The Psychometrics of Several Neural Measures of Lexical Access in Children with Autism

By
Micheal Paige Sandbank

Dissertation
Submitted to the Faculty of the
Graduate School at Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSPHY
in

Special Education

May, 2015

Nashville, TN

Approved:

Paul Yoder, Ph.D.
Sasha Key, Ph.D.
Jennifer Ledford, Ph.D.
Erin Barton, Ph.D.

TABLE OF CONTENTS

Page

Appendix

INTRODUCTION

Lexical access is the process of retrieving previously acquired words from mental storage. It can be imagined as the process of sifting through a mental rolodex of known words, in order to determine whether the word has been heard before. This cognitive process is an important step in language comprehension, and, consequently, it influences language production. It is possible that lexical access mediates change facilitated by language intervention. If this is the case, then language interventions could be adapted for nonresponders by emphasizing treatment components that improve lexical access. Before we can test the potential importance of lexical access to children with autism spectrum disorders (ASD), we need measures of lexical access with good psychometric properties in this population.

Various experimental studies have attempted to capture behavioral indicators of lexical access in young children. Although these behavioral indicators are problematic for developmentally young children with ASD, reviewing what is known from these methods provides a context for examining other potential measures of lexical access. Multiple investigations have indicated that children as young as 8 months exhibit head-turning preference for familiar words over nonsense words that are phonetically similar (i.e. cup vs. tup; Hallé & de Boysson-Bardies, 1994; Jusczyk & Aslin,1995), and prefer words that are phonetically similar to familiar words over unfamiliar words (Hallé & de Boysson-Bardies, 1996). By 14 months of age, children look more quickly at picture representations of words when those words are correctly pronounced, suggesting they can distinguish between correct and incorrect pronunciations of well-known words (Swingley & Aslin, 2000; 2002). Between 15 and 24 months, latency of looking at a picture of the word's referent after the word is spoken decreases reliably with age

(Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998). These results suggest that the development of lexical access occurs in typical children across the first two years of life.

To date, only one study has investigated lexical access using preferential looking techniques in children with ASD (Venker, Eernisse, Saffran, & Weismer, 2013). These children were between the ages of 3 and 6 years. Results suggested that the developmental trajectory of lexical access in children with ASD is delayed, but follows a pattern similar to that of typically developing children. However, there is much greater variability at the individual level. Accuracy of looking was strongly and significantly correlated with latency of looking, and with receptive and expressive language.

While behavioral experiments can provide some insight into the development of lexical access in young children, there is reason to doubt the validity of using gaze as measure of lexical access in young children, especially those with ASD. Even if their lexical access is intact, preferential looking is likely to be impaired in children with ASD, because they exhibit marked difficulty disengaging attention from one object to shift focus to another, when compared to typically developing peers (Elsabbagh et al., 2009) or other clinical populations (Landry & Bryson, 2004). In addition, experimental paradigms used to assess lexical access actually require children to engage in multiple cognitive processes to accomplish the task of accessing lexical information (e.g., attend to linguistic stimulus, detect auditory stimulus, remember the phonemic string, and finally, find the word in the mental lexicon). Thus, to isolate this process from other stages of auditory processing, investigators must use a measurement instrument that has high temporal resolution. Such fine-grained temporal resolution is difficult to achieve with preferential looking experiments. Even when videos are coded frame-by-frame, the resulting data amounts to 33 ms momentary time sampling (Fernald, Zangl, Portillo & Marchman, 2008). Thus, a measure of lexical access that has better temporal resolution is needed.

EEG for Measuring Lexical Access

The high temporal resolution of electroencephalography (EEG) allows investigators to conduct analyses of the brain's response to stimuli at much higher temporal resolution than preferential looking. EEG is the direct measurement of the brain's electrical activity, measured in microvolts, using electrodes placed at the scalp level. These electrodes record voltage fluctuations associated with spontaneous neural firing. EEG recordings can be time-locked with the onset of a stimulus to examine the neural activity associated with stimulus processing, and these are referred to as *event-related potentials* (ERPs). Event-related voltage values are recorded with high temporal frequency (e.g. once every four milliseconds). Thus, ERPs provide a useful way for investigators to capture lexical access *as it occurs*, rather than indirectly using preferential looking. Unlike some other neuroimaging methods, passive ERP procedures do not require a participant's absolute stillness, task-comprehension, or motivation to do a task. This makes passive ERP a preferable method for measuring cognitive activity in populations that exhibit difficulty with active participation in behavioral paradigms. For all of these reasons, ERPs have potential value for measuring lexical access in young children with ASD.

Current evidence. Previous ERP studies of word vs. nonword processing in typically developing children have shown differential patterns of neural activity emerging as early as 13 months of age and unfolding over the second year of development as acquisition of vocabulary progresses (Mills, Coffey-Corina, & Neville, 1993, 1994, 1997). These studies document increased negative average amplitude responses to known versus unknown words occurring between 200 and 500 ms after stimulus onset. In young children (13-17 mos), this amplitude difference was largest at temporal and parietal electrodes in both hemispheres. In older children (20 mos) with larger vocabularies (>150 words on parent report), the observed amplitude difference was limited to electrodes in the left-hemisphere temporal and parietal regions. Investigations by the same research group have since replicated and extended these results to known words versus nonsense words (Mills et al., 2004), newly learned vocabulary words (Mills, Plunkett, Prat, & Schafer, 2005), and to late-talkers, before and after vocabulary spurts (Mills, Conboy, & Patton, 2005). In summary, these findings suggest that voltage potentials associated with children's neurological responses to known words are likely to be detected between 200-500 ms post-stimulus, at left hemisphere temporal and parietal electrode sites.

Only one study has examined lexical access in children with ASD using this word versus nonword ERP paradigm. Kuhl and her colleagues (2013) compared the neurological responses of 2 year-old children with ASD, dichotomized into two groups using scores that reflected the level of social skills deficit, to those of same-aged typically developing children. Similarly to previous results, typically developing children exhibited significantly different mean amplitude responses to known versus unknown words between 200-500 ms at a single left lateralized temporal electrode (T3, see Figure 1). Children with ASD who had more adaptive social skills exhibited similar activation patterns, except that significant amplitude differences were observed at the left lateralized parietal electrode (P3). In contrast, children with ASD who had less adaptive social scores exhibited a more diffuse, right-localized amplitude difference between the two stimulus types (at right frontal (F8) and temporal (T4) electrodes). This latter finding points out that the electrode(s) at which the neural responses to words are detected may differ in typically developing children and those with ASD. More importantly, in children with ASD aged 2 years old, mean amplitude values observed at the left parietal electrode between 200-500 ms after exposure to known words significantly and strongly predicted receptive language scores 2 ($r = -.671, p < .01$) and 4 ($r = -.785, p < .001$) years following the ERP.

Limitations of prior ERP investigations. While recent neurological investigations of lexical access are beginning to deepen our understanding of the development of this process in young children, we have several reasons to believe that ERP measurement of lexical access is still in its infancy. Lab-specific differences in the location of the reference electrode can affect the observed results and reduce replicability of findings. The reference electrode is the sensor from which voltage readings are subtracted from other electrodes (Luck, 2005). The location of the reference electrode may be anywhere on the scalp, or even on the nose. In the case of many ERP components, including those associated with lexical access, there is no agreed-upon convention for the 'best' reference electrode location. Many investigators have shown that several aspects of the data from ERPs are affected by the location of the reference electrode (Nunez et al., 1997; Murray, Brunet, & Michel, 2008). These include (a) the shape of the ERP waveform, (b) variance around the ERP, (c) the latency of the observed response, and (d) the scalp location of the observed maximum neural response to experimental stimuli. The fact that ERP variable values can be largely altered by variations in data acquisition methods should reduce our confidence that investigators are likely to replicate findings from ERP studies when different references are used, even when previously documented neural responses are present.

In addition, the most common approaches to ERP variable derivation are prone to yield sample specific results (Murray, Brunet, & Michel, 2008). Any approach to ERP data analysis begins with two questions: (a) what is the temporal focus of the analysis (when?) and (b) what is the locational focus of the analysis (where?). Because ERP experiments can yield more than a million values, initial attempts to identify replicable aspects of brain activity associated with the construct of interest are often affected by type I errors associated with multiple statistical tests. Subsequent investigations of a construct typically limit the temporal and locational focus of the analysis to electrodes and time windows that yielded statistically significant results in previous investigations, even if the results of the previous investigation were associated with a different population. Because this approach, which we will refer to as the *extant literature approach,* usually involves methods that lead investigators to focus on a limited number of electrodes and time windows, they may ignore findings from other time windows or electrode locations that might be more informative. This is especially the case when our investigations include children with ASD in the early stages of language acquisition, who may exhibit neural activity that is atypically located and timed.

Because of the issues outlined above, it is reasonable to expect that this approach to ERP data reduction has contributed significant amounts of measurement error to neural investigations of cognitive constructs such as lexical access. This measurement error has likely caused results that cannot be replicated in young children with ASD at a sufficiently specific level to afford future theory testing. In science, replication is the foundation of the accumulation of knowledge (Lachenmyer, 1971). Consequently, our knowledge about difficult-to-observe cognitive responses like lexical access in children with ASD is advancing very slowly.

