest for children who reportedly understood all the word stimuli and weakest for children who understood none of the word stimuli.

Method

Participants

Thirty-four children with ASD (27 boys, seven girls), of ages 2–5 years (*M* age = 45.40 months, *SD* = 9.63), participated in the study. All of them were enrolled in a larger, longitudinal study of language development in young children with ASD (Yoder, Watson, & Lambert, 2015). At the outset of the larger study, all participants were “preverbal,” meaning that they used five words or fewer during a 15-min language sample and were reported by their parents as using fewer than 20 words total using the MCDI. Because funding for the ERP procedure was acquired 2 years after the onset of the larger study, ERPs were collected at various measurement periods for each participant in the larger study’s design. Thus, by the time of ERP data collection, 17 participants had begun to use more than 20 words and no longer met the criteria to be described as “preverbal.” The receptive and expressive vocabulary scores collected concurrently with the ERP are characterized in Table 1. Ten additional participants were consented but excluded due to excessive movement noise or lack of cooperation with ERP testing procedures. Children who entered the study with a previous diagnosis of ASD had their diagnosis confirmed using the revised diagnostic algorithm for the Autism Diagnostic Observation Schedule Module I (ADOS; Gotham, Risi, Pickles, & Lord, 2007; Lord et al., 2000) by research staff who were research reliable on this instrument. Children who entered the study without a previous diagnosis were diagnosed with ASD by a licensed clinician on the research team who had experience evaluating young children with ASD and who was research reliable on the ADOS. Diagnoses were provided based on the clinician’s judgment that the child met the criteria for autism or pervasive developmental disorder not otherwise specified outlined in the Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition–Text Revision (American Psychiatric Association, 2000), which was confirmed by the ADOS clinical interview. Children with comorbid sensory motor impairments, metabolic or progressive neurological disorders, or genetic syndromes were excluded. This information was obtained through parent report. Although no intervention was provided as a part of the study, parents reported that their children received an average of 4.85 hr of speech therapy per month and an average of 12.46 hr of all other therapies (e.g., occupational therapy, physical therapy). According to parent report, none of the participants were being treated with medications commonly prescribed to children with ASD. Other descriptive participant information is presented in Table 1.

Table 1. Demographic Information for Participants.

Variable	*n*	*M*	*SD*	Min	Max
Gender					
Boys	27				
Girls	7				
Race					
African American	9				
White	25				
Chronological age in months[a]		45.40	9.63	25.53	62.29
Level of education of primary caregiver[b,c]		6	1.57	1	9
ADOS diagnostic score[b]		23.53[d]	3.62	16	28
Mullen[b]					
Standard score		51	4.94	49	68
Age equivalency in months[e]		11.32	5.10	3.75	26.50
Expressive language age in months		7.69	3.72	2	21
Receptive language age in months		5.82	7.52	1	30
MCDI[a]					
Raw Expressive Vocabulary		44.76	74.25	0	254
Raw Receptive Vocabulary		139.90	123.27	3	396

Note. ADOS = Autism Diagnostic Observation Schedule; Mullen = Mullen Scales of Early Learning; MCDI = MacArthur-Bates Communicative Development Inventories.

[a]At time of ERP. [b]At entry in the larger longitudinal study. [c]Taken from a lab-specific survey of caregiver education. A score of 6 represents some years of college. [d]ADOS Module 1 Diagnostic Cut-off score is 12. [e]Age equivalency in months is the average age equivalency across Fine Motor, Visual Reception, Expressive Language, and Receptive Language subscales of the Mullen.

Ethics Statement

Ethical approval for this study was obtained from the institutional review board of Vanderbilt University. Written informed consent was obtained from parents or legal guardians for all participants, and evidence of assent (e.g., child behaviors indicating willingness to participate) was also obtained for all participants. The rights of participants were protected according to the principles explained in the Declaration of Helsinki.

Electroencephalogram (EEG) Data Acquisition Procedures

Stimuli

A set of 10 English words that are typically among the first learned by young children and 10 pronounceable nonwords, which were matched to words on duration and number of syllables, were used as the stimuli. The lists of stimuli were identical to the ones used in a prior word processing investigation (Mills et al., 2004), with the exception of two words that were replaced to better match the age of typical acquisition. Two nonwords that sounded like real words were also changed (see Table 2). All stimuli were recorded by a young, female, native English speaker and had prosodic features consistent with those reported in previous studies of English child-directed speech (Cooper & Aslin, 1990; Fernald et al., 1989). Stimuli will be made available for replication studies on request.

EEG Acquisition

EEG data were collected using a 128-channel Hydrocel net (EGI, Inc.), 250 Hz sampling rate, 0.1–100 Hz filters, and Cz reference. Impedances were adjusted to < 40 kΩ just before data acquisition. Words and nonwords were presented at 75 dB SPL in random order, with equal probability (three times each, 60 trials total), with a varied intertrial interval of 1500–2500 ms, which prevented habituation to stimulus onset. The entire session lasted approximately 10 min. No behavioral responses were required. To facilitate cooperation during data acquisition, an age-appropriate video with muted sound was shown to participants. Previous work has shown that muted videos presented in combination with auditory stimuli do not inhibit the quality of auditory ERP data in children (Mahajan & McArthur, 2011).

EEG Data Processing

During offline data processing, signals were smoothed using a 30-Hz low-pass filter. Each EEG record was divided

Table 2. Word and nonword stimuli presented during electroencephalogram data acquisition.

Word	Nonword
ball	kobe
car[a]	lif
book	neem
bottle	fipe
cup	mon
drink[a]	neps[a]
dog	riss
milk	towd[a]
nose	jud
shoe	zav

[a]Stimuli that have been changed from the original list used by Mills and colleagues (2004), replaced to better match age of acquisition (words) or to reduce possible familiarity of nonwords.

into 700-ms sections time-locked to stimulus onset. The first 100 ms reflected baseline (prestimulus) neural activity (measured in microvolts), and the subsequent 600 ms reflected poststimulus potentials. Single trial data were screened for "bad" channels (voltage shift in excess of 150 µV) and ocular artifacts. Trials with 15 or more bad channels, eyeblinks, or movements were rejected. For the remaining trials, data for electrodes characterized by consistently high noise levels were replaced using the spherical spline interpolation algorithm (Perrin, Pernier, Bertrand, & Echallier, 1989). Next, data were averaged across trials for each stimulus condition, re-referenced to an average reference, and baseline-corrected. A minimum of 10 artifact-free trials per stimulus condition was required for participant inclusion in statistical analyses. Across our sample, the average numbers of trials obtained were 18.6 ($SD = 6.4$) for the word condition and 18.12 ($SD = 6.6$) for the nonword condition.

