CERTAINTY RATINGS OF LEXICAL KNOWLEDGE ON VOCABULARY CHECKLISTS FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

By

Alexandra Hanania

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Communicative Sciences and Disorders -- Master of Arts

ABSTRACT

CERTAINTY RATINGS OF LEXICAL KNOWLEDGE ON VOCABULARY CHECKLISTS FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

By

Alexandra Hanania

Children with autism spectrum disorder (ASD) display language deficits, making it difficult to assess linguistic knowledge using standardized assessments. Parent report measures, such as vocabulary checklists, are commonly used across settings. While current parent measures do not capture data regarding parents' certainty about their judgments, this insight may be informative. The current study examined the use of a custom parent vocabulary checklist to assess parent certainty of their responses regarding their child's vocabulary knowledge. The participants were children with ASD and typically developing children (TD) of similar ages. In addition to answering whether their child says and understands a word, understands, or neither, parents were also asked to rate the certainty of their judgment about each word, with response choices ranging from Very Uncertain to Very Certain. Parents of TD children reported higher levels of certainty than parents of children with ASD. Within the ASD group, certainty ratings were higher for words classified as "Understands and Says" than for words classified as "Understands" or "Neither". Certainty ratings varied substantially across children with ASD. This thesis is dedicated to Dad. Thank you for being my biggest supporter.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Dr. Courtney Venker, for all of your support throughout this entire process. I would not have been able to accomplish and complete this project without your words of encouragement and insight. Thank you for helping me grow as a student and future clinician. I would also like to thank all of the members of the Lingo Lab at Michigan State University, who helped collect and analyze the data for this project. Your dedication and effort in the lab, both with participants and behind the scenes, made this project possible. Additionally, thank you to the members of my thesis committee, Dr. Bridget Walsh and Dr. Sarah Douglas, for you feedback and support.

I would like to thank my Mom for providing me with encouragement and support throughout my entire graduate school experience. Your constant belief in me allowed me to persevere throughout the most difficult period of my life. Lastly, I would like to thank my Dad. Although you were not able to see my completed thesis, I know you are always with me. I hope this project makes you proud.

TABLE OF CONTENTS

LIST OF FIGURES	vi
INTRODUCTION	1
ASD DIAGNOSIS AND CHARACTERISTICS	2
STRUCTURAL LANGUAGE IN ASD	2
VOCABULARY SKILLS	3
EXPRESSIVE AND RECEPTIVE LANGUAGE	4
ASSESSING RECEPTIVE AND EXPRESSIVE VOCABULARY IN	
ASD	7
STANDARDIZED, EXAMINER-DIRECTED ASSESSMENTS	7
LANGUAGE SAMPLES	9
PARENT VOCABULARY REPORTS	9
PSYCHOMETRIC PROPERTIES OF PARENT REPORTS	11
GOALS OF THE PRESENT STUDY	17
METHOD	19
EXAMINER-DIRECTED ASSESSMENTS	20
MACARTHUR-BATES COMMUNICATIVE DEVELOPMENT	
INVENTORIES	20
CUSTOM VOCABULARY CHECKLIST	20
ANALYSIS PLAN	21
RESULT 22	
DISCUSSION	
AIM 1	
AIM 2	
AIM 3	
LIMITATIONS	
FUTURE RESEARCH	
CONCLUSION	41
APPENDIX	42
BIBLIOGRAPHY	48

LIST OF FIGURES

Figure 1. This figure illustrates the mean certainty response for each vocabulary word	23
Figure 2. This figure illustrates the mean certainty response for each parent	24
Figure 3. Parent certainty ratings for words children were reported to say	25
Figure 4. Parent certainty ratings for words children were reported to understand (but not say)	26
Figure 5. Parent certainty ratings for words children were reported to neither say nor understand	26
Figure 6. Relationship between parent mean certainty ratings and child chronological age	27
Figure 7. Relationship between parent mean certainty ratings and child autism severity level from the ADOS-2	28
Figure 8. Relationship between parent mean certainty ratings and child receptive language growth scale value from the PLS-5	28
Figure 9. Relationship between parent mean certainty ratings and child expressive language growth scale value from the PLS-5	29
Figure 10. Relationship between parent mean certainty ratings and parent number of years of education	29
Figure 11. The above figure demonstrates the correlation between the mean certainty across each individual parent's ratings on the custom vocabulary checklist and their consistency/reliability of their responses across the two checklists	31

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder associated with social skill deficits and restricted and repetitive behaviors and interests (American Psychiatric Association, 2013). Although language deficits are not currently part of the DSM-5 diagnostic criteria for ASD (American Psychiatric Association, 2013), they are frequently associated with impairments in language skills, including receptive and expressive vocabulary. Standardized, examiner-directed assessments are one of the most commonly used types of language assessments. However, because children with ASD exhibit complex behaviors and language profiles, assessing their language skills presents many challenges when using traditional, standardized language assessments. For this reason, parent report measures are a frequently used method of evaluating language in a way that addresses the concerns of standardized assessments, especially in children who are difficult to assess.

In the form of a checklist, parent vocabulary reports ask parents or caregivers whether a child produces and/or understands certain vocabulary words. While these assessments have numerous advantages, there are some concerns over their validity and reliability (Brady et al., 2014; DiStefano & Kasari, 2016; Law & Roy, 2008; Muller & Brady, 2016; Tomasello & Mervis, 1994). For example, parents may have a difficult time determining whether their child truly understands or spontaneously produces a certain word. However, there is limited research on vocabulary checklists for children with ASD. Additionally, there is no research investigating parents' level of certainty of their responses regarding their child's knowledge of a given term, which may be a valuable source of information about emerging vocabulary skills. The current study examines the use of a newly developed, custom parent vocabulary checklist that asks

parents not only whether their child produces and/or understands certain words, but also how certain they are about their response.

ASD DIAGNOSIS AND CHARACTERISTICS

ASD is a neurodevelopmental disorder that currently affects 1 in every 54 children in the United States (Centers for Disease Control and Prevention, 2020). ASD is diagnosed based on the presence of social skill deficits as well as restricted and repetitive behaviors and interests (American Psychiatric Association, 2013). These deficits are present from an early age and typically persist across the lifespan. Additionally, for an individual to be diagnosed with ASD the deficits must not be attributed solely to an intellectual disability (American Psychiatric Association, 2013). Social deficits can include limited eye contact, decreased sharing of interests and emotions, difficulties forming social relationships (e.g., with peers), and difficulties with appropriate turn-taking during conversation. Restricted interests and repetitive behaviors (RRBs) may include repetitive non-functional play (e.g., spinning the wheels on a car), repeated motor movement (e.g., hand flapping), intense interests in a particular toy or topic (e.g., United States presidents), and immediate or delayed echolalia of a television show or another person's speech. Individuals with ASD may have rigid routines and become distressed when these routines are changed and may be unusually sensitive to sensory input from their environment.

STRUCTURAL LANGUAGE IN ASD

As previously mentioned, though language deficits are not currently part of the DSM-5 diagnostic criteria for ASD, ASD is frequently associated with impairments in structural language skills, including vocabulary (lexical abilities) and grammar (morphological and syntactic abilities; Ellis Weismer et al., 2011). Though some children with ASD develop strong language skills, the vast majority of children with ASD show considerable delays in early

vocabulary development (Tager-Flusberg et al., 2005), which is one of the main reasons for investigating early vocabulary assessment in this project. In fact, language delays are often an early warning sign of ASD (Wetherby et al., 2004) and may prompt further evaluation.

VOCABULARY SKILLS

Researchers and clinicians have aimed to examine how vocabulary knowledge and development may differ between typically developing (TD) children and children with ASD. A review conducted by Arunachalam and Luyster (2016) specifically addressed whether the underlying mechanisms used to develop vocabulary in TD children are present in children with ASD. The authors found that the mechanisms are present, but the efficacy of the mechanisms differ (Arunachalam & Luyster, 2016). Further, they argued that interruptions in other language systems that support these mechanisms, as well as the differences in the language intake of children with ASD, may be responsible for the disruptions in language development (Arunachalam & Luyster, 2016).