An alternative approach to ERP data analysis. Topographic ERP analysis is another approach to ERP data analysis that is less vulnerable to many of the problems outlined above and may produce more replicable results than the traditional approach already discussed. Topographic analysis involves identifying statistically distinct, relatively stable patterns of neural firing called *microstates*. The timing and presence of the microstates that differ between conditions in the reference group (e.g., a typically developing group) then becomes the temporal focus of the data analysis. This approach has several advantages over traditional ERP data analysis. First, topographic analysis results are reference independent; they do not change based on the location of the reference electrode. Second, data from the entire high-density map of electrodes across the entire post-stimulus time window is used to conduct topographic ERP analysis, but rigorous correction for multiple significance testing is applied. Thus, the topographic approach prevents the introduction of experimenter bias associated with selective data analysis or "data peaking" and corrects for repeated significance testing (sources of type I error), while still ensuring investigators do not overlook potentially relevant scalp locations or time segments of the post-stimulus response (i.e. sources of type II error). Moreover, when topographic analysis of ERP data from a typically developing sample is used to identify the temporal window of interest, which is then applied to the atypical sample, this approach retains a high level of falsifiability in tests of the hypotheses in the atypical sample (Yoder, Molfese, Murray, & Key, 2013). We call the application of information from the typically developing sample to the ASD sample the *normative topographic analysis* (NTA) approach. However, because the use of the NTA is predicated on the analysis of data from the entire scalp, investigators must have the capability to measure the neural response across the entire scalp. While a recommendation has not been made regarding the minimum number of electrodes required for NTA, it stands to reason that at least 21 electrodes (consistent with the 10-20 system) are needed, and more would be better. Perhaps for this reason, NTA is used less frequently than more traditional approaches to ERP quantification. Consequently, the approach still requires further evaluation to be established as a good approach for deriving measures of lexical access, particularly in children with ASD.

The need for psychometric evaluation. Before addressing whether ERP measures of lexical access can be used to test theories of receptive vocabulary development in children with ASD, it is important to examine the psychometric properties of competing ERP measures of lexical access in this population. The goal of psychometric evaluation is to establish the quality of a measure of a given construct (Nunnally, 1978). Such evaluation is necessary if scores from a measure are to have utility as the basis for clinical interpretations and actions (Messick, 1989).

Approaches to Psychometric Evaluation

There are multiple ways to establish the scientific utility, or validity, of a measurement instrument as a measure of a given construct (Yoder & Symons, 2010). The type of validation process that is relevant to this investigation is nomological validity. Cronbach and Meehle (1955) proposed this particular type of construct validity as relevant when a measure lacks an "ultimate standard" to which it can be compared. There is not presently a gold standard measure of lexical access for young children with ASD in the early stages of language acquisition. The process of establishing the nomological validity of a measure involves testing theoretically predictable (a) associations with measures of other constructs and (b) differences between groups that differ on the construct of interest. The set of predicted associations and differences is referred to as the *nomological net* (Cronbach & Meehle, 1955).

Theoretical associations with measures of related constructs. Among the constructs relevant to the nomological net that surrounds the construct of lexical access are (a) receptive language, (b) the amount of parent-provided input that is likely to be processable by the child (processable parent input), and (c) the child's attention towards such input. The theories supporting the associations between these constructs and lexical access are described here.

The relationship between lexical access and receptive language. Theoretically, a child's ability to sift through and retrieve mental representations of words is positively related to receptive language. To the extent that preferential looking paradigms measure lexical access, studies using this method support this prediction. Fernald, Perfors, and Marchman (2006) detected significant concurrent correlations between reaction times and accuracy scores on preferential looking experiments and scores of receptive language in young children who were 25 months old. In addition, these investigators found that speed and accuracy of looking at 15 months of age predicted accelerated growth in expressive vocabulary at 25 months of age. Though the evidence supporting the relationship between lexical access and receptive language stems from investigations of typically developing children, there is no reason to think that this association differs in children with ASD.

The relationship between lexical access and processable parent input. Moreover, we can speculate that a child's lexical access might be influenced by the amount of understandable child directed speech that a parent uses. Recall that lexical access can be imagined as the ability to efficiently sift through a mental rolodex of known words. The construction of this mental rolodex occurs with exposure to words. In order for this exposure to occur, the parent must provide linguistic input that is accessible to the child. Whether a child attends to parental input, and thus processes it, varies according to the nature and type of linguistic input.

A wealth of evidence supports the idea that young children benefit from a specific type of linguistic input in order to attend to and learn from language models. That is, mere exposure to linguistic input is insufficient for language learning. For example, multiple studies have shown that prelinguistic children exhibit preference for "child directed speech" (CDS; Cooper, Abraham, Berman, & Staska, 1997; Fernald, 1985; Pegg, Werker, & McLeod, 1992), which is differentiated from adult-directed speech by its higher pitched, shorter, syntactically simpler utterances that feature longer pauses and expanded intonation contours (Ferguson, 1964; Grieser & Kuhl, 1988; Sherrod, Crawley, Peterson, & Bennett, 1978; Snow, 1972).

In addition, research suggests that child word learning depends, in part, on certain interactional behaviors exhibited by the parent. For example, in studies of typically developing children, parental references to objects already within a child's focus were positively correlated with child vocabulary at 21 months, and children were more likely to learn words that referred to the object of a child's current focus (Tomasello & Farrar, 1986). In children with ASD, parent linguistic utterances that referred to the child's attentional focus or communicative referent were identified as a value-added predictor of expressive and receptive language growth (Yoder, Watson, & Lambert, 2014). Thus, scores from a valid neurological measure of lexical access should be positively correlated with measures of parental input that is directed to the child and referring to the child's current attentional focus or communicative referent.

The relationship between lexical access and attention to CDS. Finally, linguistic information is not likely to be processed if it is ignored. This issue is especially salient for young children with ASD. The preference for CDS that is typically observed in young children has been shown to be generally lower and more variable in those with ASD, and this may contribute to the social and linguistic developmental delay observed in this population. Children with ASD who fail to exhibit preference for CDS also exhibit deficient auditory processing, a preceding cognitive step for lexical processing (Kuhl, Coffey-Corina, Padden, & Dawson, 2005). Studies tracking attention to child directed speech (aCDS) in children with ASD have found that it is positively associated with concurrent and follow-up scores of receptive language (Paul, Chawarska, Fowler, Cichetti, & Volkmar, 2007; Yoder, Watson, & Lambert, 2014). Similarly, in a study of young boys with ASD, Watson, Baranek, Roberts, David, and Perryman (2010) found that vagal activity (a possible measure of attention) during CDS accounted for unique variance in communication skills measured one year later. Taken together, this evidence suggests that scores from a valid neurological measure of lexical access should be positively associated with scores measuring a child's attention to CDS.

Discriminating between groups. In addition to confirming predictable associations, evidence for nomological validity can also be strengthened by showing that a measurement discriminates between diagnostic groups known to differ on the construct the variable of interest purports to measure. Given the relationships between the constructs outlined above, it is reasonable to expect that children with ASD who have deficits in receptive and expressive language will also exhibit impaired lexical access. This difference should be apparent when these children are compared to age-matched children that exhibit typical expressive and receptive language abilities. Thus, a neurological measure of lexical access should distinguish between children with ASD and an age-matched typically developing comparison group.

Reliability. A scientifically and clinically useful neurological measure of lexical access should have acceptable test-retest reliability. The validity of a measure cannot exceed the square of its reliability (Nunnally, 1978). Thus, reliability limits validity, and should therefore be examined as a part of a measure's psychometric evaluation. There are multiple types of reliability. The one that is relevant to the current investigation is test-retest reliability – whether lexical access scores from multiple testing occasions similarly rank participants. Importantly, test-retest reliability is the type of reliability that is most likely to constrain nomological validity (McCrae, Kurtz, Yamagata, & Terraciano, 2011). What qualifies as acceptable in this regard is somewhat arbitrary, as is the case with all threshold levels of statistical values. For the current investigation, we chose .6 as a minimum for acceptable test-retest reliability. This level of stability has been called “good” (Mitchell, 1979). Also relevant is the relative stability of different approaches to quantifying ERP measures of lexical access –whether one approach yields measures that are more stable than others.

Current Investigation

The purpose of this project was to investigate and compare the psychometric properties of multiple ERP variables of lexical access. These variables were derived from the two separate approaches to ERP quantification discussed above. Our investigation of the psychometrics of these variables as early neurological measures of lexical access involved examination of test-retest reliability, discriminative validity, and concurrent and predictive nomological validity. Table 1 details the variables and corresponding research questions of the current investigation. The research questions were as follows:

1. For each putative ERP measure of lexical access: (a) What is the ranking of the test-retest stability of a single ERP data collection session? (b) How many ERP sessions is it necessary to average scores across to meet or exceed a threshold level of test-retest stability (i.e. 0.6)?
2. For each putative ERP measure of lexical access, what is the significance of and ranking of the effect size for between-group mean differences on the ERP variables between young children with ASD and typically developing children matched on chronological age?
3. For each putative ERP measure of lexical access: (a) In typically developing children, what is the significance and ranking of the effect sizes for the association between the ERP variable and concurrently-measured receptive vocabulary? (b) In children with

ASD, what is the significance and ranking of the effect sizes for the association between the ERP variable and concurrent receptive vocabulary, (c) concurrent processable parent input, (d) concurrent attention to child directed speech (aCDS), and (e) later receptive vocabulary?

METHOD

Study Design

The research questions in the present investigation were answered using a subset of the data initially collected for a larger, longitudinal, correlational study that examined value-added predictors of spoken language in children with ASD (Yoder, Watson, & Lambert, 2014). For the larger study, young children with ASD were given a battery of assessments at study entry, and at 4 subsequent measurement periods, spaced apart by approximately four-months. Table 2 indicates the assessments that were relevant to the proposed investigation and the periods at which they were given. The measurement periods of MacArthur Bates Communicative Development Inventory (MCDI; Fenson et al., 2007) scores used in the current study were linked to the period at which the ERP procedure was collected. See the measurement section for more detail on this point. For the current investigation, a group of young typically developing children were recruited to serve as a age-matched comparison group to enable the normative topographic approach and afford tests of research questions involving the typically developing group. Typically developing participants completed a single measurement session, during which ERP and receptive vocabulary data were collected. The research designs for the current investigation varied according to research question, and are described as follows:

1. A repeated measures nonexperimental design was used to (a) evaluate and rank the test-retest reliability of each ERP variable, and (b) project the number of assessment sessions across which it is necessary to average to achieve at least .6 reliability.
2. An intact group comparison design was used to assess the extent to which each ERP measure discriminates between diagnostic groups.