Next, mean amplitudes in response to word and nonword stimuli were derived for left temporal (T3) and left parietal (P3) locations within 200–500 ms after stimulus onset. These scalp locations and temporal range were selected on the basis of the past research on typically developing children (Mills et al., 1993, 1994, 1997; Mills, Plunkett, et al., 2005) and children with ASD (Kuhl et al., 2013). To capitalize on the rich data set offered by a high-density array, we used a priori determined electrode clusters rather than single electrodes corresponding to the locations of interest. Five spatially adjacent electrodes comprised the T3 electrode cluster, and six spatially adjacent electrodes comprised the P3 electrode cluster (see Figure 1). Data within each cluster were averaged across the individual electrodes. The average of highly correlated values from multiple spatially proximal electrodes was expected to be more stable than values from a single electrode and, therefore, thought to reduce the likelihood of Type II error. Using guidance from theory and previous literature, word processing was quantified in two ways for each participant: (a) as the mean amplitude to the word condition and (b) as the within-participant amplitude difference between the word and nonword conditions. Therefore, the statistical analyses were conducted on the average amplitudes to word and the averaged difference values (word–nonword) across the prespecified time window (200–500 ms) at T3 and P3 clusters. Thus, a total of four putative ERP measures of word processing were derived for each participant.

Behavioral Measures

Descriptive Measures

The Mullen Scales of Early Learning (MSEL; Mullen, 1995) was administered to our sample at entry into the larger study (Yoder et al., 2015) and used to describe the developmental and language levels of our participants, though the time between the administration of this assessment and the ERP procedure varied across participants. The MSEL is used to assess early cognitive development across four scales (Visual Reception, Fine-Motor, Receptive Language, and Expressive Language) in children ages 0–5 years. Strong convergent and criterion-related validity has been reported for scores provided by the MSEL for children with ASD (Bishop, Guthrie, Coffing, & Lord, 2011; Farmer, Golden, & Thurm, 2016; Nordahl-Hansen, Kaale, & Ulvund, 2014; Swineford, Guthrie, & Thurm, 2015). Descriptive statistics for the MSEL standard score, mental age equivalency score, and receptive and expressive language age equivalency scores are presented for our sample in Table 1.

The Autism Diagnostic Observation Schedule–Generic Module 1 (Lord et al., 2000) was used to verify ASD diagnoses of participants at entry to the larger study (Yoder et al., 2015). The ADOS is considered a "gold standard" assessment for ASD diagnostic purposes (De Bildt et al., 2004). The diagnostic reliability and validity of the ADOS are well established (Gotham et al., 2007). The calibrated severity scores from the ADOS (Gotham, Pickles, & Lord, 2009) were used to describe our sample in Table 1.

Receptive Vocabulary

The MCDI: Words and Gestures form (Fenson et al., 2007) was administered concurrently with ERP data collection to index receptive vocabulary in all participants. This form is a checklist of 396 words commonly understood by young children between the ages of 8 and 16 months, which parents can use to indicate which words their child "says and understands" or "understands only." The raw scores were summed across the two response categories to comprise "total words understood" for each participant. The raw score is the recommended metric for developmentally delayed children older than 16 months and is sufficient to test the association of interest.

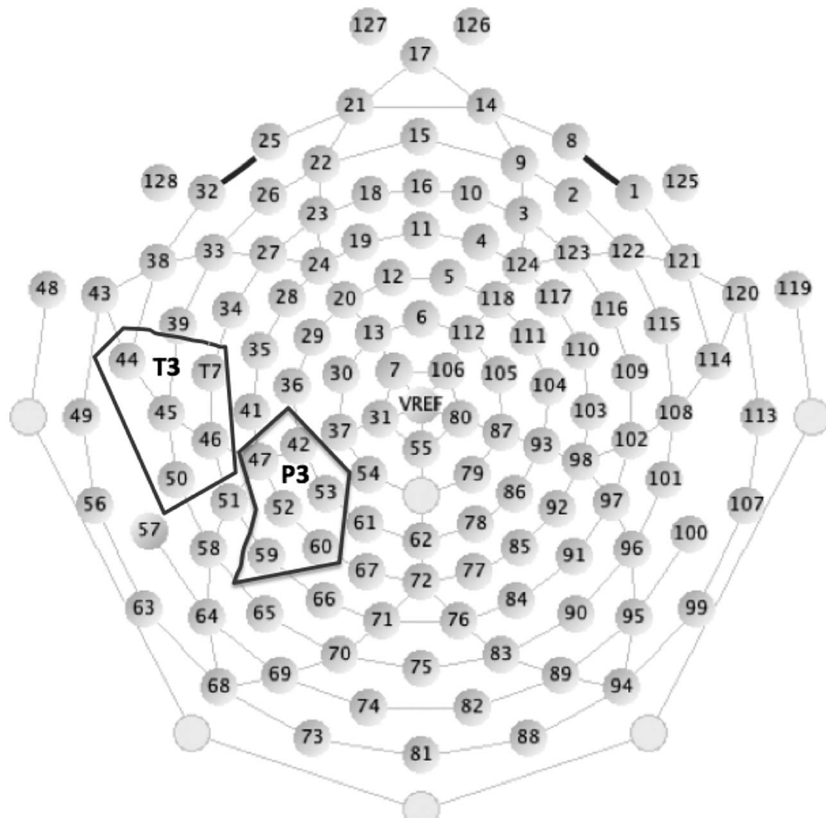
Word Stimuli Understood

Familiarity with words featured in the ERP experimental stimuli was not required for inclusion in this study. As such, participants' familiarity with word stimuli was not assessed during the ERP procedure. However, because all of the words featured in the experimental stimuli are included on the MCDI: Words and Gestures form and because this assessment was administered concurrently with ERP data collection, a detailed record of each participant's understanding of the word stimuli was kept. Completed MCDI forms were hand-checked by a research assistant to document word familiarity for each participant. Word stimuli marked in either the "understands" or "understands and says" columns were counted as understood. Word stimuli that were neither marked in the "understands" nor "understands and says" column were counted as not understood. The total number of word stimuli counted as understood was used as the metric for this construct in post hoc analyses.