A 2003 study compared patterns of vocabulary development in infants with ASD and TD children in the following domains: understanding phrases, word comprehension and expression, and the production of gestures (Charman et al., 2003). The authors discovered variation in language acquisition between the two groups. In the ASD group, comprehension of words was delayed compared to word production, and production of early gestures were delayed compared to late gestures, which are atypical patterns. However, both groups demonstrated some similar patterns, "…including word comprehension being in advance of word production in absolute terms, gesture production acting as a 'bridge' between word comprehension and word production and the broad pattern of acquisition across word categories and word forms" (Charman et al., 2003, p. 214).

Another study compared vocabulary and language development in toddlers with ASD and toddlers without ASD who are considered to be "late-talkers" (Ellis Weismer et al., 2011). They found that children with ASD who used verbal language were severely delayed in productive vocabulary development compared to late-talkers (Ellis Weismer et al., 2011). In terms of the content of the vocabulary, there were similarities. Toddlers with ASD used comparable semantic categories to late-talkers and the two groups exhibited no differences in their use of vocabulary related to psychological states (Ellis Weismer et al., 2011). Additionally, there were no variations in the level of grammatical complexity and word combining between the two groups (Ellis Weismer et al., 2011).

EXPRESSIVE AND RECEPTIVE LANGUAGE

Some children with ASD show great disparities in spoken language even when compared to other children with ASD. These children are often classified as being pre-verbal or minimally verbal (Kasari et al., 2013). Minimally verbal children have very small spoken word and fixed phrase lexicons they use to communicate (Kasari et al., 2013). However, this definition of minimally verbal does not address the use of augmentative and alternative communication (AAC) to communicate, such as iPad applications, which are commonly used by children with ASD (Kasari et al., 2013). Minimally verbal children often show floor effects on standardized assessments even though they may demonstrate other language skills in contexts outside of standardized assessments, such as during play, making some standardized assessments less effective in assessing some children with ASD (Kasari et al., 2013).

While expressive language deficits are often more noticeable in children with ASD, there are greater deficits compared to TD children in terms of receptive language abilities. Receptive language is critical in the development of expressive language (Muller & Brady, 2016).

According to Fenson and colleagues (1994), TD children are found to have stronger receptive language skills than expressive language skills, such that they produce fewer words than they internally know (Muller & Brady, 2016). Children with ASD display both receptive and expressive deficits. However, research shows that these children may demonstrate their expressive abilities on assessments better than their receptive abilities (Muller & Brady, 2016). For example, a 2010 study conducted by Ellis Weismer, Lord, and Esler revealed that children with ASD exhibited stronger expressive language scores than receptive scores.

A study by Davidson and Ellis Weismer (2017) investigated whether severe receptive language impairments are a characteristic of ASD in the early stages. The basis for their research was from studies in which young children with ASD exhibited discrepancies in their comprehension and production. Specifically, research has found that there is a greater delay in language comprehension than spoken language in ASD (Barbaro & Dissanayake 2012; Barneveld et al. 2014; Charman et al. 2003; Ellis Weismer et al. 2010; Hudry et al. 2010; Kover et al. 2013; Loucas et al. 2008; Luyster et al. 2008; Maljaars et al. 2012; Seol et al. 2014; Volden et al. 2011; Davidson & Ellis Weismer, 2017). However, the study also discussed some research that contradicted this notion. Specifically, a 2007 study by Luyster and colleagues found that this comprehension-production discrepancy is not present in children with ASD compared to TD peers, though this research seemed to be an outlier.

Additional studies demonstrated that children with ASD did exhibit a comprehensionproduction discrepancy on standardized assessments across different subtests (Barbaro & Dissanayake 2012; Barneveld et al. 2014; Hudry et al. 2010; Kjelgaard & Tager-Flusberg 2001; Kover et al. 2013; Luyster et al. 2008; Davidson and Ellis Weismer, 2017). A study by Woynaroski and colleagues in 2016 compared expressive and receptive Communicative

Developmental Inventories (CDI) scores in children with ASD. The children were assessed at four time points between ages 24 and 48 months. The children's expressive scores were higher than receptive scores at all four time points. One thing to note is that this discrepancy is unique to ASD, and it persists longer in ASD than in other developmental language disorders (Davidson & Ellis Weismer, 2017).

The 2017 study by Davidson and Ellis Weismer looked into the specificity and sensitivity of comprehension versus production for classifying 30-month-old children who are late talkers compared to those with ASD. They hypothesized that there would be a larger discrepancy between comprehension and production scores in ASD and that it would be a sensitive and specific marker for diagnosing autism (Davidson & Ellis Weismer, 2017). The study looked at change over time, measuring comprehension and production at 30, 37, 44, and 66 months (Davidson & Ellis Weismer, 2017). The results revealed that the comprehensionproduction discrepancy decreased over time throughout the study (Davidson & Ellis Weismer, 2017). Additionally, the profiles of late-talkers and the children with ASD differed vastly. The difference in scores in the ASD group were negative, revealing that they had more severe delays in comprehension compared to production (Davidson & Ellis Weismer, 2017). The difference in scores in the late-talkers group were positive, so their comprehension was in line with age expectations, but production was lower than age expectations (Davidson & Ellis Weismer, 2017). The majority of the children in the ASD group displayed the comprehension-production discrepancy profile while none in the late-talkers group had this profile (Davidson & Ellis Weismer, 2017). The study concluded that this is a marker for ASD, but only in early ASD because it becomes less prevalent with time (Davidson & Ellis Weismer, 2017).

ASSESSING RECEPTIVE AND EXPRESSIVE VOCABULRY IN ASD

Studies of language development in children with ASD are often focused on the type and extent of delays they show. However, it is also important to consider the type of assessment being used to assess language skills, and the strengths and weaknesses it may have for this population. Many of the behaviors associated with an ASD diagnosis can make it difficult to assess language abilities in children with ASD, especially in children characterized as pre-verbal or minimally verbal. These difficulties in assessment, as well as factors related to parents' abilities to accurately report on their children's knowledge, may influence the psychometric properties of some assessments compared to others for this population (Brady et al., 2014; DiStefano & Kasari, 2016; Law & Roy, 2008; Muller & Brady, 2016; Tomasello & Mervis, 1994). Some types of assessment that may glean the most useful information for some populations may be less useful for the ASD population, and vice versa. Additionally, because language deficits in ASD vary greatly from person to person, clinicians may need to make decisions about what forms of assessment to use on an individual client basis. This is important because accurate assessment of language skills remains critical for determining appropriate comorbid diagnoses, planning treatment goals and strategies, tracking progress, and re-evaluating eligibility for services.

STANDARDIZED, EXAMINER-DIRECTED ASSESSMENTS

Standardized, examiner-directed assessments are one of the most commonly used types of language assessments. They are conducted by a trained adult and the administration procedures are consistent. One advantage is that standardized assessments glean direct information regarding children's expressive and receptive language abilities. Test items are designed to elicit specific responses in order for the child to demonstrate individual skills in each

of these domains. Two of the most common standardized assessments used in the young ASD population are the Peabody Picture Vocabulary Test (PPVT; Dunn, 2019) and the Expressive Vocabulary Test (EVT; Williams, 1997). The PPVT is used to assess a child's language comprehension. The examiner provides a word label that describes one of four pictures on a page. To indicate comprehension, the child must point to (or state the associated number for) the named image. Contrarily, the EVT is used to assess expressive skills. During the EVT, the examiner elicits verbal labels from the child by asking, "What is this?". The child must provide a correct spoken response to get each item correct. Another advantage of examiner directed assessments is that most provide normative data, in which an individual's performance is compared to their peers in the form of standardized scores.