3. A concurrent correlational design was used to test the associations between the ERP putative measures of lexical access and (a) receptive language in typically developing children, (b) receptive language in children with ASD, (c) processable parent input in children with ASD, and (d) aCDS in children with ASD.(e) A longitudinal correlational design was used to test the predictive associations.

Participants

Selection criteria for the larger sample of children with ASD. For the larger study, recruited participants (a) were between 24 and 48 months of age, (b) had been diagnosed with an ASD based on criteria listed in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition-Text Revision (American Psychiatric Association, 2000) and the Autism Diagnostic Observations Schedule (ADOS; Lord, Rutter, DiLavore & Risi, 1999), (c) scored 20 words or less on the productive scale of the Words and Gestures form of the MCDI, and (d) said 5 different words or less during a 15 min language sample. Children with comorbid sensory motor impairments, metabolic or progressive neurological disorders, and genetic syndromes were excluded.

Children with ASD in the proposed study. Because funding for the ERP procedure was acquired 2 years after the onset of the larger study and because the ERP procedure was administered at only one of the two sites for the larger study, a subgroup of 34 children (27 boys) was available for the proposed study. The mean age of the participants in this ASD group for the proposed study was 45.40 months (*SD* = 9.57). Other descriptive information on these participants with ASD is presented in Table 3. When the measure was administered at the period at which the ERP was collected, reported scores are from that period; otherwise, reported scores are from Time 1 as defined in the larger study.

Typically developing children. A group of 60 age-matched typically developing children also experienced the EEG procedure. Because the comparison group was matched on age and had no developmental delay, their receptive language scores were significantly higher than those of participants with autism (*d*=2.47). Additional descriptive information about the typically developing comparison group is also presented in Table 3.

EEG Data Acquisition Procedures

EEG data were collected using a hydrocel 128-channel net (EGI, Inc., Eugene, OR), a 250 Hz sampling rate, .1-100 Hz filtering, and Cz referencing. This electrode net can be quickly and easily placed on a participant's head in approximately 2 minutes. Impedances were adjusted to <40 KOhms just before data acquisition. All auditory stimuli were presented at 75 dB SPL in random order, 3 times each, with a varied intertrial interval of 1500-2500 ms, which prevented habituation to stimulus onset. The entire session consisted of 60 trials, and lasted approximately 10 min. The participants were not asked to respond to stimuli in any particular way. To facilitate cooperation during data acquisition, an age-appropriate video with sound muted was shown to participants. Video stimuli did not differ between conditions. Thus, it is assumed that the neural response elicited by the video stimuli would not contribute to observed significant differences between conditions within participants.

Stimuli. Word stimuli comprised a set of 10 English words that are typically understood by infants 8-12 months of age. Nonword stimuli were 10 pronounceable nonsense words matched on word duration and number of syllables. Experimental stimuli were chosen based on procedures used in precedent investigations (Mills et al., 2004), with some slight alterations (see Table 3). Two words were changed from the list used in the preceding investigation because the age of typical acquisition for these words was later than our developmental period. Two nonwords were changed from the list used in the preceding investigation because these two nonwords sounded like real words. During EEG data acquisition, children were exposed to naturally spoken stimuli from the two randomly alternating experimental conditions, word and nonword.

Post-collection data processing. Impedances were rechecked immediately after data acquisition. During data processing, signals were smoothed using a 30 Hz low pass filter. Each EEG record was divided into 700 ms sections, which were time-locked with the onset of a single stimulus. The first 100 ms reflected baseline (pre-stimulus) neural activity (measured in microvolts, μV), and the subsequent 600 ms reflected post-stimulus potentials. Single trial data were screened for channels that had more than the criterion level of allowable noise using NetStation. A "bad" channel was defined as one that featured a voltage shift in excess of 150 μV within a trial. An eye-movement artifact was defined as an eye-channel difference that exceeded 70 μV. Trials with 15 or more bad channels 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). This method uses the data from adjacent electrodes to reconstruct the value of the channel in question. After artifact removal and data correction, data were averaged across trials for each stimulus condition, re-referenced to an average reference, and baseline corrected.

ERP Measures

The aspects of neural response measured. We attempted to quantify individual differences on two dimensions of neural response to words: (a) response strength and (b) speed. Response strength refers to the amount of electrical energy detected on the scalp. High levels of response strength can occur because of (a) highly simultaneous firing of individual neurons and/or (b) highly consistent across-trial timing of neural firing relative to the stimulus onset. Results of previous ERP investigations of lexical access suggest that higher response strength associated with familiar word stimuli is more adaptive (Mills, Coffey-Corina, & Neville, 1993, 1994, 1997; Kuhl et al., 2013). Speed refers to the timing of maximum post-stimulus-onset neural firing associated with stimulus processing. Theoretically, individuals who exhibit faster neural responses associated with known words will be better at lexical access.

Approaches to ERP variable derivation. This study explored the relative validity and stability of several different putative neural measures of lexical access, derived from two separate approaches to ERP quantification: (a) the extant literature approach (ELA) and (b) the normative topographic approach (NTA). This multiplicity of potential variables from ERP procedures is common amongst ERP studies, and represents one of the major obstacles to using ERP procedures as scientific tool. That is, the many potential variables available from a single measurement period's data collection session reduce the falsifiability of hypothesis testing. This exploratory study is needed to pave the way for confirmatory uses of the most psychometrically sound ERP measure of lexical access.

The extant literature approach. When precedent empirical literature or theory provides an *a priori* hypothesis regarding the timing and location of the neural response associated with a cognitive process, investigators sometimes use such information to choose the post-stimulus time window and electrode that will serve as the focus of analysis (Handy, 2005). Several EEG studies of lexical access in children have found between-condition (i.e., known word versus unknown word, or word versus nonword) differences in the neural response between 200 and 500 ms after the onset of the stimulus (Mills, Coffey-Corina, & Neville, 1993, 1994, 1997; Mills, Plunkett, Prat, & Schafer, 2004). The one EEG study that compared lexical access in children with ASD to that of typically developing children built upon these findings by examining this time window exclusively (Kuhl et al., 2013). Thus, the time window between 200-500 ms post stimulus was chosen as the 'when' at which the ERP variables for the extant literature approach were derived.

Two scalp locations have emerged from previous ERP literature as potentially useful electrode sites for deriving ERP variables that index lexical access: the left lateralized temporal site (T3), and the left lateralized parietal site (P3). Kuhl and her colleagues (2013) found that the response strength of typically developing children differed significantly between conditions (i.e. known words versus unknown words) at T3. In children with ASD, the average amplitude response to known words at P3 predicted receptive language two and four years after the ERP. The electrode location of these results was consistent with those of previous ERP studies of lexical access in typically developing children (Mills, Coffey-Corina, & Neville, 1993, 1994, 1997; Mills, Plunkett, Prat, & Schafer, 2004).

Since we used an electrode map that was more dense (i.e., more electrodes per cm) than that used by Kuhl and her colleagues, and because electrodes around the max-difference electrode are almost always highly intercorrelated, we averaged the neural response strength values from the target electrodes (e.g. T3 and P3) with those from the spatially adjacent electrodes. This electrode cluster method for defining the 'where' for the extant literature approach was chosen because it was expected that an average of values from multiple spatially-proximal and highly correlated electrodes would be more stable than values taken from a single electrode. This procedure was thought to reduce the likelihood of type II error because measures with less measurement error tend to produce stronger associations and larger differences than measures with more measurement error. Thus, we derived a variable representing response strength and one representing speed from each of the two electrode clusters indicated by previous literature, T3 and P3. This yielded a total of 4 variables derived from the ELA.

Normative topographic approach (NTA). The topographical approach and the benefits of its use are described in detail in a tutorial by Murray, Brunet, and Michel (2008). The *normative* application of the topographical approach to examine speech processing is described by Yoder, Molfese, Murray, and Key (2013). Using data from the typically developing group, the time window of interest was identified via repeated significance testing of between-condition differences on two measures of neural activity that aggregate information across the electrodes: Global field power (GFP) and Dissimilarity (DISS). GFP is a measure of response strength associated with a particular stimulus, which is calculated within each time frame as the standard deviation of amplitudes across all electrodes relative to an averaged reference (Murray, Brunet, & Michel, 2008). DISS is a measure of the between-stimuli difference in the pattern of neural firing measured on the scalp (i.e., topography) averaged across electrodes. These tests of significance were conducted on all post-stimulus time samples between 60 and 600 ms. Sixty milliseconds was selected as the beginning of the analyzed epoch because ERP waveforms associated with the first 60 ms are thought to reflect strictly sensory processing (Naatanen & Winkler, 1999). Six hundred milliseconds was selected as the end of the analyzed epoch because the Kuhl et al. (2013) study suggests that lexical access is detected prior to 600 ms. To minimize Type I error associated with multiple significance testing, attention was restricted only to the time ranges in which significant between-condition differences occurred across 10 consecutive time samples (40 consecutive ms).

ERP variables. For each participant, the condition used to calculate each variable was the word condition. This condition was chosen based on methods used in extant literature and theory regarding which stimulus condition elicits lexical access.

For the ELA, response strength was quantified as the average amplitude at the target electrode cluster across the relevant time window. Average amplitude was calculated by averaging microvolt values across the time samples in the relevant time window and then averaging across the relevant electrodes surrounding the selected electrode within condition. This method was chosen because it is frequently used to index response strength in EEG experiments (Picton et al., 2000) and it is the metric that indexed response strength in the investigation by Kuhl and her colleagues (2013). Speed was quantified as the time frame of the minimum (e.g. most negative) amplitude at the target electrode cluster, within the specified time window (Handy, 2005). The minimum value was chosen because previous studies found that the word condition was associated with more negative voltage values than the nonword condition (Kuhl et al., 2013). Thus, the four variables derived from the ELA, representing response strength and speed at each of two specified scalp locations were, (a) average amplitude across 200 to 500 ms at the T3 electrode cluster (ELA T3 Avg Amp), (b) time frame of the minimum value between 200 and 500 ms at the T3 electrode cluster (ELA T3 Latency), (c) average amplitude across 200 to 500 ms at the P3 electrode cluster (ELA P3 Avg Amp), and (d) time frame of the minimum value between 200 and 500 ms at the P3 electrode cluster (ELA P3 Latency).