Data Analysis Plan

Discrimination between words and nonwords was analyzed using planned comparisons of grand-averaged data for each condition at the time window (200–500 ms poststimulus onset) and scalp locations (T3 and P3) identified in previous literature. On the basis of the expectation

Figure 1. Map of the electrode layout of the EGI net used to acquire electroencephalogram data. Electrode clusters corresponding to left temporal (T3) and left parietal (P3) are indicated. VREF = reference voltage.



of increased negative amplitudes in response to word compared to nonword stimuli, the significance of the condition differences at T3 and P3 scalp locations was tested against zero using paired one-sided *t* tests. To determine whether age needed to be controlled when testing between-conditions differences on ERP variables, we examined the associations between age and the average amplitudes to word and average word–nonword differences at T3 and P3 using Pearson correlation coefficients.

Brain–behavior associations between word processing and receptive vocabulary were documented using Pearson correlation coefficients for receptive vocabulary and each of four putative ERP measures (the mean amplitude to word and the word–nonword amplitude difference at T3 and P3). On the basis of the directional prediction of better receptive vocabulary being associated with more negative values to word and greater differences between the conditions, the significance of these associations was tested using one-tailed significance tests. Previous ERP investigations have documented higher response strength (more negative values) to the word condition than the contrast condition. Thus, greater negative amplitudes and negative difference scores (derived by subtracting the nonword condition from the word condition) were thought to reflect more adaptive word processing.

Results

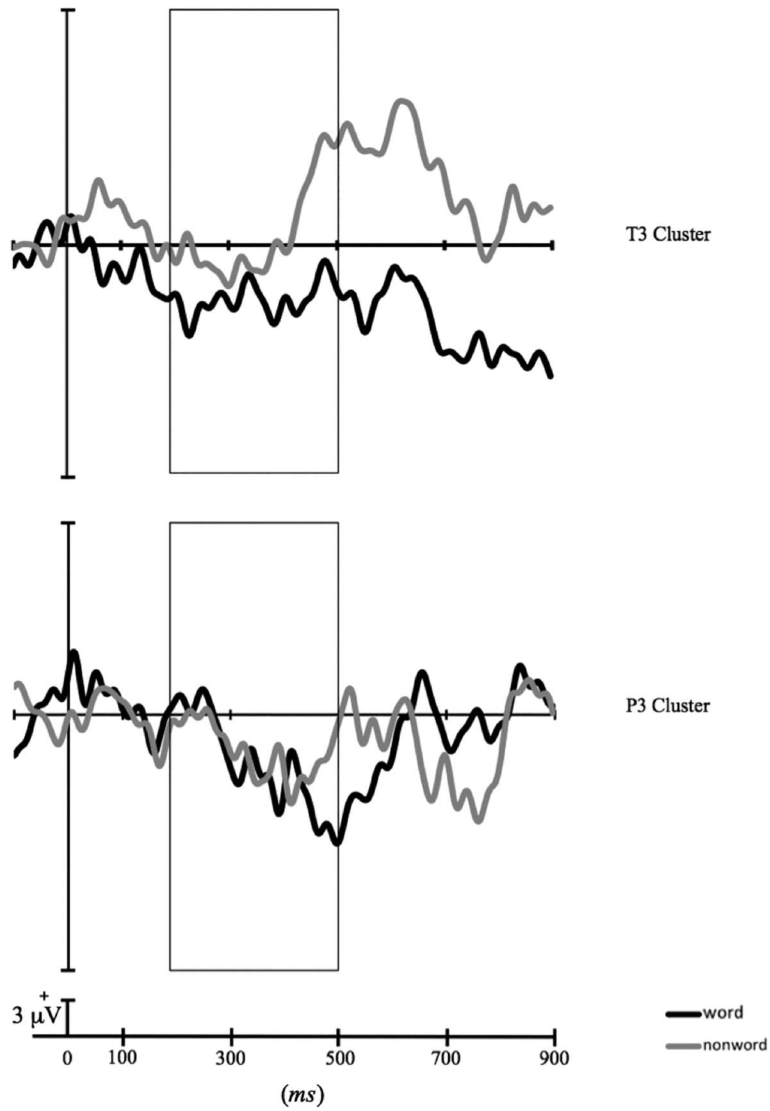
Preliminary Analyses

All analyses were conducted using the statistical program R (R Core Team, 2013), with the package erp.easy (Moore, 2016). To verify that data satisfied the assumptions of the linear model, we examined plots of residuals and standardized residuals against fitted and leveraged values, as well as normal QQ plots for each linear model. The data satisfied assumptions of multivariate normality and homoscedasticity. Consequently, no transformations were made. Potentially influential data points were handled with sensitivity analyses, which verified that their removal did not substantially change results.

RQ1. Word–Nonword Differentiation

Grand-averaged waves at T3 and P3 electrode clusters for the entire sample are presented in Figure 2. Visual inspection confirmed that, for both electrode clusters, mean amplitudes to word stimuli between 200 and 500 ms were more negative than mean amplitudes to nonword stimuli, with the average word–nonword difference larger at the T3 electrode cluster ($M = -0.89$, $SD = 4.63$) than at the P3 cluster ($M = -0.28$, $SD = 2.95$). However, neither the average

Figure 2. Grand-averaged waveforms for word and nonword conditions, averaged across the left temporal (T3) and left parietal (P3) electrode clusters. The temporal window of the word processing response documented in previous literature (200–500 ms after stimulus onset) is highlighted.



word–nonword difference at T3, $t(33) = -1.11$, $p = .13$, $d = -0.231$, nor that at P3, $t(33) = -0.56$, $p = .29$, $d = -0.09$, were significantly different from zero. Age in months was not significantly correlated with the average amplitude to word at T3 ($r = -.12$, $p = .74$) or P3 ($r = -.08$, $p = .32$) nor the size of the between-conditions mean amplitude difference at T3 ($r = .04$, $p = .83$) or P3 ($r = -.07$, $p = .67$).