While standardized assessments can provide useful information about a child's performance on a specific task, many clinicians and researchers have found some issues when using standardized assessments to evaluate children with ASD. For example, most standardized assessments require a relatively long administration time, so keeping a child engaged provides many challenges. This is an issue that is present when using standardized assessments to assess all children, but the issues may be amplified in children with ASD and other developmental disorders. Behavioral concerns for some children with ASD during administration include limited attention or lack of motivation (Koegel et al., 1997). Additionally, standardized assessments often require a minimum level of cognitive, metalinguistic, and social skills that are not meant to influence scores but can do so just the same. For example, asking a child to look at pictures and answer questions about those pictures requires joint attention, a developmental social skill that is sometimes weak or even not present in young children with ASD. Further, some assessments look at skills that are later developing in children with ASD (Muller & Brady,

2016). Other measures of language assessment may be more valid for the ASD population including language sampling and parent report measures (Kasari et al., 2013).

LANGUAGE SAMPLES

Language samples provide rich information about children's linguistic abilities and can be used in both clinical and research settings. Language samples are brief instances of language use that are meant to be representative of the individuals' overall language abilities. Some examples of language sample contexts include play sessions, story-retells, or simply conversation. The sample is typically recorded and analyzed. Clinicians and researchers may analyze the number of vocabulary terms used, the mean length of utterance, or the number of spontaneous utterances the child makes, among other things. One major benefit of language samples is that they capture naturalistic information (Muller & Brady, 2016). However, the samples require extensive of training and time to analyze and do not typically measure receptive language skills. Volden and colleagues (2011) argued that language samples may be best used in conjunction with other forms of assessment.

PARENT VOCABULARY REPORTS

Another frequently used assessment in the ASD population are parent report measures, in which parents or caregivers provide information about their child's abilities in a certain language domain. Parent vocabulary reports are used to measure children's vocabulary knowledge by having a parent report whether their child knows, knows and says, or does neither for a given word from a list of many early learned words. The MacArthur-Bates Communicative Development Inventories (CDI) are the most commonly used parent report measures (Mayor & Plunkett, 2011). There are two main forms of the CDI: the CDI Words and Gestures (CDI-WG) and the CDI Words and Sentences (CDI-WS). The CDI-WG is used for children ranging from 8 and 18 months of age and the CDI-WS is used for children ranging from 18 to 30 months of age.

Parent report measures have several advantages. They use naturalistic data and do not require extensive training to administer (Law & Roy, 2008). Other standardized assessments require considerable time to understand the nuances of the assessment protocol and scoring procedures. Parent reports require no administration training and only a brief explanation of the directions to parents and caregivers. These reports tap into parent-specific information regarding a child's language abilities. Parents observe their children in perhaps the most naturalistic setting, in the home, and in a variety of other settings. Parents also see their children more than anyone else so these reports tap into exclusive information that cannot be found in other forms of assessment (Law & Roy, 2008). While examiner-directed assessments provide a snapshot of a child's abilities, parents can provide more in-depth information regarding their child's abilities based on their time spent with their child. Rather than evaluating a child's performance in a rigid, unfamiliar setting (e.g., a clinical testing room), these reports capture a child's behaviors in "reallife" scenarios, with all of the confounding factors that impact their abilities on a daily basis. Lastly, the information gleaned from these reports can be used both clinically and for research purposes. For example, an SLP may target goals based on skills that parents report as being weak for their child. Additionally, researchers may use parent report data in studies.

While there are many benefits of these parent measures, there are many drawbacks, as with any assessment method. First, comprehension of language can be especially difficult to assess when children produce limited spoken language. Expressive language may be the strongest indicator of vocabulary understanding, and it may be extremely difficult for parents to assess these two linguistic elements independent of each other. Another concern regarding

parent reports is the accuracy of a parent's response. Parents can be unintentionally biased when estimating their children's knowledge and abilities (Law & Roy, 2008). Some parents may tend to over- or underestimate their child's vocabulary knowledge. Parent report measures frequently direct parents to only consider their child's spontaneous spoken language, not language that is prompted or imitated (e.g., immediate echolalia). While it may seem easy to discern the difference between these instances, it may actually require a more nuanced look on the parent's part. Some of the cons related to these assessments may be even more apparent for children with language deficits. Most vocabulary reports specify that the parent should only consider spoken language output, and not consider the use of any AAC devices. Because some children with ASD produce limited spoken language, the validity of parent reports may be weaker than when they are used to assess typically developing children's lexicons. Additionally, individuals frequently rely on non-verbal behaviors to confirm understanding. Subtle facial expressions, eye-contact, and nods and gestures can convey meaning, but these behaviors may be ambiguous in children with ASD. Because of deficits in these types of social behaviors, parents of children with ASD may be missing these cues when determining if their child understands certain words, which can lead to uncertainty in their responses. Another concern is the length of these checklists. Since they are time consuming, some parents may not be able to take the time to fully evaluate their answers and may answer similarly across items.

PSYCHOMETRIC PROPERTIES OF PARENT REPORTS

Studies have been conducted to evaluate the psychometric properties of parent report measures. In 1993, Fenson and colleagues reported that the CDI does have good test-retest reliability and a high correlation between receptive language and language production.

Additionally, Fenson and colleagues also reported that the CDI does have high internal consistency (1993).

A 2000 study evaluated the psychometric properties of both versions of the CDI in relation to a diverse sample of children (Feldman et al., 2000) The researchers compared scores on the CDI-WG given at age 1 with scores on the CDI-WS given at age 2 in relation to demographic variables, such as maternal level of education and socioeconomic status (SES; Feldman et al., 2000). The study found that the CDI did in fact vary as a function of demographic variables (Feldman et al., 2000). The most consistent variable was sex, according to the authors, with girls scoring higher than boys on all but one scale of the CDI-WG and all scales of the CDI-WS (Morisset, Barnard, & Booth, 1995; Feldman et al., 2000). The variable of race was associated with scores on every scale of the CDI-WG and on three out of five of the scales of the CDI-WS, but race was responsible for minimal variance in scores (Feldman et al., 2000). One variable that was associated with more drastic differences in scores on both the CDI-WG and the CDI-WS was maternal education and health insurance status (Feldman et al., 2000). However, the directions of the associations with maternal education and insurance status were not consistent within and across the two versions of the CDI (Feldman et al., 2000). There were negative correlations observed between the parental reports of understanding and SES, which provided evidence that parents who have lower education and income levels use more liberal definitions of comprehension to judge their children's vocabulary knowledge (Feldman et al., 2000). Maternal education and health insurance status had negative associations with scores on the Vocabulary Production Scale of the CDI-WG but was positively correlated with the CDI-WS version of that scale (Feldman et al., 2000). The authors theorized that at age 1, judgements about vocabulary production do require more interpretation than judgements made at age 2

(Feldman et al., 2000). In turn, parents may overestimate language at age 1, but accurately estimate language at age 2 (Feldman et al., 2000). In terms of the overall correlations between the CDI-WG and the CDI-WS, the study contradicted previous research because the correlation was only moderate (Feldman et al., 2000).

Authors stated that the evidence suggests that scores on the different sections of the CDI may be prone to reporting bias (Feldman et al., 2000). Some of the CDI sections require more reporter interpretation, while some ask the reporter to provide information that is more concrete (Feldman et al., 2000). On sections that demand interpretation, parents with lower education and income seem to overestimate the abilities of their children (Feldman et al., 2000). A 1978 study by Chapman did state that generally parents have a difficult time differentiating between instances on comprehension that are heavily influenced by nonlinguistic cues and environmental factors that influence meaning (Feldman et al., 2000). Further, parents may have a tendency to confuse exposure to certain vocabulary terms with understanding of said terms (Feldman et al., 2000). This may provide an explanation for why parent reports of vocabulary for young children provide higher scores than other direct tests (Tomasello & Mervis,1994).