For the NTA, the evaluated measures of response strength and speed were (a) average GFP across each time window of interest and (b) time sample of maximum GFP within each time window of interest. GFP is the commonly-used metric for response strength in the topographic approach (Murray, Brunet, & Michel, 2008). The time sample of maximum GFP is much less frequently used than peak or average GFP, but was evaluated here as measure of speed that is analogous to the time sample of minimum amplitude, which is measure of speed used in the other approach. Maximum (not minimum) GFP was used because GFP only yields positive values, so higher values are indicative of greater response strength, which was presumed to be adaptive.

Measures of Theoretically Related Constructs

Receptive vocabulary. Raw scores from the receptive scale of the MCDI: Words and Gestures (Fenson et al., 2007) were used to index receptive vocabulary in all participants. This form of the MCDI (Fenson et al., 2007) is a checklist of 396 words that are commonly known by young children, which parents can use to indicate which words their child “says and understands” or “understands only”. These raw scores were summed to comprise “total words understood” for each participant. The words on the MCDI are reflective of vocabulary commonly known by typically developing children between the ages of 8 to 16 months. The authors recommend the use of the raw score as a metric for language in older developmentally delayed children. Investigations of the validity of the MCDI as a measure of receptive vocabulary have demonstrated that MCDI scores at age 2 are predictive of language skills at age 3 in typically developing children (Feldman et al., 2005) and those with ASD (Luyster, Qiu, Lopez, & Lord, 2007).

Relative to the measurement schedule of the larger study, the period at which the proposed study analyzed the MCDI scores varied by participant because the period at which the ERP procedure was conducted varied by participant. Concurrent analyses were conducted using MCDI scores taken at the same time as EEG data acquisition. Predictive analyses were conducted using MCDI scores taken four months after the ERP data acquisition. Typically developing participants completed a single assessment session, during which both the MCDI and the EEG data acquisition were completed.

Processable parent input. The variable chosen to reflect the construct of processable parent input (PPI) was meant to quantify the general extent to which a parent responded to a child's lead with linguistic information relevant to the child's current attentional or immediately preceding communicative focus. This variable was created by averaging the z scores of two similar variables, which were coded from videos of two parent-child interaction sessions. Measurement theory asserts that an average of scores taken from multiple contexts designed to measure the same construct is more likely to be representative of a person's true score than any score taken from a single assessment. This is because the aggregate score has averaged out some amount of random measurement error (Yoder & Symons, 2010). The two measurement contexts across which variables were aggregated were the Parent-Child Free Play (PCFP) and the Parent-Child Snack (PCS).

Parent-Child Free Play. The PCFP is a lab-created assessment context meant to capture a naturalistic interaction between parents and their children. At the start of this procedure, a researcher provided two toy sets and asked parents to "play as they would at home" with their children. After 10 min, the researcher returned and provided parents three age-appropriate books before leaving again. Again, parents were left to interact with the materials and their children as they were naturally inclined to. This portion of the sample lasted 5 min.

Parent-Child Snack. The PCS is a lab-created assessment context that is meant to mimic a common parent-child routine that naturally contains several opportunities for child requesting and parental response: snack time. During this procedure, the parent and child were seated at a table with a small cup, a pitcher of juice, and several single-bite cookies or crackers (or a parent-provided snack that accommodated preferences or dietary restrictions). Parents were told to interact with their children as they would at home. The procedure lasted 10 min.

Coding Procedures. Videos of the Parent-Child Free Play (PCFP) and the Parent-Child Snack (PCS) were coded using multiple passes with the aid of Procoder DV (Tapp, 2003). For the PCFP, coders used a 5-second momentary interval coding method to identify intervals during which both the parent and child were visible. Of these intervals in the PCFP and for all PCS intervals, a 5-sec partial interval method was used to identify intervals that contained at least one instance in which a child touched or looked at an object (i.e., attentional leads). These intervals were then flagged. For the PCS only, coders also used a 5-sec partial interval method to mark intervals during which children exhibited intentional communication (i.e., communicative leads). Intentional communication was defined as (a) use of nonconventional gestures, non-word vocalizations, or imitative symbols while coordinating attention between object and adult, (b) use of conventional gestures while attending to adult, or (c) use of spoken words or American Sign Language approximations. In the last pass, coders marked intervals that had already been flagged if the parent spoke about the object that served as the focus of the child's attention or put the child's presumed meaning into words. These were the intervals that contained PPI. A secondary coder recoded a random sample of 20% of all coded sessions. Interobserver reliability was documented using two-way random intraclass correlation coefficients (ICC). The ICCs reflecting interobserver reliability for variables derived from PCS and PCFP at Time 2 were both .98. The ICCs for variables from PCS and PCFP at Time 4 were both .97.

Variable derivation. Our goal in deriving a variable from multiple procedures across multiple time periods was to obtain scores that reflected to the fullest extent possible each parent's generalized tendency to provide linguistic input that was relevant to her child's communicative or attentional referent. To do this, we drew on the assertions of measurement theory that the average of scores from multiple contexts and time periods are expected to be more representative of a participant's generalized tendency or ability, as opposed to a single point estimate from one context at one period in time. For each participant, the number of 5-sec intervals coded as containing PPI were tallied for each procedure at each time period. Prior to aggregation, correlations between scores from PCFP and PCS were calculated to ensure that participant ranking from both scores met a threshold level of similarity, .40. This level was chosen *a priori*, given the assumption that scores taken from two different procedures that are purported to measure the same construct should be at least moderately correlated. The Pearson correlations between PPI scores from PCFP and PCS at time 2 and time 4 were .53 and .46, respectively. Scores from each procedure were then transformed into z-scores and averaged within each time period (times 2 and 4). Correlations between averaged z-scores were calculated to ensure the comparability of scores from two different time periods. Because this correlation exceeded our chosen threshold level of stability (r = .78), the aggregated scores from both time periods were averaged to create a single variable that reflected PPI for each participant dyad. When scores from either observation sample were missing for any participant at any time point, we used available scores to create the aggregate score for that participant. This was the case for two participants.

Attention to Child Directed Speech (ACDS; Watson, Baranek, Roberts, David & Perryman, 2010**)**. This construct was assessed at times 1 and 3 via a short procedure previously used in other studies (Watson et al., 2010, Watson, Roberts, Baranek, Mandulak & Dalton, 2012). During the procedure, the child was seated at a table facing a puppet theater and exposed to four conditions. First, children watched a three-minute age-appropriate music video. After a ten second break, children were exposed to a series of one-minute vignettes featuring child-directed speech: (a) a video of a woman reading a child's picture book, (b) a live puppet show presented by an adult female research assistant, and (c) a video of a woman playing with and describing a novel toy. In each vignette, speakers exhibited vocal qualities consistent with the characteristics of CDS (i.e exaggerated intonation, high pitch, and short utterance length). All stimuli were presented from the window of the puppet theater. Child behavior during this procedure was captured using a small video camera.

Coding and variable derivation. Videos of participants during the ACDS procedure were coded using timed-event behavior sampling in Procoder DV (Tapp, 2003). Child behavior during the portion of the procedure which featured CDS stimuli was coded as either (a) "looking" at CDS stimuli or (b) "not looking" at CDS stimuli. Scores were the proportion of seconds during which the child was looking at CDS stimuli divided by the total number of seconds during which CDS stimuli were presented. Interobserver reliability was documented using the same process as for PPI variables. The ICC for aCDS at Times 1 and 2 were .99 and 1, respectively. Following the same logic stated previously, scores from each time period were checked for comparability, and the resulting correlation exceeded the chosen threshold level (r = .58). Thus, scores were averaged across time 1 and 3 to create an aggregate score that was more likely to be representative of a child's general tendency to attend to CDS. When a child's aCDS score was not available for either time 1 or time 3, then only one value was used to calculate aCDS for that child. This was the case for eight participants.

Analysis

Research question 1: test-retest stability. Test-retest stability of both ERP measures was examined using data from a subset of typically developing participants (n=13, 22% of the

TD sample) that underwent two sessions of EEG data acquisition separated by less than 2 weeks. This question was tested using two types of analyses that emanated from generalizability theory: (a) the generalizability (G) study and (b) the decision (D) study. Generalizability theory uses the general linear model to parse variance associated with true score from that associated with measurement error. These estimates are then used to calculate the level of stability achieved with single measures and the average of scores from multiple measures (Brennan, 1992). The outcome of a generalizability study is an intraclass correlation coefficient (ICC), known as the G coefficient, reflecting the level of stability across the testing occasions already collected by the investigator. Assuming the same variance structure as was observed for two ERP sessions, D studies can project past the number of observed testing sessions to estimate the number of sessions needed to provide a stable ranking of participants on the across-session average of the ERP variable. To conduct the decision study, one must posit a desired threshold of stability one might reasonably expect. This is an arbitrary decision, but published benchmarks for ICC indicate that ICCs > .6 are ‘good’ (Mitchell, 1979). For this investigation, a G coefficient greater than or equal to 0.6 was chosen as the level of “acceptable” test-retest stability.

Research question 2: discriminative validity. For each ERP variable, discriminative validity was tested using independent t-tests, which compared ERP scores from the typically developing comparison group to those of the participants with ASD.

Research question 3: nomological validity. To examine the nomological construct validity of each ERP variable, relationships between each ERP variable and variables representing theoretically related constructs were statistically examined. In typically developing children, this included (a) receptive vocabulary. In children with ASD, this included (b) concurrent receptive vocabulary, (c) concurrent PPI, (d) concurrent aCDS, and (e) later receptive vocabulary. The effect size of these associations was estimated using Pearson's product moment correlations and the accompanying one-tailed t-tests were used to test the significance of the associations. The expected sign of each correlation varied depending on whether the aspect of neural firing was response strength or speed, and whether the variables were derived from the ELA (which yielded negative response strength values) or NTA (which yielded positive response strength values). Previous ERP investigations suggested that higher response strength (more negative values) to the word condition was more adaptive. Thus, for response strength variables derived from the ELA, negative associations with each theoretically related construct were anticipated. Since GFP values can only be positive, positive associations with theoretically related constructs were anticipated for all NTA-derived response strength variables. In regards to speed of lexical access, theory asserts that faster is better. Thus, we expected (and statistically tested for) negative correlations between all speed variables and theoretically related constructs of interest.