RQ2. Unconditional Associations Between ERP Scores of Word Processing and Receptive Vocabulary

There were no significant concurrent correlations between any of the four ERP measures we examined and receptive vocabulary scores, either for T3 average amplitude to word ($r = -.20$, $p = .13$), P3 average amplitude to word ($r = -.21$, $p = .11$), T3 average word–nonword difference ($r = -.17$, $p = .16$), or P3 average word–nonword difference ($r = -.04$, $p = .39$). However, the direction of the associations between putative word processing measures and receptive vocabulary was consistent with our expectations in every case.

Post Hoc Analyses

A series of exploratory post hoc analyses was conducted to identify potential explanations for our failure to confirm our a priori hypotheses related to our primary research questions. First, we verified that word–nonword amplitude differences at T3 and P3 were not significantly associated with autism symptom severity, developmental level, or the number of trials retained for each participant. We also verified that between-conditions differences were not significant at other electrode sites of interest (i.e., right hemisphere homologues and frontal and occipital sites). Further information on these analyses is reported in Supplemental Material S1.

Because the word stimuli list was purposefully kept static in order to minimize variance in ERP responses that could be attributable to stimuli differences between participants and because familiarity with word stimuli was not required for inclusion in this study, we investigated the extent to which participant word stimuli knowledge may have accounted for our failure to document significant word–nonword differences and associations with receptive vocabulary. To this end, we examined the group-level data on participant familiarity with word stimuli and then tested whether the between-conditions differences on ERP measures might vary by the number of stimulus words understood. Because between-conditions differences on ERP measures might be detected only in children who understand all of the stimulus words, the significance of the condition differences at T3 and P3 scalp locations against zero was tested exclusively within participants who reportedly knew all 10 word stimuli, using paired one-sided *t* tests.

We then examined whether the association between ERP measures of word processing and concurrent receptive language was conditional on participant familiarity with word stimuli. A linear regression model was used to test the statistical interaction between each putative ERP measure of word processing (average word–nonword difference at T3 and P3 and average amplitude to word at T3 and P3) and the total number of word stimuli understood by the participant to predict concurrent receptive vocabulary. Significant interactions were followed up with the Johnson–Neyman test to identify the minimum number of understood word stimuli for which the ERP response was significantly associated with receptive vocabulary (Aiken, West, & Reno, 1991). The effect of each predictor in each linear model was quantified by the partial eta-squared effect size (η_p^2), which represents the proportion of variance in a dependent variable associated with a given predictor, with variance from other predictors and interactions partialled out (Richardson, 2011). Values of .01, .06, and .14 have been suggested for this effect size as benchmarks indicating small, moderate, and large effects (Cohen, 1988).

Participant Understanding of Word Stimuli

Participants varied widely in their reported understanding of the words featured in the ERP paradigm. The median number of word stimuli understood was 7, and the mean was 6.4 (*SD* = 3.56). Nine participants reportedly knew all 10 word stimuli. One participant reportedly knew none of them. No single word was understood by all participants. The word stimulus understood by the most participants (n = 28) was "ball." The word stimulus understood by the fewest participants (n = 13) was "bottle."

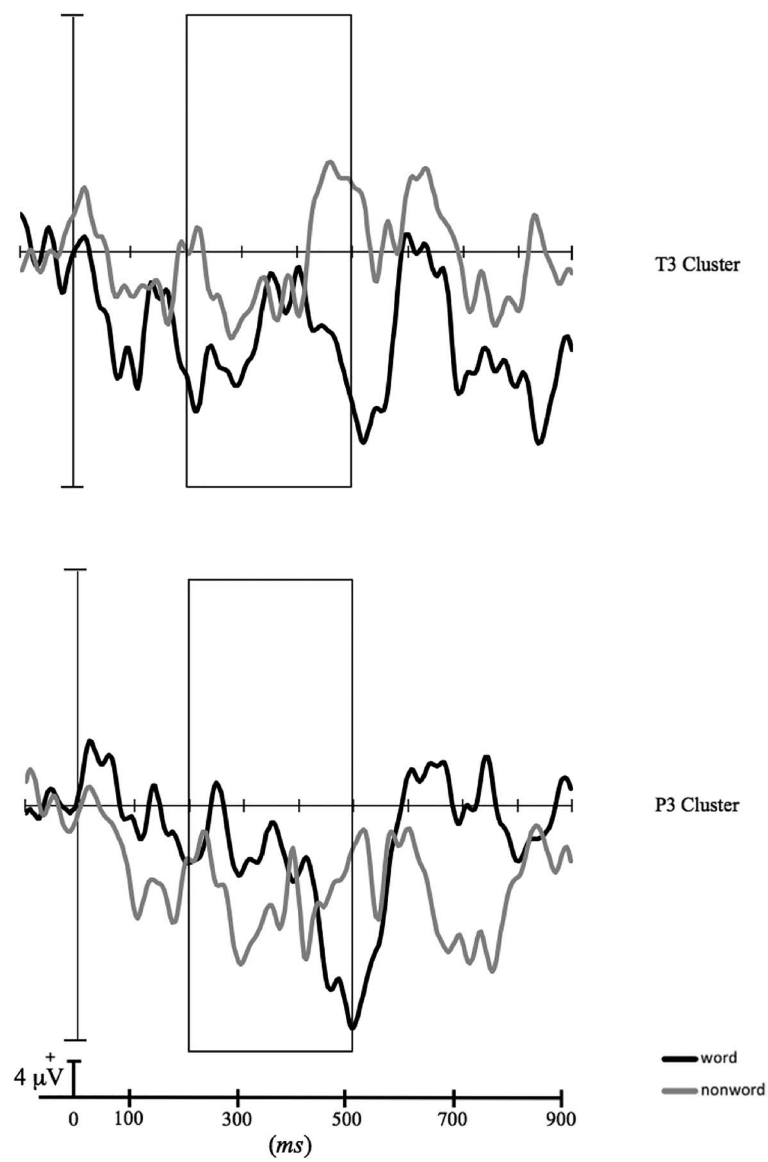
Between-Conditions Differences for Participants Who Understood All Word Stimuli