A 2005 study examined the predictive and concurrent validity of parent reports at two time points in TD children (Feldman et al., 2005). The researchers compared children's scores on the CDI at ages 2 and 3 with scores on cognitive and receptive language tests conducted at age 3 (Feldman et al., 2005). The study looked at three versions of the CDI: the CDI-WG and the CDI-WS, and the CDI-III, which is used for children ages 30-42 months and assesses productive vocabulary, the maturity of the child's syntax, and their use of language (Feldman et al., 2005). The researchers found that girls in the study had significantly higher scores on all components of the CDI-WS except for the Sentence Complexity section compared to their male peers (Feldman

et al., 2005). When comparing performance on the assessments of subgroups based on maternal level of education, differences in scores were statistically significant on the Vocabulary Production and the Three Longest Sentences portions of the assessment of the CDI-WS and the Sentences and Using Language portions of the CDI-III (Feldman et al., 2005). Correlations between scores on the CDI-III and standardized assessments were statistically significant and correlations of scores on the CDI-WS Vocabulary Production and Three Longest Sentences scales and the standardized assessment scores were statistically significant as well (Feldman et al., 2005). However, the correlations between the CDI-WS Sentence Complexity section were not statistically significant compared to scores on the standardized assessments used (Feldman et al., 2005). The researchers did find concurrent validity between the CDI-III and the standardized assessments (Feldman et al., 2005). The CDI-WS was found to have predictive validity on two of its three subtests when compared with other measures (Feldman et al., 2005). Overall, the authors stated that the results of their study were consistent with previous research at the time and state that parents are reasonably good at assessing their children's vocabulary knowledge between the ages of 18 and 30 months (Feldman et al., 2005).

Pan, Rowe, Spier and Tamis-Lemonda (2004) specifically looked at children from lowincome families in order to assess how the CDI gathered information about their expressive vocabulary abilities. The sample compared the children's performance on assessments at two different time points, when the children were 2 and when they were 3 (Pan, Row, Spier & Tamis-Lemonda, 2004). Similar to the previous study, this was done to assess the predictive nature of the CDI. At age 2, the children were given the CDI and a spontaneous speech sample as well as standardized expressive assessments so that the CDI could be compared to other types of assessments (Pan, Row, Spier & Tamis-Lemonda, 2004). Then, at age 3, the children were given a receptive language assessment, the Peabody Picture Vocabulary Test, to determine if expressive scores on the CDI predicted receptive scores on the PPVT (Pan, Row, Spier & Tamis-Lemonda, 2004). The researchers found that there was only a moderate association between the CDI scores and the other assessments given at age 2 and the CDI scores "…accounted for unique variance in the PPVT scores at age 3;0…" (Pan, Row, Spier & Tamis-Lemonda, 2004; p. 587-588). One interesting finding in the study was that when the children in the study were divided into two subgroups of being from rural versus urban areas, predictive validity between the CDI and the PPVT was stronger for the rural group (Pan, Row, Spier & Tamis-Lemonda, 2004).

While the previous studies looked at the psychometric properties of the assessment in TD children, other studies did examine its psychometric properties for children with ASD and other developmental disorders. A 1997 study by Yoder, Warren, and Biggar aimed to evaluate the CDI infant scale (CDI/I), also called the CDI-WG, and its stability over a two-week period of time in children with developmental delays. The study looked at two types of reliability: overall reliability and item-by-item reliability (Yoder, Warren, & Biggar, 1997). The overall reliability refers to whether the overall amount of words a child was reported to know and/or say stayed the same when mothers fill out the CDI-WG twice over a two-week period. Item-by-item reliability refers to consistency in responses for individual words when retested twice across a two-week period. The study found that there was good overall reliability, meaning that the CDI-WG is a good instrument for assessing overall vocabulary deficits. The authors interpreted these findings as indicating that the use of the CDI-WG to assess the approximate amount of words children with developmental delays know or say is reliable (Yoder, Warren, & Biggar, 1997).

However, the study did not find strong item-by-item reliability, meaning that mothers' responses on individual words changed over the two-week period (Yoder, Warren, & Biggar,

1997). The study stated that one thing that could impact these inconsistencies is the confidence of the mothers' responses, or their certainty (Yoder, Warren, & Biggar, 1997). Some factors that the study found to contribute to the mothers' confidence are the nonverbal behaviors that the children use to express communication and the mothers' socioeconomic status (Yoder, Warren, & Biggar, 1997). The authors concluded that the CDI-WG is not an effective and accurate tool to use when selecting specific vocabulary words to target in therapy due to inconsistencies in itemby-item scores, but it is a good tool to use to determine if a child with developmental delay has overall expressive vocabulary deficits (Yoder, Warren, & Biggar, 1997).

A 2013 study by Nordahl-Hansen, Kaale, and Ulvund examined the reliability of parent responses on the CDI-WG for children with ASD. The researchers had both parents and preschool teachers of 55 children with ASD fill out the CDI. The aim of the paper was to assess the inter-rater reliability and to look at whether parents tend to over or underestimate their child's abilities. The authors found that there was in fact strong inter-rater reliability between the parent responses and the teacher responses. Unsurprisingly, parents did tend to give their children higher ratings to an extent, however there was only a minimal variance. The authors noted that these differences could likely be attributed to contextual differences. It was concluded that the findings imply that parents are fairly good at evaluating their own children's vocabulary, making the CDI-WG a good assessment for the pediatric ASD population.

Another, later study by the same authors evaluated the McArthur-Bates CDI in children with ASD (Nordahl-Hansen, Kaale, & Ulvund, 2014). The goal of the study was to measure the concurrent validity between the CDI and other assessments, including the MSEL, RDLS, and the CDI when both parents and teachers fill it out (Nordahl-Hansen, Kaale, & Ulvund, 2014). The study found that there was a high correlation between the parent reported

CDI, the teacher reported CDI, the MSEL, and the RDLS in terms of expressive vocabulary scores (Nordahl-Hansen, Kaale, & Ulvund, 2014). It was no surprise that the CDI filled out by the parents and other direct measures had a high concurrent validity because this is in line with previous research (Luyster et al., 2008; Miniscalco et al., 2012; Nordahl-Hansen, Kaale, & Ulvund, 2014). These results suggest that using a parent report such as the CDI can give us some information comparable to standardized assessments. When looking at receptive language, the study found that there was concurrent validity between the parent and teacher CDIs and direct tests, but the validity was higher for teacher reports than parent reports (Nordahl-Hansen, Kaale, & Ulvund, 2014). There are many reasons why this may be the case. For example, some hypothesize that parents may overestimate their children's language abilities (Nordahl-Hansen, Kaale, & Ulvund, 2014). This is one consistent criticism of parent reports (Tomasello & Mervis, 1994).

GOALS OF THE PRESENT STUDY

These studies taken together demonstrate that there are some domains in which parent reports have strong psychometric properties, and some domains where the reports do not. One reason for this is that it may be difficult for many parents to discern true vocabulary knowledge, especially when their child has minimal expressive vocabulary and produces ambiguous nonverbal responses to parents' questions and comments. We proposed that the decreased reliability of parent report measures may be related to parents' varying confidence in their responses, as well as their child's emerging knowledge (Fernald et al., 2006). In this project, we examined responses of parents of children with ASD on a novel vocabulary checklist that measures the certainty of parents' responses for individual items. We also looked at responses on our novel checklist from both parents of children with ASD and parents of TD children and assessed

differences across these two groups. Lastly, we compared these responses from the parents of children with ASD with their responses for the same set of words on the MacArthur-Bates CDI-WG, to determine reliability. This thesis project had three goals:

1. To characterize the information gathered by a custom vocabulary checklist completed by parents of young children with ASD.

We predicted that certainty ratings would vary considerably across individual children and across individual words. We also predicted that certainty ratings would be significantly lower for children with ASD than for a comparison group of TD children of the same age.

2. To determine whether parent certainty ratings differ based on response (i.e., whether the child says the word vs. understands the word vs. neither) or child characteristics (e.g., language skills, autism severity).

We predicted that parent certainty ratings would be higher for words that children produce than for words they only understand or neither understand nor produce. We predicted that certainty ratings would be lower for children with weaker language skills and higher autism severity. We also expected to identify links between reliability and parent education level.

3. To examine the relationship between parent certainty ratings and parent reliability in reporting vocabulary knowledge across two separate vocabulary checklists.

We predicted that parent certainty and parent reliability would be significantly and positively correlated.

METHOD

This study included 19 children with ASD (2 to 5 years old) and their parents. A comparison group of 28 TD children (2 to 5 years old) and their parents were also included to help address the first aim. This data was previously collected as part of an existing project in Dr. Venker's lab. My role included organizing and characterizing the data, deriving relevant variables (e.g., mean certainty rating by parent), helping to analyze the data to address the project aims, interpreting the results, and writing the thesis document.