RESULTS

Preliminary Results for NTA Variable Derivation

No significant between-condition differences were found in the TD data for GFP, but two time ranges of significant between-condition difference in DISS were detected. The first time range spanned from 316 ms to 392 ms after stimulus onset. The second time range spanned from 444 ms to 600 ms after stimulus onset. These time ranges of interest were segmented according to topographic similarity, using the results of a hierarchical cluster analysis, which groups time samples with similar topographies within each condition. Topographical maps were combined if they were correlated at or above .9 for 10 consecutive time frames (40 ms), and then segmented according to several statistical error criteria. The optimal segmentation of the data into relevant topographic maps was chosen based on examination of three criteria: the Global Explained Variance (GEV), the Cross-Validation (CV) criterion, and the Krzanowski-Lai (KL) criterion. The nature of these values and their use for identifying the optimal segmentation is addressed by Murray, Brunet, and Michel (2008). The chosen segmentation was that with the fewest number of topographic maps, which (a) explained at least 80% of the variance (GEV>=.8), (b) was associated with a decreased CV value (compared to other segmentations), and (c) was associated with a peak in KL value (compared to other segmentations).

Segmentation revealed that each time range of interest was associated with a different topographical map, or “microstate”, which represented the composition of stable neural firing associated with the cognitive processes that are unique to lexical access. Figure 3 illustrates the segmentation of the word and nonword grand-average waves and relevant topographic maps for each condition within each group. Because two relevant time windows were identified by the NTA, the four variables derived from this approach were (a) average GFP from 316 to 392 ms

(NTA Avg GFP 1), (b) time sample of the maximum GFP between 316 and 392 ms (NTA Latency 1), (c) average GFP from 444 to 600 ms (NTA Avg GFP 2), and (d) time sample of the maximum GFP between 444 and 600 ms (NTA Latency 2).

Research Question 1. Test-retest stability

Results of the G and D studies examining the questions of test-retest stability are presented in Table 5.

1a.What is the ranking of test-retest stability of a single ERP data collection session? Across all examined ERP variables, G Coefficients ranged .04 to .71. Only one proposed lexical access variable reached the threshold level of stability for a single session, ELA T3 Avg Amp. Other measures of response strength, though lower, exhibited moderately high stability levels. The G Coefficients associated with NTA Avg GFP 2 and ELA P3 Avg Amp were .51 and .50, respectively. G Coefficients for variables reflecting speed of lexical access were generally lower, ranging from .04 to .22.

1b. How many ERP sessions is it necessary to average scores across to meet or exceed a threshold level of stability (i.e. 0.6)? When variables have high test-retest stability as a result of low error variance, fewer test session scores are needed to average out error in scores. Consequently, for variables that had the highest observed G coefficients, D study results indicated that fewer sessions are required to produce a stable estimate of the construct the ERP variable measures. For ELA T3 Avg Amp, scores from a single session met the criterion stability level. For both the NTA Avg GFP 2 and ELA P3 Avg Amp, an average of two sessions was needed to obtain scores with a threshold level of stability. Latency variables required averaging the most sessions for scores to become temporally stable. Averaging six ERP session scores were needed for the ELA T3 Latency variable and the NTA Latency 2 to reach criterion-level stability.

Thirteen were needed for the ELA T3 Latency. For NTA Latency 1, the observed retest stability was so low that the projected number of retest scores needed was likely an inaccurate estimate, which far exceeded the number of ERP sessions that would be feasible for any investigator to collect.

Research Question 2: Discriminative Validity

Effect sizes associated with the tests of mean differences between participants with ASD and age-matched TD controls for each ERP variable are presented in Table 5. Significant and large effect sizes were observed for both NTA Avg GFP 1 and NTA Avg GFP 2. However, the direction of these effect sizes suggested that participants with ASD exhibited, on average, more adaptive (higher-powered) scores. These results depart from theory, which suggests that children with ASD, on average, exhibit lower-response strength and slower responses to familiar word stimuli than TD children. Although, theoretically predictable algebraic signs for the difference were observed for two variables: ELA T3 Latency and NTA Latency 2, these differences were very small ($d = 0.029$ and 0.048, respective) and their confidence intervals included zero. Thus, there is no evidence suggesting that any of the examined lexical access variables can predictably discriminate between diagnostically separate groups.

Research Question 3: Nomological Validity

Correlations indicating the observed associations between each ERP variable and scores of the theoretically related constructs are presented in Table 5.

3a. In typically developing children, what is the association between the ERP variable and concurrently-measured receptive vocabulary? Concurrent receptive vocabulary scores were available for all 60 TD participants. However, nearly half of the TD participants were assigned the maximum raw receptive vocabulary score possible on the MCDI. Thus, to control for ceiling effects, participants with maximum receptive vocabulary scores were removed from the analysis. Scores for the remaining 32 participants were analyzed for associations. Only one significant association was observed between an ERP variable measuring lexical access and scores of receptive vocabulary in the group of age-matched TD children, and this was for ELA T3 Avg Amp (*r* =-.428, *p* = .014). As we predicted, the direction of this association was negative. Thus, the expected relation between receptive language and lexical access was only observed in one putative measure of lexical access, out of the eight that were examined.

3b. In children with ASD, what is the association between the ERP variable and concurrent receptive vocabulary? Concurrent receptive vocabulary scores were available for all 34 participants with ASD. Only one significant correlation was observed between the examined ERP variables and concurrent scores of receptive vocabulary in children with ASD, and this was for NTA Latency 1 (*r* = -.334, *p* = .026). The expected direction of this association was negative. Thus, the expected relation between receptive language and lexical access in children with ASD was only observed in one of the eight examined ERP variables.

3c. In children with ASD, what is the association between the ERP variable and concurrent processable parent input? Concurrent PPI scores were available for 31 of the participants with ASD. None of the observed correlations between putative ERP measures of lexical access and scores of PPI were significant. The strongest correlation was observed for ELA T3 Avg Amp (*r* =-0.275, *p* = 0.06), which was in the theoretically predicted direction. Only one of the correlations associated with measures of latency was even in the theoretically predictable direction, ELA T3 Latency (*r* = -0.140, *p* =.22), and this association was nonsignificant.

3d. In children with ASD, what is the association between the ERP variable and concurrent attention to child directed speech? Scores of aCDS were available for 29 participants with ASD. Only one of the correlations reflecting the relationship between putative ERP measures of lexical access and aCDS in children with ASD was significant and in the expected direction, and this was for ELA P3 Avg Amp (r =-0.353, p = 0.03). Of the four measures of latency, only one correlation was in the expected direction, but it was nonsignificant (NTA latency 1, r = -0.177, p = 0.18). Thus, the expected relation between lexical access and aCDS was only observed in one putative ERP measure of lexical access.

3e. In children with ASD, what is the association between the ERP variable and later receptive vocabulary? Scores of receptive vocabulary collected four months after the ERP data collection were only available for 23 of the participants with ASD. None of the correlations reflecting the association between ERP measures of lexical access and scores of later receptive vocabulary in children with ASD were significant. Two ERP measures were associated with later receptive vocabulary in the expected direction and exceeded .3 in absolute value, NTA Avg GFP 1 (r = 0.330, p = 0.06) and ELA T3 Avg Amp (r = 0.315, p = 0.07). However, these correlations were not significant. This was likely a consequence of the lower number of scores available to test this question. The remaining measures of response strength were associated with correlations of slightly smaller magnitude in the expected directions. Only one of the measures of latency, NTA Latency 1, was associated with a correlation in the expected direction (r =-0.211, p = 0.17), but it was nonsignificant. Thus, the expected association between lexical access and later receptive language in children with ASD was not reliably observed in ERP measures of lexical access.

DISCUSSION

Although it is possible that the ERP variables examined here are valid measures of lexical access for other populations and other purposes, the current investigation, which admittedly featured a small sample, did not yield unconditional evidence of their nomological or discriminative validity as measures of lexical access in young children in the early stages of language development. In general, all of the ERP measures failed to discriminate between typically developing children and those with autism in a theoretically predictable manner. In addition, most of the tests of the associations between each variable and theoretically related constructs yielded small correlations that were not statistically significant. Those correlations that were significant were distributed across different measures. Thus, significant results did not converge to highlight any single ERP variable as a valid measure of lexical access.

Though most were not significant and small, directionally predictable associations were observed for all tested relations for two ERP variables, ELA T3 Avg Amp and ELA P3 Avg Amp. Correlations reflecting associations with related constructs ranged in absolute value from 0.143 to 0.368 for ELA T3 Avg Amp, and 0.116 to 0.353 for ELA P3 Avg Amp. In addition, both of these variables had relatively high test-retest stability. The G Coefficients for ELA T3 Avg Amp and ELA P3 Avg Amp were 0.71 and 0.50, respectively. These results, together with significant associations found for other investigations, suggest that these two variables might accrue sufficient validity support in larger studies. However, our failure to detect statistically significant effect sizes for most associations between theoretically-related constructs and these ERP variables limits the extent to which we can assert their validity.