Grand-averaged waves for participants who reportedly understood all 10 word stimuli (n = 9) are presented in Figure 3. Within this subset of the sample, mean amplitudes to word stimuli between 200 and 500 ms were more negative than mean amplitudes to nonword stimuli at T3 and P3 electrode clusters. However, the between-conditions difference was much larger at T3 (M = −1.27, SD = 3.63) than at P3 (M = −0.02, SD = 3.09). Visual inspection suggests that the response to words at P3 was initially more positive than the response to nonwords within this subset, yielding average amplitudes for each condition that were virtually equivalent. Within this subset, average between-conditions amplitude differences were not significantly different from zero, either at T3, $t(8)$ = −1.05, p = .16, d = −0.31, or P3, $t(8)$ = −.02, p = .49, d = −0.008, though this may have been a function of low power. Achieved power for these analyses was 0.21 for the between-conditions difference test at T3 and 0.05 for that at P3 (G*Power; Faul, Erdfelder, Buchner, & Lang, 2009). The number of word stimuli understood by participants was not correlated with the size of the between-conditions mean amplitude difference at T3 (r = .09, p = .71) or at P3 (r = .027, p = .56), nor with the average amplitude to word at T3 (r = −.07, p = .35). However, the number of word stimuli reportedly understood was moderately correlated with the average amplitude to word at P3 (r = −.33, p = .02). The extent to which this variable (word stimuli understood) influenced the potential utility of this ERP word processing measure was examined in a follow-up analysis.

Statistical Interaction of ERP Measures With Number of Word Stimuli Understood Predicting Receptive Vocabulary

Results for all four linear models testing the influence of word stimuli understanding on the association between each putative ERP measure and receptive vocabulary are presented in Table 3. Only the ERP measure derived from between-conditions mean amplitude differences at T3 (T3 average difference) statistically interacted with understanding of word stimuli (β = −1.57, p = .03, η_p^2 = .1543) to predict concurrent receptive vocabulary scores. Figure 4 illustrates this interaction and charts the threshold level of word stimuli understanding for which predictive slopes are significant. The lower bound of the region of significance was 4.05. Thus, children with ASD who reportedly knew at least five of the word stimuli featured in the ERP paradigm had T3 word–nonword difference scores that were significantly associated with concurrent receptive vocabulary ($r \leq -.47$), such that greater word–nonword differences, with stronger negative amplitude values for word, were associated with better receptive vocabulary. For children who understood all 10 word stimuli, this association was the strongest (r = −.53). In contrast, for children who knew less than half of the word stimuli featured in the ERP paradigm, no ERP measure was significantly associated with receptive vocabulary. Table 3 also indicates that reported understanding of word

Figure 3. Grand-averaged waveforms for word and nonword conditions, averaged across the left temporal (T3) and left parietal (P3) electrode clusters for participants who reportedly knew all 10 word stimuli featured in the event-related potential paradigm. The temporal window of the word processing response documented in previous literature (200–500 ms after stimulus onset) is highlighted.



stimuli significantly predicted receptive vocabulary in all models, as would be expected.

Discussion

The purpose of this investigation was to examine word processing in young children with ASD and to document the utility of four putative measures of word processing derived based on findings in prior literature. Across the entire sample, brain responses between word and nonword stimuli did not significantly differ at the times and locations identified by previous studies. Moreover, none of the four measures of word processing derived using the parameters identified in previous literature were significantly correlated with concurrent scores of receptive vocabulary, without consideration of the participants' word stimuli understanding. However, we documented a significant association between one putative measure of word processing (average word–nonword difference at T3) and concurrent receptive vocabulary that was conditional on the number of word stimuli understood by the child. These preliminary results provide a strong rationale for quantifying word processing as the average word–nonword difference at T3 in future studies, after first ensuring that the ERP paradigm primarily features words understood by each participant.

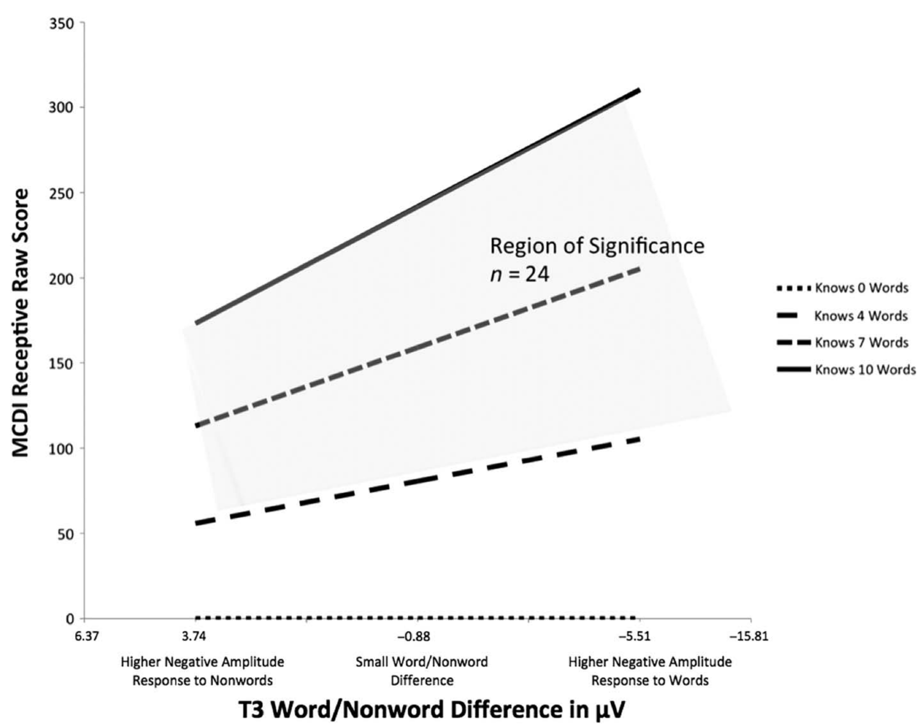
Table 3. Predicting concurrent receptive vocabulary with measures of word processing and number of word stimuli understood.