Participants were recruited through parent email listservs, social media groups, doctors' offices and clinics, and flyers posted in the community. The children in the ASD group had all received a diagnosis of ASD. They were reported by their parents not to have any other known genetic disorders. Based on clinical observation, the children in the TD group did not show any signs of ASD. In addition, their parents filled out the Social Communication Questionnaire (Rutter et al., 2003), which presented yes-or-no questions regarding their child's social communication. The questionnaire was designed for children ages 4 and older, who have a mental age of over 2 years old. Parents in both groups completed a Background Information Form that collected information such as how many siblings the children have, race and ethnicity, and medical history.

The study was prospectively approved by the Michigan State University Internal Review Board. Parents provided written informed consent for their child to participate. Children in both groups came into the lab for a two-day visit. Parents in both groups completed several questionnaires, including vocabulary checklists and a background information form that collected information about the child's race and ethnicity, medical history, and intervention

services (for the children with ASD). Children participated in standardized assessments, eyegaze tasks, and two play sessions used to gather language samples. The eye-gaze protocol was a looking-while-listening task that assessed language comprehension. Children and their parents participated in two play sessions. In one session the children used traditional, non-electronic toys to play with, and in the other they were given all electronic toys to play with.

EXAMINER-DIRECTED ASSESSMENTS

Children were given the Preschool Language Scales, 5th Edition (PLS-5; Zimmerman, Steiner, & Pond, 2011) to assess receptive and expressive language and the Visual Reception scale from the Mullen Scales of Early Learning (Mullen, 1995) to assess memory and problem solving. The children in the ASD group were also given the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2; Lord et al., 2012; Luyster et al., 2009) to confirm ASD diagnosis and measure autism severity.

MACARTHUR-BATES COMMUNICATIVE DEVELOPMENT INVENTORIES

Parents of the children with ASD filled out the McArthur-Bates CDI Words and Gestures Form (CDI-WG) and the McArthur-Bates CDI Words and Sentences (CDI-WS) form. Parents of children in the TD group were not asked to complete the CDIs because the children with typical development likely would have been at ceiling on these measures, given that they are designed for children with typical development between 8 and 18 months of age (CDI-WG) or between 18 and 30 months of age (CDI-WS). The two CDI forms gathered information regarding the children and the ASD group's vocabulary knowledge, similar to our vocabulary checklist.

CUSTOM VOCABULARY CHECKLIST

Our custom vocabulary checklist (see Appendix) consisted of eight verbs and twenty-four object nouns. These were all common vocabulary words that are frequently in children's early

lexicons and on the McArthur-Bates CDI. As a first step, this study focused only on object nouns. Parents were given the following instructions: "Please circle whether your child understands, understands and says, or neither understands nor says the following words. For each word, also circle how certain you are about your response". Response options were: Very Certain, Certain, Neutral, Uncertain, and Very Uncertain.

ANALYSIS PLAN

The first aim was addressed primarily through the reporting of descriptive statistics, including mean, standard deviation, and range. An independent samples *t*-test compared the mean certainty ratings for the children with ASD and the TD children. For the second aim, independent samples *t*-tests were used to compare certainty ratings by response. In addition, Pearson correlation coefficients were calculated to assess the relationship between mean certainty ratings and child and parent characteristics. For the third aim, we conducted correlational analyses to assess the relationship between parent certainty ratings and parent reliability in reporting vocabulary knowledge.

RESULTS

Our study had three research aims. Aim one was to characterize the information gathered by a custom vocabulary checklist completed by parents of young children with ASD. We hypothesized that certainty ratings would vary across individual children and words. Additionally, we predicted that certainty ratings would be significantly lower for children with ASD than for a comparison group of TD children of similar ages. This aim was accomplished by gathering descriptive data about parental certainty ratings on our custom checklist. Specifically, for each vocabulary item, we gathered the mean certainty rating across parents, the standard deviation, and the minimum and maximum response. Our analyses revealed variability in certainty across parents and across words. The vocabulary term with the lowest mean certainty was "truck", with a mean certainty rating of 3.72 and the largest range of parental responses, from 0 to 5, with 0 being "very uncertain" and 5 being "very certain". The vocabulary term with the highest mean certainty rating and the smallest range of responses was "slide", with a mean certainty rating of 4.42 and responses ranging from 4 to 5. Figure 1 shows the mean certainty rating for each word across all parents.



Figure 1. This figure illustrates the mean certainty response for each vocabulary word.

Additionally, we gathered each parent's mean certainty rating, illustrated in Figure 2. The parent with the lowest mean certainty across all vocabulary items had a mean certainty rating of 3.17. The parents with the highest mean certainty rating had a mean of 5. This means that they rated their responses for every word on our custom checklist as "very certain".



Figure 2. This figure illustrates the mean certainty response for each parent.

Lastly, we examined the difference in parental certainty in parents of TD children compared to parents of children with ASD. We found that parents of TD children reported significantly higher levels of certainty (M = 4.94, SD = 0.07) than parents of children with ASD (M = 4.09, SD = 0.20, p < .001).

The second aim of our project was to determine whether parent certainty ratings differ based on response (i.e. whether the child says the word vs. understands the word vs. neither) or child or parent characteristics (e.g. language skills, autism severity, parent education level). We predicted that certainty ratings would be higher for words that children produce than for words they only understand or neither understand nor produce. We also predicted that certainty ratings would be lower for children with weaker language skills and higher autism severity. To address the first part of this aim, we used t-tests to compare certainty ratings by response, regardless of the individual word. We found that the mean certainty rating (MCR) for "says" (M = 4.613, *SD* = 0.545; between "certain" and "very certain") was significantly higher than the MCR for "understands" (M = 3.810, SD = 0.929; p < .001; between "neutral" and "certain"). More specifically, for all but two words, "says" had the highest MCR. The two exceptions were the words "shoe" and "sock", for which the "neither" response had the highest certainty rating. Additionally, the MCR for "says" was significantly higher than the MCR for "doesn't know" (M = 3.658, *SD* = 1.024, p< .001; between "neutral" and "certain"). Lastly, there was not a significant difference found between the MCRs for "understands" and "neither" (p = .249).



Figure 3. Parent certainty ratings for words children were reported to say.



Figure 4. Parent certainty ratings for words children were reported to understand (but not say).



Figure 5. Parent certainty ratings for words children were reported to neither say nor understand.

To address the second part of this aim, we tested the correlation between parent certainty ratings and select child and parent characteristics, including autism severity, child age, expressive and receptive language skills, and parental level of education (see Figures 6-10). Autism severity was determined by ADOS-Severity scores, language skills were determined by Expressive Communication and Auditory Comprehension growth scale value scores on the PLS, and parent level of education was determined by years of education. Our results indicated that there was a significant correlation between child PLS Auditory Comprehension scores and certainty (r = .479, p = .038). There was a marginal association between child PLS Expressive Communication scores and certainty (r = .418, p=.075). There was not a significant correlation between parental years of education and certainty (r = .253, p=.297), child age and certainty (r = .235, p=.332), or autism severity and certainty (r = .019, p=.939).



Figure 6. Relationship between parent mean certainty ratings and child chronological age.







Figure 8. Relationship between parent mean certainty ratings and child receptive language growth scale value from the PLS-5.



Figure 9. Relationship between parent mean certainty ratings and child expressive language growth scale value from the PLS-5.



Figure 10. Relationship between parent mean certainty ratings and parent number of years of education.

The final aim of this project was to examine the relationship between parent certainty ratings and parental reliability in reporting vocabulary knowledge across two different vocabulary checklists: the CDI-WG, which is a traditional parent vocabulary checklist, and our custom vocabulary checklist. We predicted that certainty and reliability would be significantly and positively correlated. This aim only focused on the ASD participant group because parents of TD children only filled out our novel checklist (as the TD children would have been at or near ceiling on the CDI). The variables being compared were parent consistency across the two checklists and the mean certainty rating across child. To determine consistency, we compared parents' responses for each word across the two checklists and counted the number of instances in which parents had the same response to obtain a percent consistency for each parent.