Alternative Explanations

It is possible that some of the examined ERP variables with poor validity evidence do have scientific utility, and that alternative explanations account for our failure to confirm hypotheses regarding predicted relations. Several alternative explanations exist. First, the number of participants for which data was available to test the research questions was insufficient to detect small but significant associations. In fact, in order to detect a correlation that would be deemed large by Cohen's 1988 guidelines (0.50) with 95% confidence, a sample size of 34 would be required. Though we had ERP data for this number of participants, less data was available for measures of receptive language, PPI, and aCDS. Thus, for several research questions, we had insufficient sample size with present data available to detect even large associations with 95% confidence. Second, it is possible that the constructs examined in the current study are strongly related, but that those relations are not linear. If this were the case, then observed Pearson's *r* values (that are meant to quantify linear relations) would be diminished. Third, the relation between the tested constructs may vary according to the level of a third unexamined variable. In other words, unexamined third variables may have a statistical interaction with the ERP variable predicting the behavioral measure, thereby limiting the magnitude of the observed associations between the ERP variables and the theoretically related constructs when the 3rd variable is ignored.

Finally, there is reason to believe that our approaches to variable derivation led us to the "wrong" temporal windows and scalp locations, despite our efforts to do otherwise. Although there were topographical differences between conditions in the TD sample, between-condition differences were not observed for average amplitudes at T3 or P3 for typically developing children or those with autism. T3 and P3 were selected on the basis of the scalp region at which

Kuhl and colleagues (2013) detected between-condition differences in her TD and ASD samples. However, these results may have been sample specific. Alternatively, it is possible that the samples studied by Kuhl and her colleagues (2013) were not similar enough to the samples used in the current investigation to warrant replication attempts. Both of the samples featured in the current investigation were, on average, nearly two years older than those investigated by Kuhl and her colleagues (2013). To examine whether a study sample difference in chronological age was a factor that contributed to our failure to replicate significant between condition differences at T3 and P3 electrode clusters, we calculated correlations between chronological age in months and average amplitudes at T3 and P3 for both TD and ASD samples. We did not detect associations of significant magnitude between these two variables in the ASD sample. However, in the TD sample, a significant correlation between age and T3 Avg Amp was detected (r = -0.29, p =.011). In addition, the ASD sample featured in the current investigation was more severely impaired than that of Kuhl and her colleagues (2013), both in terms of cognitive ability and symptom severity. To investigate whether these factors contributed to our failure to replicate the results of previous literature, we examined correlations between relevant ERP variables and ADOS diagnostic scores, as well as those with composite standard scores from the Mullen Scales of Early Learning (Mullen, 1995). No significant correlations were detected. However, there was a correlation of notable magnitude between P3 Avg Amp and Mullen Composite scores (r =-.25, p =.07). Thus, it is possible that the timing and/or location of the signal changes as children age and develop, as is the case for other cognitive processes associated with language processing (Courchesne, 1978; Courschesne, 1990). If this is the case, sample differences across studies in age and cognitive ability may have contributed to our failure to observe significant between-condition differences for the T3 and P3, because the time windows and scalp locations chosen based on previous literature were too broad and imprecise to exclusively capture the neural signal associated with lexical access in our sample.

On the other hand, our difficulty identifying the timing and location of the signal associated with lexical access within the ASD group could be the result of different compensatory neural activity exhibited by children with ASD. It is possible that the timing and location of the neural activity associated with lexical access in children with ASD differs greatly from that of TD children. If this is the case, a normative approach will not be useful in identifying a good measure of lexical access in children with ASD. Moreover, it is possible that each child with ASD exhibits different and individualized compensatory neural activity. If the neural activity and corresponding ERP signal associated with lexical access differs in each child with ASD, identifying a good measure of lexical access for this population will require a different approach than was taken in the current study. For example, an approach that uses trials as the basis for degrees of freedom when testing the between condition differences in response strength, topography, or latency within each participant might be useful.

Strengths of the Investigation

Despite these limited results, there are several strengths of this investigation. First, it employed the use of a novel approach to ERP variable derivation, the NTA, which was born of the desire to use a theoretically and mathematically sound approach to ERP analysis that would be more likely to generate replicable results. Surprisingly, variables derived by the NTA were not found to have psychometric properties that were superior to ELA-derived response strength variables. However, it should be noted that the derivation of these two ELA-derived variables *depended* on results found in previous investigations. If relevant previous literature on lexical access had not been available, we might have used a more traditional approach to variable derivation, one that depended on detecting significant differences at an electrode site. Because the average amplitudes at T3 and P3 across 200 to 500 ms did not significantly differ between word and nonword conditions in either group, these variables would not have been derived and tested for validity. Though no significant between-condition differences were found for response strength for either approach (ELA or NTA), significant between-condition differences were found for DISS in TD participants. The time windows of significant between-condition DISS overlap with the time window used to derive ELA variables. This suggests that differential topography associated with familiar words versus nonwords is responsible for the utility of the variables derived from ELA. These results provide insight into ways that the NTA can be combined with more traditional ERP analysis methods to enhance the strength of observed evidence, and contribute to theory for future investigations.

The second strength of this investigation is that it involved an in-depth examination of ERP variables on several aspects of scientific utility, which included test-retest reliability and multiple types of validity. Few investigations examine and compare ERP variables on even a single aspect of scientific utility. Such investigations are necessary to identify variables that are likely to yield replicable results. Replication of ERP results is needed if ERP investigations are to substantially contribute to knowledge about cognitive processes of clinical populations, which can inform future intervention practices.

Finally, the results of this investigation indicated the extent to which ERP variables of lexical access yield scores that are stable over short periods of time, and provide guidance regarding the number of retest scores needed for averaging to reach an acceptable level of stability. In the case of ELA T3 Avg Amp, NTA GFP 2, and ELA P3 Avg Amp, only one or an average of scores from two sessions are needed to obtain a score set that is stable and thereby more likely to be scientifically useful than the less stable ERP variables. Investigators can use these results to plan the number of ERP measurement sessions needed to obtain a stable set of scores. More reliability investigations like this one are needed for such guidance regarding other commonly used ERP measures of language processing.

Future Investigations

Future investigations should build on the current results and test the extent to which a combination of traditional and novel approaches to ERP analysis can yield better measures of lexical access in young children. For example, the topographic maps of segments associated with significant between-condition differences in DISS could be used to visually identify electrode clusters associated with maximum difference, and the time windows identified by the approach could function as the time windows across which amplitudes might be averaged. Visual inspection of the topographic maps associated with word and nonword conditions in each sample (Figure 3) suggests that maximal between-condition differences may have been found at electrode clusters associated with P4 and F3 scalp locations (see Figure 1). Variables derived from amplitude values at these electrode clusters that are averaged across the time windows associated with significant DISS might serve as more valid measures of lexical access.

Additional work could be done in future investigations to examine the alternative explanations for diminished associations observed in the current investigation. This might entail developing and testing theories that include nonlinear relations between cognitive processes and related constructs, or statistically testing interactions with related third variables. These questions should be tested with a larger sample size, to ensure adequate power to detect the statistical significance of an observed effect. Efforts to ensure the sample is more closely matched to those used by previous investigators may also increase the likelihood of replication regarding timing and location of the neural signal associated with lexical access. Such work may contribute to a more extensive documentation of the ERP signal associated with lexical access, one which more precisely maps the timing and location of the signal, as well as the extent to which it changes with chronological age and developmental level. This would facilitate more accurate lexical access variable derivation for both typical and clinical populations, paving the way for more in depth investigations of the development of lexical access and its role in language acquisition. Such investigations are important, because they may highlight intervention characteristics associated with improved lexical access, and therefore, intervention approaches that may improve lexical access and by extension receptive language. This is needed if ERP investigations of lexical access are to ever be useful in the field of Special Education.

Conclusion

The results of the current investigation add to the evidence suggesting that these putative ERP measures of lexical access are limited in their scientific utility. Though ERP methods have the potential to illuminate new information about cognitive processes that are difficult to observe through behavior, they are inhibited by error variance that may be attributable to several factors. If ERPs are ever to be clinically useful, such that they highlight potential moderators and mediators of change, or identify subgroups of disorders that were previously categorized as a single group, then more must be done to identify and reduce these sources of noise. More work is needed to refine the science of ERP investigations of clinical populations.

APPENDIX

Table 1.

List of Constructs, Measures, and Corresponding Research Questions

Construct	Procedure	Variable derivation Approach	Variable	Research Questions
LA Speed	ERP	ELA	Time sample of maximum amplitude within the relevant time window and electrode cluster	All
LA Speed	ERP	NTA	Time sample of maximum GFP within relevant time window	All
LA Response Strength	ERP	ELA	Average amplitude across relevant electrode cluster and time window	All
LA Response Strength	ERP	NTA	Average GFP across the time window of interest	All
Receptive Vocabulary	MCDI	Parent report	Sum of words "understood only" and "said" reported by parent	3a, 3b, 3e
Processable parent Input	PCFP & PCS	Direct observation and interval coding	Number of 5 sec intervals containing processable parent input, transformed into z-scores	3c
aCDS	CDS assessment procedure	Direct observation an timed event coding	Proportion of seconds looking at CDS stimuli out of total seconds CDS stimuli presented	3d

Note. aCDS= Attention to Child Directed Speech, CDS= Child Directed Speech, ELA = Extant Literature Approach, ERP= Event Related Potentials, GAvg = Grand Average Approach, LA=Lexical Access, MCDI = MacArthur-Bates Communicative Development Inventory, NTA= Normative Topographic Approach, PCFP = Parent-Child Free Play, PCS=Parent-Child Snack.

Table 2

Measures Relevant to the Current Investigation and Time Period of Administration in the Larger Study

Measure	Time Periods
Receptive Language	
MCDI	All 5
Processable Parent Input	
Parent-Child Snack	2, 4
Parent-Child Free Play	2, 4
Attention to Child Directed Speech	1, 3
EEG data	Varied for each participant

Note. MCDI=MacArthur-Bates Communicative Development Inventory.