Variable	Coefficient	*SE*	*t*	*p*	η_p^2
T3 average amplitude to word model					
Intercept	−41.92	29.75	−1.41	.169	.05
T3 Avg Amp	−1.82	8.81	−0.21	.838	.0007
Words Understood	27.44	4.08	6.72	.000***	.61
T3 Avg Amp × Words Understood	−0.35	1.07	−0.33	.741	.005
P3 average amplitude to word model					
Intercept	−43.22	29.87	−1.44	.158	.06
P3 Avg Diff	−1.34	14.24	−0.09	.926	.0002
Words Understood	28.73	4.31	6.66	.000***	.61
P3 Avg Amp × Words Understood	0.43	1.81	0.24	.815	.002
T3 average word–nonword difference model					
Intercept	−37.20	26.53	−1.40	.1711	.05
T3 Avg Diff	0.97	4.41	0.22	.8259	.003
Words Understood	26.58	3.61	7.36	.000***	.65
T3 Avg Diff × Words Understood	−1.58	0.70	−2.26	.0309*	.15
P3 average word–nonword difference model					
Intercept	−42.19	30.25	−1.39	.173	.05
P3 Avg Diff	−1.93	8.13	−0.24	.814	.0004
Words Understood	28.00	4.14	6.75	.000***	.62
P3 Avg Diff × Words Understood	−0.35	1.22	−0.29	.774	.005

Note. T3 = left temporal electrode cluster; Avg Amp = average amplitude to word between 200 and 500 ms after stimulus onset; Words Understood = number of word stimuli featured in the event-related potential paradigm reportedly understood by the participant; P3 = left parietal electrode cluster; Avg Diff = average difference between word and nonword amplitudes between 200 and 500 ms after stimulus onset.

$^*p < .05$. $^{***}p < .001$.

Figure 4. Interaction between the left temporal (T3) difference measure of word processing and the number of word stimuli reportedly understood by the participant predicting concurrent receptive vocabulary. Word processing values reflect the word–nonword average difference score in microvolts from 200 to 500 ms after stimulus onset. Microvolt values on the *x*-axis correspond to the minimum, maximum, mean, and 1 *SD* above and below the mean, in sequential order. Receptive vocabulary is measured by the raw scores from the MacArthur–Bates Communicative Development Inventories (MCDI; Fenson et al., 2007). The gray shaded region reflects the region of significance, which ranges from 4.05 (i.e., 5) to 10 word stimuli understood.



Although the results from our post hoc analyses are perhaps the most important and useful findings of this investigation, further consideration of the other results is merited. At the group level, we observed mean amplitudes to word stimuli that were more negative than mean amplitudes to nonword stimuli at left temporal and parietal sites, which is consistent with waveform patterns documented in previous literature (Kuhl et al., 2013; Mills, Conboy, et al., 2005; Mills et al., 1997; Mills, Plunkett, et al., 2005). However, these differences were not statistically significant. The most salient explanation for this unexpected finding is the fact that participants in our sample varied widely in their familiarity with the word stimuli featured in the ERP paradigm. However, the number of word stimuli understood by participants did not have a linear association with the size of the between-conditions difference at either location. It is possible that the above association is nonlinear, but our small sample size did not allow testing of this hypothesis. Even within the subset of the sample that reportedly knew all the word stimuli, significant between-conditions differences were not observed at left temporal or parietal sites. However, this nonsignificant difference could be due to insufficient statistical power due to the small sample size ($n = 9$). We also considered, but discounted, the contributions of age or excessive motion to our results, as neither the age in months nor the number of trials retained was significantly correlated with the size of the between-conditions differences at left temporal and parietal locations, nor did they interact with any of the putative word processing measures to predict MCDI scores.

Given that previous word processing investigations have documented significant between-conditions differences that were broad and bilateral in younger children with less word comprehension (Mills et al., 1993, 1997), in those who were newly familiar with word stimuli (i.e., after novel word training; Mills, Plunkett, et al., 2005), and in children with ASD with less adaptive social scores (Kuhl et al., 2013), readers may wonder if word processing might have been better indexed at right frontal, temporal, parietal, or occipital sites, at least within the subset of our sample that reportedly knew fewer than five word stimuli, or were "low comprehenders" (i.e., those who had receptive vocabulary scores that were below the sample median). However, additional post hoc exploratory analyses at right hemisphere homologues did not provide any significant findings. This was the case for the entire sample, as well as for participants who knew fewer than five word stimuli, and those who were "low comprehenders."

Evidence for Cerebral Specialization

What conclusions can be drawn regarding the significant interaction between the T3 word–nonword difference measure and the number of word stimuli understood in predicting concurrent receptive vocabulary scores? These results could be interpreted as more precise evidence supporting cerebral specialization associated with increased vocabulary size and word experience (Mills, Conboy, & Paton, 2005). Previous word processing investigations documented word–nonword differences that were left-lateralized and more focally distributed in typically developing children with greater language experience (Mills et al., 1993, 1997). Mills, Conboy, and Paton (2005) argued that those results were indicative of changes in lateral organization of brain activity to words, which were driven by the attainment of specific language milestones (i.e., substantial increases in receptive vocabulary) as well as the amount of experience with the word stimuli. Previous findings that support this hypothesis relied primarily on evidence that shows brain response differences between subsets of participants divided using a median split of scores of receptive vocabulary knowledge. This practice forces a continuous variable into a dichotomous one, resulting in the loss of information. The analyses in the current investigation afforded consideration of the interaction between two continuous variables and demonstrates that the between-conditions differences at T3 were associated with both word experience and receptive vocabulary. Instead of using an arbitrary and sample-specific method to dichotomize word experience, an empirical method was used in the current study to identify the number of understood stimuli needed before the expected association between the ERP measures at T3 (left hemisphere electrode) and receptive vocabulary was detected. In other words, larger word–nonword differences at T3 are not exclusively indicative of participant familiarity with the words presented. Participants who were reportedly familiar with word stimuli but had low total receptive vocabularies had smaller between-conditions differences at left temporal electrodes. In contrast, participants who reportedly knew more of the word stimuli and had higher total receptive vocabularies had higher between-conditions differences at T3, with more negative amplitudes to word. Thus, this study adds more precise evidence to support the hypothesis that greater word experience and greater comprehension abilities together contribute to changes in the lateral organization of the brain response to words, even for children with ASD.