The minimum percent parental consistency was .51 and the maximum was 1. The mean of parent consistency ratings was .74 with a standard deviation of .129. The minimum mean certainty rating on the custom checklist was 3.16 with a max mean certainty of 5. The average of all the mean certainty ratings across children was 4.09 with a standard deviation of .57. We conducted a Pearson correlation analysis to examine the relationship between the average certainty rating on our custom parent vocabulary checklist across each child with parent certainty across both our custom checklist and the McArthur-Bates CDI. Once analyzed, the results yielded an r value (Pearson correlation coefficient) of .56 with a p-value of .013. According to Cohen (1988), an effect size of .50 or more is a large correlation. Therefore, according to this



threshold, there is a strong correlation between parent certainty and parent reliability. The scatter plot below illustrates the relationship between parent reliability/consistency and certainty.

Figure 11. The above figure demonstrates the correlation between the mean certainty across each individual parent's ratings on the custom vocabulary checklist and their consistency/reliability of their responses across the two checklists.

DISCUSSION

AIM 1

Our results regarding differences between parents of TD children compared to parents of ASD were as predicted. Parents of children with ASD were significantly less certain about their responses in reporting their children's vocabulary knowledge than parents of TD children of

similar ages. There are many possible reasons for this pattern. First, the mere diagnosis of ASD could make some parents second guess their abilities to make judgements about their child's skills. Second, part of the diagnostic criterion for an ASD diagnosis is social impairments and differences. Deficits in social interactions, particularly reduced social interactions, could lead parents to be more uncertain of their child's vocabulary knowledge. For example, if a child walks up to their parent independently, holding an apple, and says the word "apple", a parent may confidently believe that their child knows this word. If a child does this multiple times a week, this may lead them to be even more certain. However, if a child only says "apple" when their parent asks them what they want to eat multiple times with fleeting attention, and when they do respond they exhibit limited eye contact, a parent may be more uncertain. While it is impossible to discern whether the social aspects of the interaction are what led to the parent's uncertainty, we can presume that it may have some influence.

To our knowledge, this study is the first to use a novel vocabulary checklist that asks parents to report how certain they are about their judgments of their child's knowledge of specific vocabulary words. The findings for aim one revealed that most of the vocabulary items had certainty ratings ranging from 1 or 2 to 5. Mean certainty ratings ranged from 3.72 to 4.42 across vocabulary items, and from 3.17 to 5 across parents, revealing great variability amongst parents for each item. This variability supports the potential usefulness of measuring parental certainty of children's vocabulary knowledge. If variability in these domains had been limited, this would suggest that our custom checklist had limited construct validity. The most common certainty rating across parents was "certain" and the least common was "very uncertain". Because "very uncertain" was the least common rating, regardless of the response, we can see that parents generally lean more towards certainty than uncertainty when completing the

checklist. However, because our checklist included a "neutral" response, this provided parents with an opportunity to not pick either side of the spectrum. Future studies could eliminate the "neither" rating to force parents to pick between either side of this spectrum of certainty.

In addition to overall findings, it is interesting to consider specific items that fell on the extreme ends of the certainty ratings. "Truck" was the vocabulary term with the lowest certainty rating and the largest range in minimum and maximum certainty ratings. This was the only item to receive a certainty rating of 0, meaning that parents were very uncertain about their responses. There are many reasons why "truck" could lead to such variability. First, "truck" could refer to both a literal automobile as well as a toy truck, making it somewhat ambiguous. Additionally, it is a word that could require a certain level of exposure to understand. If a child's parent drives a truck, they are more likely to know this word than a child whose parent drives a car or SUV. Additionally, if a child is very interested in toy cars and trucks, they are more likely to know this word than a child who does not have any toy trucks. The word with the highest certainty rating and lowest standard deviation, as well as the smallest minimum and maximum range, was "slide". Certainty responses ranged from 4 to 5 for this word, making it the only word that all parents were certain about their responses. This may be because slides are exciting and can elicit strong, visceral memories. This may mean very positive and happy memories, if a child loves slides, but also could mean very negative memories if a child is afraid of slides. One thing to note about this word is that it is ambiguous in that "slide" can be a noun or a verb. However, because of the high certainty responses for this word, we predict that parents viewed the word as a noun rather than a verb.

Upon additional analysis, we did not find patterns in certainty ratings based on the categories of the words. For example, parents did not have noticeably higher certainty ratings for

foods or toys. This is surprising because one may predict that more exciting items, such as toys, would elicit more responses, either verbally or behaviorally, from children, compared to clothing, for example, which may inform parental certainty. Additionally, we did not find patterns based on the level of production difficulty for each word. For example, certainty ratings did not correlate with syllable length or early versus later developing phonemes. While such findings could be informative, they would also decrease the content validity of our custom assessment, because the aim of the checklist is to have parents respond based on their child's knowledge and understanding of the word itself rather than their ability to *accurately* produce the word.

AIM 2

Our analysis confirmed our hypothesis that parents would have the highest level of certainty for words their child says compared to words they are reported to know or are reported to be unknown. This is not surprising because production is the most tangible demonstration of internal knowledge. We also found no significant difference in mean certainty ratings between the responses of "understands" and "neither". Because of these findings, when looking at certainty varying as a function of response, it appears that the child saying the word has the greatest influence on parental certainty. This finding also helps to explain the significant difference in certainty ratings between parents of TD children and parents of children with ASD because the TD children were almost exclusively reported to say all of the words on our custom checklist.

As mentioned in the section discussing the findings of aim 1, we did find that there was great variability in the certainty ratings across words. However, our overall findings suggest that the variability is not random. We found that certainty ratings differed based on

response (I.e. understands vs says vs neither) and correlated with language skills. Additionally, parents of TD children were significantly more certain of their responses compared to the parents of children with ASD. These findings taken together suggest that the certainty ratings were not random and were in fact meaningful. Previous research has discussed concerns over the validity of parental reports (Brady et al., 2014; DiStefano & Kasari, 2016; Law & Roy, 2008; Muller & Brady, 2016; Tomasello & Mervis, 1994). The current findings suggest that validity and reliability of parent report measures may differ across families. Additional studies are needed to examine the psychometric properties of using certainty ratings on parent report measures.

As previously discussed, our data revealed that there is a significant correlation between language skills, specifically auditory comprehension skills, and parental certainty, but no significant correlation between autism severity and parental certainty. Upon examining individual participants in the ASD group, we can understand this finding more clearly. Participant 4122, or Participant A., had the highest mean certainty rating of all the participants in the ASD group, with a mean of 5, and a parent reliability score of 1. Participant 4129, or Participant B, was the child with the lowest mean certainty rating, with a mean of 3.17 and a reliability score of .61. When looking at the ADOS autism severity ratings for each of these children, Participant A received a severity rating of 9 while participant B received a severity rating of 8, meaning that the child whose parent had higher certainty ratings also had higher ASD severity according to this measure. However, Participant A received stronger language scores on the PLS-5, with an Auditory Comprehension standard score of 98 and an Expressive Communication standard score of 93, both of which are within normal limits, while Participant B received an Auditory Comprehension standard score of 50 and an Expressive Communication standard score of 63, both of which are significantly below the average range. This implies that

perhaps autism severity is not as strong of an indicator of parent certainty and reliability compared to language skills.

It is unsurprising that language skills could influence how certain parents are of their children's vocabulary knowledge. While we did not find a significant correlation between expressive language scores certainty, a positive trend was observed. This is perhaps the least surprising pattern that emerged from correlating certainty with other child characteristics. If a child is producing a variety of words, this can provide evidence to parents that their child internally has strong vocabulary knowledge. Even if a child does not verbally produce all of the words on a vocabulary checklist, parents may tend to give their child higher certainty ratings because they have demonstrated their knowledge of a variety of other, possibly related, words. It seems surprising that auditory comprehension was significantly correlated with certainty, but expressive communication was not. However, auditory comprehension is another way that internal vocabulary knowledge may be demonstrated to parents. For example, if a child consistently performs an action when a parent gives their child verbal directions, such as "get the ball", a parent may be very certain that their child knows the word, even if it is not being produced. Additionally, children may express words they do not comprehend. For example, if a child produces the word "apple" but calls every round red object an "apple" (i.e., overgeneralization), they are producing the word with a lack of in-depth understanding. Parents may be able to separate vocabulary knowledge demonstrated through comprehension from vocabulary knowledge demonstrated through expression, which may explain the significant correlations found in this study.