Table 3

Demographic Information for Participants

Measure	ASD	TD
Number of Participants	34	60
Males	27	26
Females	7***	34***
Race[a]		
Black	9	16
Asian	0	2
Caucasian	25	46
Hispanic	0	2
More than 1 race	0	6
Mean Age in Months at the ERP session	45.40 (9.57)	43.19 (12.19)
Mean MCDI at the ERP session		
Expressive Raw Score	53.38 (86.35)***	342 (99.81)***
Receptive Raw Score	145.20 (128.47)***	367.83 (58.08)***
Level of Education of Primary Caregiver[b]	6 (1.57)***	7.31 (1.28)***
	5.14 (14.16)	n/a
Mean ADOS Score (SD) at Time 1	23.63 (3.69)	n/a
Mullen at Time 1		
Standard Score	51 (4.95)	n/a
Age Equivalency (mos)	11.16 (5.12)	n/a
Expressive Language Age(mos)	7.79 (3.72)	n/a
Receptive Language Age (mos*)*	5.75 (7.52)	n/a

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

[a] Participants that self-identified as more than 1 race were counted in tallies for each racial demographic they identified as.

[b] Taken from a demographic questionnaire

* Test statistic testing differences between group has *p-value* <.05

***Test statistic testing differences between group has *p-value* <.001

Table 4

Word and Nonword Stimuli Presented During EEG Data Acquisition

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

Note. * denotes 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 sound (nonwords)

Table 5

Table of Results

Measure	Test-retest stability in TD participants		Cohen's *d* for lexical access difference between TD and ASD	Correlation with receptive vocabulary in TD participants	Correlation with concurrent receptive vocabulary in ASD participants	Correlation with Parent Processable Input in ASD participants	Correlation with aCDS in ASD Participants	Correlation with later receptive vocabulary in ASD participants
	G Coeff.	D Study						
ELA T3 Avg Amp	0.71	1	-0.408	-0.428*	-0.206	-0.275	-0.143	-0.315
ELA T3 Latency	0.22	6	0.029	0.353	0.052	-0.140	0.036	0.151
ELA P3 Avg Amp	0.50	2	-0.224	-0.034	-0.221	-0.199	-0.353*	-0.207
ELA P3 Latency	0.11	13	-0.063	0.260	0.224	0.212	0.301	0.426
NTA Avg GFP 1	0.39	3	-0.494*	-0.205	0.215	0.190	-0.064	0.214
NTA Latency 1	0.04	NA	-0.174	0.078	-0.334*	0.037	-0.177	-0.211
NTA Avg GFP 2	0.51	2	-0.431*	-0.207	0.227	0.195	-0.232	0.330
NTA Latency 2	0.20	6	0.048	0.341	0.091	0.011	0.0224	0.190

Note. ASD=Autism Spectrum Disorder; Avg Amp=Average Amplitude; ELA=Extant Literature Approach; GFP=Global Field Power, NTA=Normative Topographic Approach; TD=Typically Developing.Negative values of Cohen's *d* reflect means that favor the ASD group. Correlations reported as pearson's *r*. G Coefficients are reported as intraclass correlations. $*p < .05$ $**p < .01$

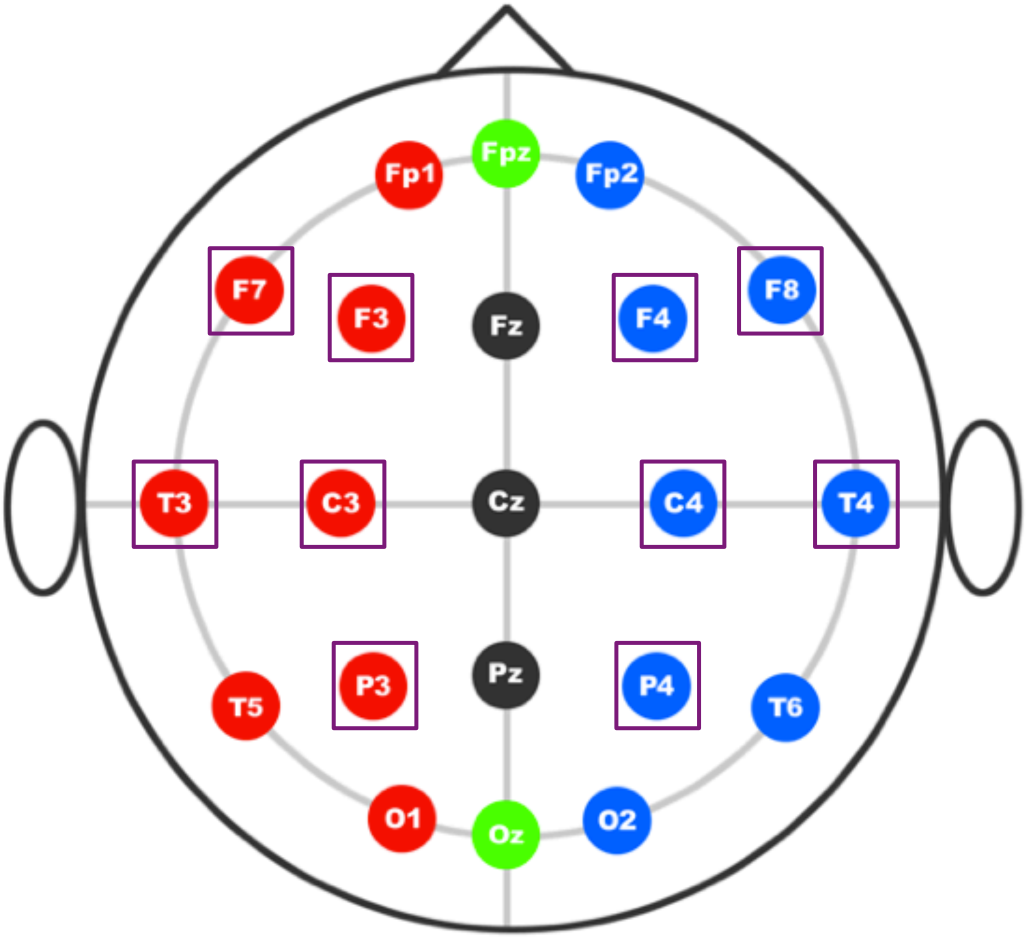

Figure 1. Illustration of the 10-20 electrode placement system. Square boxes indicate electrodes analyzed by Kuhl and colleagues (2013).

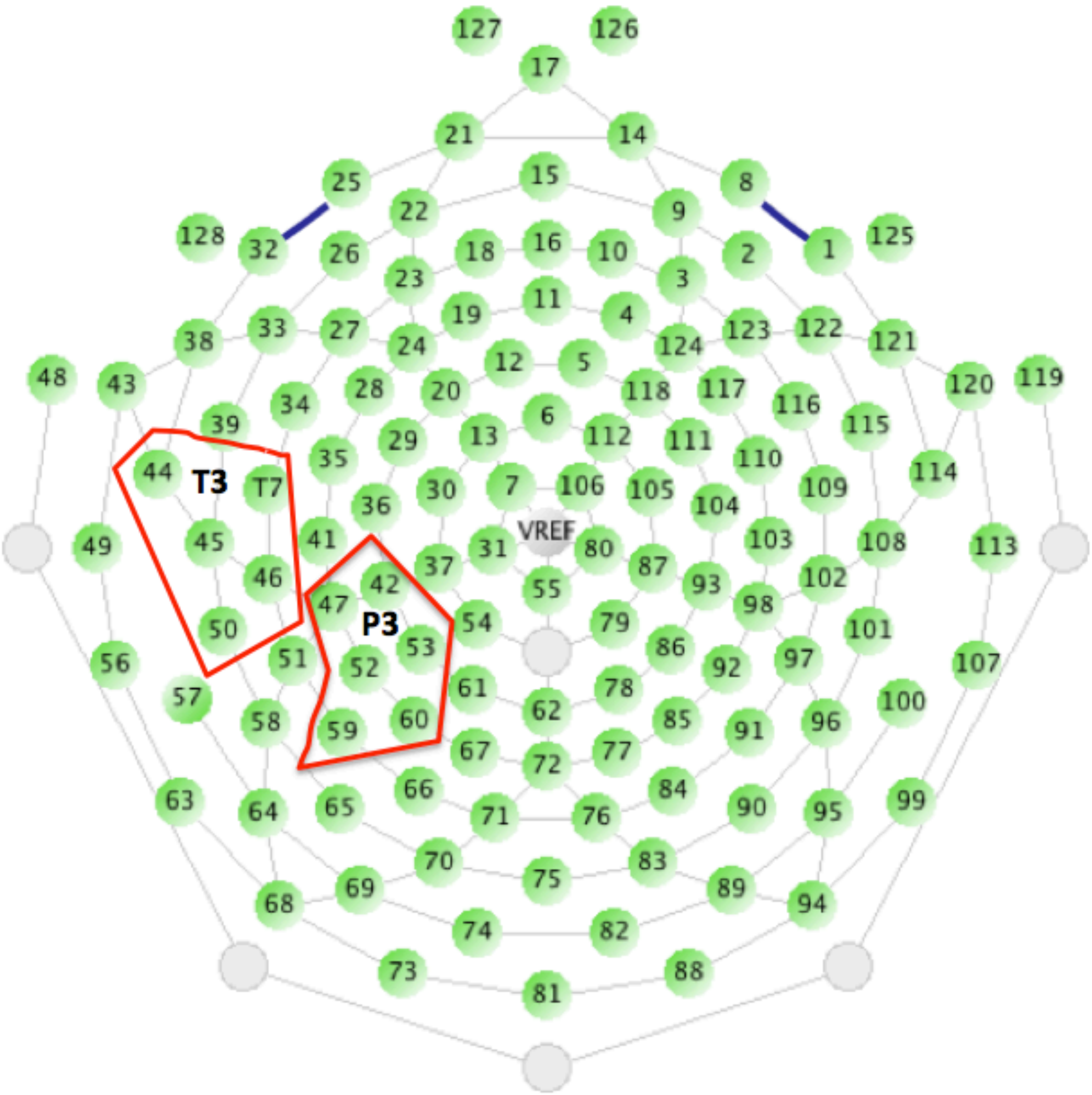

Figure 2. Map of the electrode layout of the EGI net used to acquire EEG data. Electrode clusters corresponding to T3 and P3 are indicated.