Utility of the T3 Word–Nonword Difference Measure

The current results provide evidence that word processing may be best indexed as the word–nonword average amplitude difference at left temporal electrodes, at least for children with ASD, given that the association between this measure and concurrent receptive vocabulary increased predictably based on the number of word stimuli understood. Quantifying word processing as the word–nonword difference, rather than as the average amplitude to word, also has the potential to better control for individual differences in overall average amplitude to any auditory stimuli. However, these results depart from the only other investigation of word processing in children with ASD, which documented strong predictive (though not concurrent) nomological validity of a word processing measure derived as the average amplitude to word at a left parietal electrode (Kuhl et al., 2013). Beyond differences in participant understanding of word stimuli, other potential differences in our sample may account for our divergent results. It is possible that, as a group, children with ASD in our study were qualitatively different from those in the study by Kuhl and colleagues (2013). At the time of entry in the larger study (Yoder et al., 2015), participants in our study had lower cognitive scores than those studied in the word processing investigation by Kuhl and colleagues (2013). However, post hoc analyses suggest that neither mental age nor ADOS severity scores were significantly correlated with scores from any of the four putative ERP measures of word processing, nor did they influence the associations between these scores and concurrent receptive language. Of course, because funding for the ERP procedure was acquired 2 years after funding for the larger study, the timing of the ERP procedure relative to the administration of these assessments (i.e., MSEL and ADOS) varied widely across participants. Thus, the lack of concurrent scores of mental age and autism severity limits our ability to detect associations that might be present or to definitively conclude that our sample was qualitatively different at the time of ERP from that featured in the study by Kuhl and colleagues (2013). Therefore, though there is preliminary evidence to suggest that the left temporal word–nonword amplitude difference might be a scientifically useful way to quantify word processing, replication of these results is needed.

Strengths

The primary strengths of this study stem from our reliance on prior findings to (a) derive putative measures of a cognitive process and (b) generate specific falsifiable hypotheses to test the utility of these putative measures. If electrophysiological measures are to have value for clinical populations, it is vital that researchers advance the practice of making a priori determinations regarding the timing and location of a given construct, use those parameters to derive putative measures of the given construct, and document the nomological validity of those putative measures by examining their association with theoretically related constructs. ERP investigations that document the timing and location of mean differences between stimuli are useful for the advancement of developmental theory. However, brain–behavior correlations provide additional and arguably more rigorous evidence of the validity of ERP measures than do within-procedure or mean change scores.

A second strength of the study is that we declined to exclude participants who did not reportedly know all of the word stimuli featured in the ERP paradigm, opting instead to test whether their understanding of word stimuli in analyses affected the expected association between ERP and receptive vocabulary. This allowed us to retain a larger sample size and to conduct a more specific test of the extent to which our putative measures truly indexed word processing. Although only nine participants reportedly knew all 10 words, the cut point at which the relevant word processing score was significantly associated with receptive vocabulary was substantially lower than 10 (i.e., 5). Thus, excluding participants who did not know all 10 words would have led us to exclude individuals for whom this measure still has validity and would have reduced the sample size to such an extent that statistical tests would have been greatly underpowered.

Finally, this is only the second ERP investigation to examine individual differences in this construct, word processing, in children with ASD. Deficits in language comprehension are common in children with ASD and may contribute to later deficits in social and academic functioning. Investigations of word processing have the potential to deepen our understanding of the early development of language comprehension in this population.

Limitations

Several features of our investigation limit the strength of our conclusions. First, even within the subset of the sample that knew all 10 word stimuli, we were unable to document significant between-conditions amplitude differences either at planned electrode sites or at right hemisphere homologues in exploratory post hoc analyses. Group-level discrimination has been documented in all previous word processing investigations, and we were unable to replicate this finding. The small number of participants who reportedly knew all 10 word stimuli likely limited our power to detect significant between-conditions differences. It is also possible that our small sample size limited our power to detect significant interactions between the other a priori derived measures of word processing and the number of word stimuli understood in predicting concurrent receptive vocabulary.

Second, one previous investigation documented predictive associations between an ERP measure of word processing and receptive language (Kuhl et al., 2013). In the current investigation, we were unable to test whether measures of word processing predicted later receptive vocabulary. Although our data were taken from a longitudinal study of language development in children with ASD (Yoder et al., 2015), funding for the ERP procedure was acquired 2 years after the start of the study. As such, most of our participants completed the ERP procedure toward the end of their participation in the study. Thus, later receptive vocabulary was only available for a small number of participants, only a few of whom reportedly knew at least five word stimuli. Establishing the predictive validity of an a priori derived measure of word processing would go a long way toward demonstrating the measure's potential clinical utility. Thus, further work that examines the extent to which the average difference at T3 predicts later outcomes of interest is needed.

Third, excessive data noise may have contributed error to our measure of word processing, limiting our ability to detect significant between-conditions differences and brain–behavior associations. Though we took careful steps during postcollection data processing to remove artifacts from our data, substantial between-conditions differences in our grand-averaged baseline data, particularly at left parietal sites, suggest that these procedures may not have been sufficient. We were only able to retain approximately half of the average number of trials retained in the previous word processing investigation of children with ASD (Kuhl et al., 2013). Though we verified that the number of trials retained was not significantly correlated with any of the putative measures of word processing, measurement theory dictates that more trials per condition would have stabilized ERP scores. Alternatively, it is possible that our artifact removal procedures were more strict than those of previous investigations. Without knowing the specific data cleaning procedures of Kuhl and colleagues (2013), we cannot conclude that noise was greater in our own data.

Future Investigations

ERPs have the potential to be of great clinical use, and evidence that the measure is valid at the individual level, rather than the condition mean level, is necessary for this to be the case. Thus, replication of the current results is needed both in typically developing and clinical populations. Investigators should further examine the relative validity of the average word–nonword difference at T3 and the average amplitude to word at P3 by documenting the associations between these measures and concurrent and later measures of receptive language. In addition, because validity is limited by the reliability of a measure, investigators should document the short-term test–retest reliability of these ERP measures of word processing. The number of trials necessary to obtain stable ERP scores of word processing might potentially be examined with generalizability studies comparing scores derived from differing numbers of trials. Furthermore, our results demonstrate that investigators must track participant's word stimuli understanding and either ensure that participants are familiar with at least half of the word stimuli featured in the ERP paradigm or control for participant word stimuli understanding in statistical analyses. Samples should feature enough participants who reportedly know the word stimuli in order to obtain the power necessary to detect associations of interest.