We originally predicted that higher autism severity would correlate with lower levels of certainty. However, our findings were not consistent with this prediction, as autism severity and

certainty were not significantly correlated. Using the diagnostic criteria of ASD as a lens to examine this finding may provide an explanation. Autism severity is not indicative of language abilities outside of the pragmatic domain, because language skills are not part of the ASD diagnostic criteria. Because of this, children with ASD present with a wide range of language abilities, with some being minimally verbal while others have language skills comparable to TD peers. The certainty measure we are correlating with autism severity is on a measure of language skills, not a measure of one of the diagnostic domains of ASD. Therefore, the lack of significant correlation between these two variables suggests that parents may be able to differentiate language skills from the behaviors associated with ASD.

AIM 3

Our findings for aim 3 supported our prediction that certainty and reliability would be positively correlated. Parents who were more reliable in reporting their child's vocabulary knowledge were also more certain about their responses, which provides evidence that parental certainty in their responses is in fact warranted. This finding has both clinical and research implications. Clinically, if clinicians were to provide parents with one vocabulary checklist that did have them report on their certainty, clinicians could predict that the parents with higher levels of certainty may be more reliable in reporting their child's vocabulary knowledge than parents who are less certain, which can inform how clinicians plan treatments and coach parents through developing their child's language skills. In terms of research, these findings could inform future studies where parental reliability is correlated with other variables that we found to be positively correlated with certainty. For example, researchers could correlate reliability with child language skills on a measure such as the PLS.

The finding that reliability and parental certainty were positively correlated even in this relatively small sample of children with ASD can also provide us with insight about the psychometric properties of parent report measures. Previous literature has reported the CDI-WG has good inter-rater reliability when parents and preschool teachers of children with ASD complete the checklist, concluding that the assessment itself is an effective measure of vocabulary for young children with ASD (Nordahl-Hansen, Kaale, & Ulvund, 2013). Our findings advance understanding of this issue because we found that parents who are confident in their ability to assess their child's knowledge have good intra-rater reliability across two separate vocabulary checklists.

LIMITATIONS

There were multiple limitations to the current study. First, certain aspects of the study were not explicitly controlled, such as the time between filling out the CDI and the custom vocabulary checklist. Some parents may have filled the two checklists out close together in time, while others may have filled one out at the beginning of the session and the other towards the end. The proximity between filling out the checklists could impact the reliability of parents' responses. Another limitation to the study is that the order in which the checklists were completed was not explicitly controlled for. Parents were presented a variety of forms, including the checklists, in a consistent order, with the custom checklist being presented before the CDI. Through informal observation, it appeared that most parents completed our custom checklist before the CDI, but there may have been some instances where parents chose to complete the forms in a different order.

An additional limitation of the present study was that it was difficult to compare the ASD participant group to the TD group because the TD participants could only fill out our custom

checklist, and not the CDI. The TD children in this study were out of the developmental age range for the CDI and would have performed at ceiling, preventing us from obtaining meaningful scores in order to compare the reliability of parent report measures across the two participant groups. Lastly, the present study had a relatively small sample size. Future work addressing these issues in a larger sample of young children with ASD is needed.

FUTURE RESEARCH

There are many ways our research findings may inform future research regarding a parent's ability to assess their own child's vocabulary knowledge. A possible future study could use this same project format but instead of having parents reporting on noun knowledge, the researchers could look at verb knowledge. Our custom checklist did include verbs, but for the purposes of this project, we only analyzed nouns as a first step. It would be informative to determine whether similar patterns in certainty and reliability emerge when parents are reporting on action verbs, which typically (unlike object nouns) involve ongoing movement (e.g., run, swim).

Another example of a future study that could build on our findings is cross referencing parent vocabulary reports with looking-while-listening tasks (LWL). LWL is a method for implicitly measuring vocabulary comprehension by presenting multiple images on a screen (e.g., a car and a shoe) and asking children to find one of them (e.g., Where's the car?). There has been previous literature that examined the relationship between LWL results and parent report measures. For example, a 2006 study assessed the relationship between parent reports and LWL tasks for TD infants at three different chronological ages (Houston-Price, Mather, & Sakkalou, 2006). The researchers used the British English version of the MacArthur-Bates CDI. The children were presented with pairs of images of which they were reported to either know both

names of the images or know neither of the names of the images. At all ages, the participants looked preferentially at the target images for words their parents reported as known and word reported as unknown, which suggests that parents may underestimate their children's vocabulary knowledge on reports.

Another study that examined LWL tasks and parent reports was conducted with children with ASD (Venker et al., 2016). The researchers found that the children showed comprehension of the target words, even when they were reported by parents to be unknown, again suggesting that parents may underestimate children's vocabulary knowledge, especially when their children have limited expressive language output. LWL tasks allow us to assess receptive knowledge in children even if they have limited verbal output. In such a study, we could compare words that parents report that their children know with words that their children demonstrate knowledge of on an LWL task. Then, this reliability could be correlated with parental certainty levels, to assess whether confidence correlates with parental accuracy.

Another future study could be conducted with an ASD participant group compared to TD children with similar language levels. One aim of our current study was to assess whether ASD severity or language skills is a stronger indicator of parental certainty. A study in which children with ASD's parental vocabulary reports are compared with those of similar language levels, such as children who are TD but scored similarly on standard language measures, such as the PLS-5, could help us determine whether language skills or ASD correlates with parental certainty more strongly.

CONCLUSION

The present study evaluated the use of a custom parent vocabulary checklist that asked parents to report on the certainty of their responses about their children's vocabulary knowledge. As expected, mean certainty ratings varied a great deal across words and across parents. Parents of TD children reported higher levels of certainty compared to parents of children with ASD. Additionally, parent certainty was significantly and positively correlated with auditory comprehension skills in children with ASD, such that children with strongly language skills had parents who reported higher certainty about their responses. Parental certainty was also significantly and positively correlated with parental reliability in reporting vocabulary knowledge across two separate vocabulary checklists. These findings have both clinical and research implications regarding the evaluation of vocabulary skills, the use of parental report measures, and therapeutic planning for the pediatric ASD population. Future research can build on the present findings to further explore how parental certainty can inform both research and clinical practice. APPENDIX

Appendix: Custom Vocabulary Checklist

Please circle whether your child understands, understands and says, or neither understands nor says the following words. For each word, also circle how certain you are about your response.