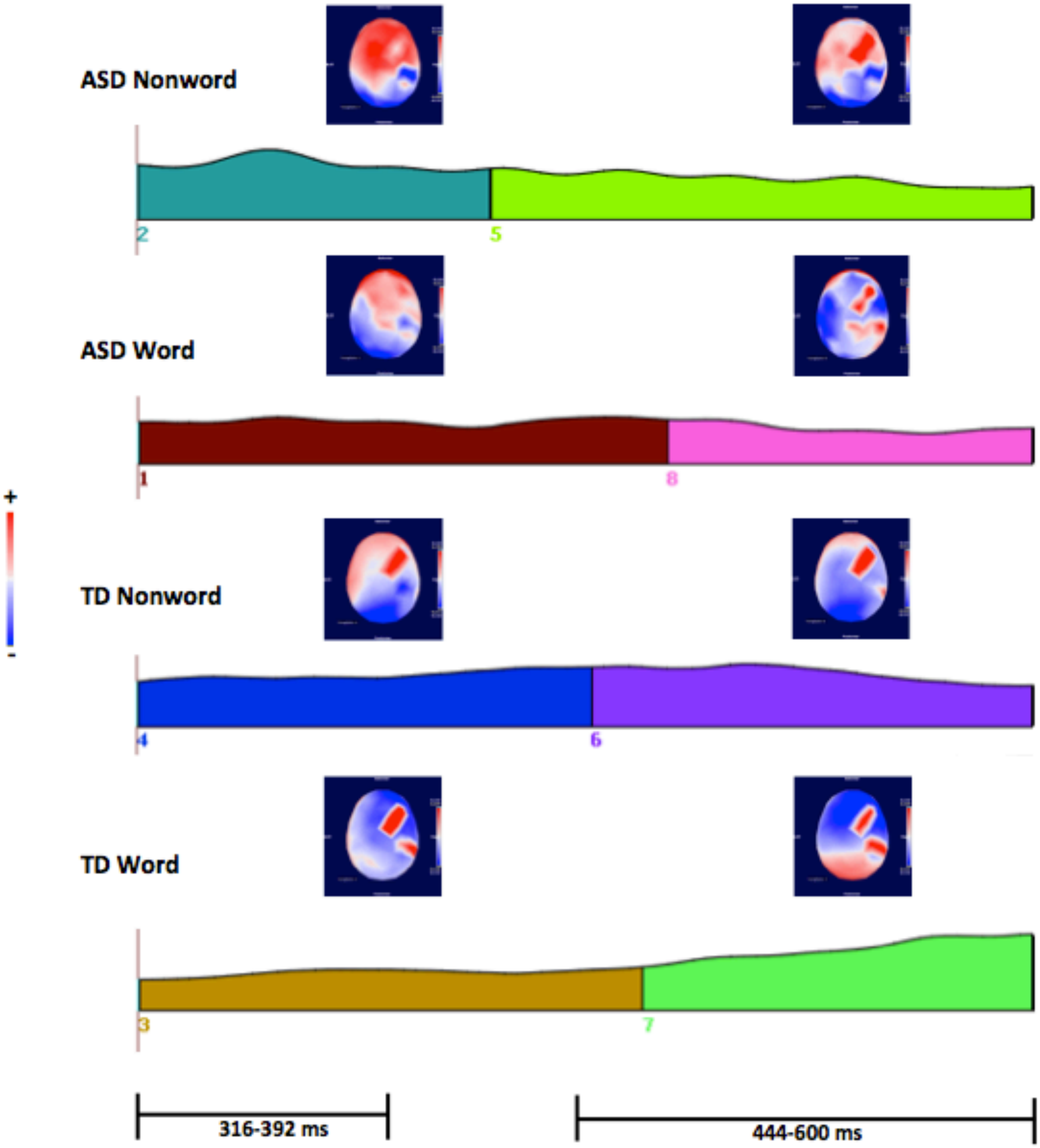

Figure 3. Microstate segmentation and corresponding template maps for the conditions of word and nonword. Microstate segments are graphed under GFP curves and labeled with differentially colored and numbered maps. Periods of significant between condition DISS found for TD participants is marked.

REFERENCES

American Psychiatry Association. (2000). Diagnostic and statistical manual of mental disorders DSM-IV-TR. *Washington, DC: American Psychiatry Association*, *371*.

Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, *11*, 27-34.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*, 477-488.

Courchesne, E. (1990). Chronology of postnatal human brain development: Event-related potentials, positron emission tomography, myelogenesis, and synaptogenesis studies. In: J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-Related brain potentials: Basic issues and applications (*pp. 210-241). New York: Oxford University Press.

Courchesne, E. (1978). Neurophysiological correlates of cognitive development: Changes in long latency event-related potentials from childhood to adulthood. *Electroencephalography and Clinical Neurophysiology, 45*, 468-482.

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

DeMyer, M. K., Barton, S., DeMyer, W. E., Norton, J. A., Allen, J., & Steele, R. (1973). Prognosis in autism: A follow-up study. *Journal of Autism and Childhood Schizophrenia*, *3*, 199-246.

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., Bolton, P., Charman, T., Baird, G., & Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype: disengagement and facilitation. *Journal of Child Psychology and Psychiatry*, *50*, 637-642.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, *76*, 856-868.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates Communicative Development Inventories*. Paul H. Brookes Publishing Company.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*, 103-114.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181-195.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*, 228-231.

Gillberg, C. (1991). Outcome in autism and autistic-like conditions. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 375–382.

Gillberg, C., & Steffenburg, S. (1987). Outcome and prognostic factors in infantile autism and similar conditions: A population-based study of 46 cases followed through puberty. *Journal of Autism and Developmental Disorders, 17*, 273 - 287.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology*, *24*, 14.

Handy, T. C. (Ed.). (2005). *Event-related potentials: A methods handbook*. MIT press.

Halle, P., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development, 17*,119–129.

Halle, P., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development, 19*, 463–481.

Howlin, P., Mawhood, L. & Rutter, M. (2000). Autism and developmental receptive language disorder – a follow-up comparison in early adult life: Social, behavioral, and psychiatric outcomes. *Journal of Child Psychology and Psychiatry, 41*, 561–578.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1–23.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry*, *45*, 1115-1122.

Liss, M., Harel, B., Fein, D., Allen, D., Dunn M., Feinstein, C., et al. (2001). Predictors and correlates of adaptive functioning in children with developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 219–230.

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. Handy (Ed.), *Event-related potentials: A methods handbook*, (pp. 17-32). MIT press.

Klinger, L., Dawson, G., & Renner, P. (2002). Autistic Disorder. In E. Mash & R. Barkley (Eds.), *Child psychopathology,* 2nd edition, (pp. 409-454). New York: Guilford Press.

Kobayashi, R., Murata, T., & Yoshinaga, K. (1992). A follow-up study of 201 children with autism in Kyushu and Yamaguchi areas, Japan. *Journal of Autism and Developmental Disorders, 22*, 395-411.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*, 233-242.

Kuhl, P. K., Coffey‐Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science*, *8*, F1-F12.

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.

Lachenmeyer, C. (1971). *The language of sociology.* New York: Columbia University Press.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A. & 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.

Lord, C. & Paul, R. (1997). Language and communication in autism. *Handbook of autism and pervasive developmental disorders,* 2, 195-225.

Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur–Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research*, *50*, 667-681.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15*, 28–50. doi:10.1177/1088868310366253

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: Guilford Publications.

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: 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.

Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin*, *86*, 376.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topography*, *20*, 249-264.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826.

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., & Cadusch, P. J. (1997). EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and clinical neurophysiology*, *103*, 499-515.

Nunnally, J.C. (1978). Psychometric Theory. McGraw-Hill. New York, NY.

Paul, R., & Cohen, D. J. (1984). Outcomes of severe disorders of language acquisition. *Journal of Autism and Developmental Disorders*, *14*, 405-421.

Paul, R., Chawarska, K., Fowler, C., Cicchetti, D., & Volkmar, F. (2007). “Listen my children and you shall hear”: Auditory preferences in toddlers with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *50*, 1350-1364.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, *15*, 325-345.

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.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Scott, J., Clark, C., & Brady, M. (2000). *Students with autism*. Thomson Wadsworth, Canada.

Sherrod, K. B., Crawley, S., Petersen, G., & Bennett, P. (1978). Maternal language to prelinguistic infants: Semantic aspects. *Infant Behavior and Development*, 1, 335-345.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 549-565.

Stets, M., Stahl, D., & Reid, V. M. (2012). A meta-analysis investigating factors underlying attrition rates in infant ERP studies. *Developmental Neuropsychology*, *37*, 226-252.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition,76*, 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word–form representations of 14-month-olds. *Psychological Science, 13*, 480–484.

Tapp, J. (2003). *Procoder for digital video* [computer software]. Nashville, TN: Vanderbilt Kennedy Center.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57,* 1454-1463.

Venker, C. E., Eernisse, E. R., Saffran, J. R., & Weismer, S. E. (2013). Individual Differences in the Real‐Time Comprehension of Children with ASD. *Autism Research*, *6*, 417-432.

Venter, A., Lord, C., & Schopler, E. (1992). A follow‐up study of high‐functioning autistic children. *Journal of Child Psychology and Psychiatry*, *33*, 489-597.

Watson, L. R., Baranek, G. T., Roberts, J. E., David, F. J., & Perryman, T. Y. (2010). Behavioral and physiological responses to child-directed speech as predictors of communication outcomes in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *53*, 1052-1064.

Yoder, P. J., Molfese, D., Murray, M. M., & Key, A. P. (2013). Normative topographic ERP analyses of speed of speech processing and grammar before and after grammatical treatment. *Developmental Neuropsychology*, *38*, 514-533.

Yoder, P. J., & Symons, F. J. (2010). *Observational measurement of behavior*. Springer Publishing Co.

Yoder, P., Watson, L. R., & Lambert, W. (2014). Value-Added Predictors of Expressive and Receptive Language Growth in Initially Nonverbal Preschoolers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 1-17.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale* (Vol. 3). San Antonio, TX: Psychological Corporation.