Future demonstrations that a left temporal ERP measure of word processing is a moderator or mediator of change effected by language intervention, which can be used to identify subgroups of nonresponders ahead of intervention or early in the intervention process, would be useful. It might even prove more useful than behavioral measures for diagnostic purposes and could accelerate diagnoses for young children who might benefit from early intervention. Still, even if valid and stable brain-based measures are identified, additional barriers to clinical translation remain. Large-scale investigations will be needed to obtain normative data for typical and clinical populations, and the current availability of relevant equipment may not be extensive enough to meet the demands of widespread clinical use.

Conclusion

This paper provides preliminary evidence that the average word–nonword difference at left temporal electrodes is the most scientifically useful way to index word processing in children with ASD. This paper also highlights the difficulties associated with the identification of neural measures of language processing that could be clinically useful. Though ERP methods have the potential to illuminate new information about cognitive processes that are difficult to observe through behavior, like behavioral measures, the degree to which they reflect the intended process is influenced by several extraneous variables. Carefully designed investigations and replication studies are needed to refine the science of ERP investigations.

Acknowledgments

This research was funded by a grant from the National Institute on Deafness and Other Communication Disorders (R01DC006893), which was awarded to Paul Yoder, and a grant from the National Institutes of Health (U54HD083211), which was awarded to Elisabeth Dykens. The first author was supported by a grant from the Office of Special Education Programs (H325D100034A), which was awarded to Ann Kaiser. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the Office of Special Education Programs. We are very grateful to our wonderful staff (Elizabeth Gardner, Nicole Thompson, Paula McIntyre, Ariel Schwartz, Tricia Paulley, and Kristen Fite) and to the families who trust us with their precious children. We would also like to acknowledge Magdalene Jacobs for her extremely useful assistance during the final stages of this work.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. New York, NY: Sage.

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

American Psychiatric Association. (2000). *Diagnostic and statistical manual* (4th ed.) *Text Revision (DSM-IV-TR)*. Washington, DC: Author.

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., . . . Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 594–604.

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.

Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the Mullen Scales of Early Learning and the differential ability scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities, 116*(5), 331–343.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30,* 213–236.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillside, NJ: Erlbaum.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship between Autism Diagnostic Observation Schedule–Generic (ADOS-G), Autism Diagnostic Interview–Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders, 34,* 129–137.

Farmer, C., Golden, C., & Thurm, A. (2016). Concurrent validity of the differential ability scales, with the Mullen Scales of Early Learning in young children with and without neurodevelopmental disorders. *Child Neuropsychology, 22,* 556–569.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur–Bates Communicative Development Inventories*. Baltimore, MD: Brookes.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477–501.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(5), 693–705.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*(4), 613.

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 & Communication Disorders, 45,* 681–690.

Hus, V., Pickles, A., Cook, E. H., Risi, S., & Lord, C. (2007). Using the Autism Diagnostic Interview—Revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry, 61*(4), 438–448.

Kuhl, P. K., Coffey-Corina, S., Padden, D., Munson, J., Estes, A., & Dawson, G. (2013). Brain responses to words in 2-year-olds with autism predict developmental outcomes at age 6. *PLoS One, 8,* e64967.

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,* 205–223.

Mahajan, Y., & McArthur, G. (2011). The effect of a movie soundtrack on auditory event-related potentials in children, adolescents, and adults. *Clinical Neurophysiology, 122,* 934–941.

Mills, D. L., Coffey-Corina, S., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology, 13,* 397–445.

Mills, D. L., Conboy, B., & Paton, C. (2005). Do changes in brain organization reflect shifts in symbolic functioning? In L. L. Namy (Ed.), *Symbol use and symbolic representation: Developmental and comparative perspectives* (pp. 123–153). London, United Kingdom: Psychology Press.

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development, 20,* 19–31.

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H., & Werker, J. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience, 16,* 1452–1464.

Mills, D. M., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience, 5,* 326–342.

Mills, D. M., Coffey-Corina, S. A., & Neville, H. J. (1994). Variability in cerebral organization during primary language acquisition. In G. Dawson & K. Fischer (Eds.), *Human behavior and the developing brain* (pp. 427–455). New York, NY: Guilford Publications.

Molfese, D. L. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain and Language, 38*(3), 345–363.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development, 8,* 197–211.

Molfese, D. L., Molfese, V. J., & Espy, K. A. (1999). The predictive use of event-related potentials in language development and the treatment of language disorders. *Developmental Neuropsychology, 16,* 373–377.

Moore, T. (2016). erp.easy: Event-Related Potential (ERP) Data Exploration Made Easy. R package version 1.0.0. http://CRAN.R-project.org/package=erp.easy

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (pp. 58–64). Circle Pines, MN: AGS.

Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E. (2014). Language assessment in children with autism spectrum disorder: Concurrent validity between report-based assessments and direct tests. *Research in Autism Spectrum Disorders, 8*(9), 1100–1106.

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43,* 418–431.

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two year follow-up. *Autism Research, 1,* 97–107.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72,* 184–187.

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech acquisition in older nonverbal individuals with autism: A review of features, methods, and prognosis. *Cognitive and Behavioral Neurology, 22*(1), 1–21.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6,* 135–147.

Rutter, M., Mawhood, L., & Howlin, P. (1992). Language delay and social development. In P. Fletcher & D. Hall (Eds.), Specific speech and language disorders in children: Correlates, characteristics and outcomes (pp. 63–78). London, England: Whurr Publishers.

Swineford, L. B., Guthrie, W., & Thurm, A. (2015). Convergent and divergent validity of the Mullen Scales of Early Learning in young children with and without autism spectrum disorder. *Psychological Assessment, 27,* 1364–1378.

Tager-Flusberg, H. (2000). Understanding the language and communicative impairments in autism. *International Review of Research in Mental Retardation, 23,* 185–205.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6,* 468–478.

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75,* 1067–1084.

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,* 1259–1273.

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer Publishing Company.

Yoder, P., Watson, L. R., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 45*(5), 1254–1270.