	-				
Turtle	Understand	ls	Understands and says	ン	Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
Drinking	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
	Understands				
Eating	Understand	ls	Understands and says		Neither
Eating Level of Certainty	Understand Very Uncertain	ls Uncertain	Understands and says Neutral	Certain	Neither Very Certain
Eating Level of Certainty	Understand Very Uncertain	ls Uncertain	Understands and says Neutral	Certain	Neither Very Certain
Eating Level of Certainty Reading	Understand Very Uncertain Understand	ls Uncertain ls	Understands and says Neutral Understands and says	Certain	Neither Very Certain Neither
Eating Level of Certainty Reading Level of Certainty	Understand Very Uncertain Understand Very Uncertain	ls Uncertain ls Uncertain	Understands and says Neutral Understands and says Neutral	Certain	Neither Very Certain Neither Very Certain
Eating Level of Certainty Reading Level of Certainty	Understand Very Uncertain Understand Very Uncertain	ls Uncertain ls Uncertain	Understands and says Neutral Understands and says Neutral	Certain	Neither Very Certain Neither Very Certain

Running	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Sitting	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Sleeping	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Swimming	Understan	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
Waving	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
Dog/Doggie	Understar	nds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
Boy	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Girl	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Apple	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Ball	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

BalloonUnderstandsUnderstands and saysNeither	
---	--

г

Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain
Bed	Understand	s	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Bib	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Blanket	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Block	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Book	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Bowl	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Chair	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Cookie	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain U	Uncertain	Neutral	Certain	Very Certain
Cup	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain U	Uncertain	Neutral	Certain	Very Certain

Certainty

Door	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Duck	Understand	ls	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Fish	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Hat	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Pants	Understand	ds	Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Pig	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Shirt	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Shoe	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Slide	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Sock	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Spoon	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

Truck	Understands		Understands and says		Neither
Level of Certainty	Very Uncertain	Uncertain	Neutral	Certain	Very Certain

BIBLIOGRAPHY

BIBLIOGRAPHY

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Author.
- Arunachalam, S., & Luyster, R. J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research*, 9, 810–828. https://doi.org/10.1002/aur.1590
- Barbaro, J., & Dissanayake, C. (2012). Developmental profiles of infants and toddlers with autism spectrum disorders identified prospectively in a community-based setting. *Journal of autism and developmental disorders*, *42*(9), 1939-1948.
- Barneveld, P. S., Swaab, H., van Engeland, H., & de Sonneville, L. (2014).Cross-Sectional Evidence for a Decrease in Cognitive Function With Age in Children With Autism Spectrum Disorders?. *Autism Research*, 7(5), 527-534.
- Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye tracking as a measure of receptive vocabulary in children with autism spectrum disorders. *Augmentative and Alternative Communication*, *30*(2), 147-159.
- Centers for Disease Control and Prevention. (2020, April). *Autism prevalence rises in communities monitored by CDC*. <u>https://www.cdc.gov/media/releases/2020/p0326-</u> <u>autism-prevalence-rises.html</u>
- Chapman, R. (1978). Comprehension strategies in children. *Language and speech in the laboratory, school, and clinic*, 309-327.
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language*, *30*(1), 213–236. https://doi.org/10.1017/S0305000902005482
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd Ed. Hillsdale, NJ: Erlbaum.
- Davidson, M. M., & Ellis Weismer, S. (2017). A discrepancy in comprehension and production in early language development in ASD: Is it clinically relevant? *Journal of Autism and Developmental Disorders*, 47, 2163–2175. https://doi.org/10.1007/s10803-017-3135-z
- DiStefano, C., & Kasari, C. (2016). The window to language is still open: Distinguishing between preverbal and minimally verbal children with ASD. *Perspectives of the ASHA Special Interest Groups*, 1(1), 4-11.

- Dunn, D. M. (2019). Peabody Picture Vocabulary Test, (5th ed). [Measurement instrument]. Bloomington, MN: NCS Pearson, Inc.
- Effect Size. (n.d.) Retrieved from

://www.psychology.emory.edu/clinical/bliwise/Tutorials/SCATTER/scatterplots/effect.h tm#:~:text=In%20psychological%20research%2C%20we%20use,A%20correlation%20 coefficient%20of%20.&text=30%20is%20considered%20a%20moderate,a%20strong% 20or%20large%20correlation

- Ellis Weismer, S., Gernsbacher, M. A., Stronach, S., Karasinski, C., Eernisse, E. R., Venker, C. E., & Sindberg, H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders*, 41(8). https://doi.org/10.1007/s10803-010-1134-4
- Ellis Weismer, Susan, Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders*, 40, 1259–1273. https://doi.org/10.1007/s10803-010-0983-1
- 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(4), 856-868.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development*, 71(2), 310–322. https://doi.org/10.1111/1467-8624.00146
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., & Al., E. (1993). MacArthur Communicative Development Inventories: User's Guide and Technical Manual. Singular Publishing.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, 59(5).
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98
- Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in preferential looking task. *Journal of Child Language*, 34(4), 701-724.
- Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., ... &

Pact Consortium. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International journal of language & communication disorders*, *45*(6), 681-690.

- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Res*, 6(6), 479–493. https://doi.org/10.1002/aur.1334
- Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and cognitive processes*, 16(2-3), 287-308.
- Koegel, L. K., Koegel, R. L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders*, 27(3), 233-243.
- Kover, S. T., McDuffie, A. S., Hagerman, R. J., & Abbeduto, L. (2013). Receptive vocabulary in boys with autism spectrum disorder: Cross-sectional developmental trajectories. *Journal of autism and developmental disorders*, *43*(11), 2696-2709.
- Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health*, 13(4), 198–206. https://doi.org/10.1111/j.1475-3588.2008.00503.x
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1
 - 4. Western Psychological Services.
- Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., & Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 49(11), 1184-1192.
- Luyster, R, Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., & Lord, C. (2009). The Autism Diagnostic Observation Schedule— Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(9), 1305–1320.
- Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of autism and developmental disorders*, *38*(8), 1426-1438.
- Luyster, Rhiannon, Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language*, *34*(3), 623–654. https://doi.org/10.1017/S0305000907008094

- Maljaars, J., Noens, I., Scholte, E., & van Berckelaer-Onnes, I. (2012). Language in low-functioning children with autistic disorder: Differences between receptive and expressive skills and concurrent predictors of language. *Journal of Autism and Developmental Disorders*, 42(10), 2181-2191.
- Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769-785.
- Miniscalco, C., Fränberg, J., Schachinger-Lorentzon, U., & Gillberg, C. (2012). Meaning what you say? Comprehension and word production skills in young children with autism. *Research in Autism Spectrum Disorders*, 6(1), 204-211.
- Morisset, C. E., Barnard, K. E., & Booth, C. L. (1995). Toddlers' language development: Sex differences within social risk. *Developmental Psychology*, *31*(5), 851.
- Mullen, E. M. (1995). Mullen Scales of Early Learning. AGS edition ed.
- Muller, K., & Brady, N. (2016). Assessing Early Receptive Language Skills in Children with ASD. *Perspectives of the ASHA Special Interest Groups*, 1(1). https://doi.org/10.1044/persp1.SIG1.1
- Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E. (2013). Inter-rater reliability of parent and preschool teacher ratings of language in children with autism. *Research in Autism Spectrum Disorders*, 7(11), 1391–1396. https://doi.org/10.1016/j.rasd.2013.08.006
- Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E. (2014). Language assessment in children with autism spectrum disorder: Concurrent validity between report-based assessments and direct tests. *Research in Autism Spectrum Disorders*, 8(9), 1100-1106.
- Pan, B. A., Rowe, M. L., Spier, E., & Tamis-Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. J. Child Lang.
- Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Quesionnaire*. Western Psychological Services.
- Seol, K. I., Song, S. H., Kim, K. L., Oh, S. T., Kim, Y. T., Im, W. Y., ... & Cheon, K. A. (2014). A comparison of receptive-expressive language profiles between toddlers with autism spectrum disorder and developmental language delay. *Yonsei Medical Journal*, 55(6), 1721-1728.
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 335–364). Wiley and Sons.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension

is still a problem. *Monographs of the Society for Research in child Development*, 59(5), 174-179.

- Venker, C. E., Haebig, E., Edwards, J., Saffran, J. R., & Weismer, S. E. (2016). Brief report: Early lexical comprehension in young children with ASD: Comparing eyegaze methodology and parent report. *Journal of autism and developmental disorders*, 46(6), 2260-2266.
- Volden, J., Smith, I. M., Szatmari, P., Bryson, S., Fombonne, E., Mirenda, P., ... & Georgiades, S. (2011). Using the preschool language scale, to characterize language in preschoolers with autism spectrum disorders. *American Journal of Speech-Language Pathology*.
- Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders*, 34(5), 473–493. https://doi.org/10.1007/s10803-004-2544-y
- Williams, K. T. (1997). Expressive Vocabulary Test. Circle Pines, MN: American Guidance Service.
- Woynaroski, T., Yoder, P., & Watson, L. R. (2016). Atypical cross-modal profiles and longitudinal associations between vocabulary scores in initially minimally verbal children with ASD. *Autism Research*, 9(2), 301-310.
- Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, 6(1), 59–64.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scales, Fifth Edition*. The Psychological Corporation.