American Journal on Intellectual and Developmental Disabilities

Patterns of Cognition, Communication and Adaptive Behavior in Children with Developmental Disabilities

--Manuscript Draft--

Manuscript Number:	AJIDD-D-20-00063R1
Article Type:	Research Report
Keywords:	developmental disabilities; language; cognition; adaptive behavior
Corresponding Author:	Phebe Albert Georgia State University Atlanta, UNITED STATES
First Author:	Phebe Albert
Order of Authors:	Phebe Albert
	MaryAnn Romski
	Rose Sevcik
	Robin Morris
Manuscript Region of Origin:	UNITED STATES
Abstract:	Young children with developmental disabilities (DD) exhibit a range of strengths and weaknesses in cognitive, language, and adaptive skills. Identifying individual patterns of abilities across these domains is important for informing interventions. This study examines how 129 toddlers with significant developmental delays and less than 10 spoken words perform across different developmental domains and assessment methods (i.e., caregiver report and clinician-administered tests). Children exhibited statistically and clinically meaningful strengths and weaknesses across developmental domains, which may have important implications for differential interventions. Caregiver-reported and clinician-rated measures of cognition, language and adaptive functioning were highly related. However, the relation between caregiver report and clinician ratings was weaker for a subgroup of children with relatively more limited expressive language compared to other children in the sample.

Abstract

Young children with developmental disabilities (DD) exhibit a range of strengths and weaknesses in cognitive, language, and adaptive skills. Identifying individual patterns of abilities across these domains is important for informing interventions. This study examines how 129 toddlers with significant developmental delays and less than 10 spoken words perform across different developmental domains and assessment methods (i.e., caregiver report and clinician-administered tests). Children exhibited statistically and clinically meaningful strengths and weaknesses across developmental domains, which may have important implications for differential interventions. Caregiver-reported and clinician-rated measures of cognition, language and adaptive functioning were highly related. However, the relation between caregiver report and clinician ratings was weaker for a subgroup of children with relatively more limited expressive language compared to other children in the sample.

Keywords: Developmental disabilities, Language, Cognition, Adaptive behavior

Patterns of Cognition, Communication and Adaptive Behavior in Children with Developmental Disabilities

Young children with developmental disabilities (DD) can experience persistent difficulties across multiple domains of functioning (Shevell, Majnemer, Platt, Webster, & Birnbaum, 2005). Previous research also indicates that young children with DD can demonstrate unique profiles of strengths and weaknesses across different domains of development that may be important for planning interventions (Caselli et al., 1998; Fidler, Philofsky, & Hepburn, 2007; Luyster, Seery, Talbott, & Helen, 2011; Singer Harris, Bellugi, Bates, Jones, & Rossen, 1997; Weismer, Lord, & Esler, 2010). However, these differences may be subtle, or attenuated as a result of how abilities are assessed. In a longitudinal study of toddlers (mean age 12 months), scores on the *Mullen Scales of Early Learning* (MSEL; Mullen, 1995) Early Learning Composite (ELC) did not show decreases over time (from 13 to 30 months), despite significant decreases over time in individual MSEL subdomain scores (Ben-Sasson & Gill, 2014), suggesting increases in other subdomain areas. These findings underscore the importance of examining each subdomain performance; such broad developmental scores as the MSEL ELC may conceal the appearance of strengths and weaknesses across different subdomains of functioning that provide valuable information for intervention planning (Milne & McDonald, 2015).

Developmental Skills Performance Profiles

Akshoomoff (2006) reported that ability profiles of children with DD vary across subdomains of developmental functioning. Akshoomoff (2006) identified specific profiles of performance across subdomains on the MSEL in children ages 16 to 43 months with autism spectrum disorder (ASD), i.e., relatively stronger fine motor ability accompanied by relatively weaker receptive language ability. Other research findings suggest that toddlers with global developmental delays and ASD demonstrate a language profile of slightly greater receptive than expressive language abilities (Calandrella & Wilcox, 2000; Weismer et al., 2010). Previous studies also have revealed specific patterns of strengths and weaknesses across subdomains of the *Vineland Adaptive Behavior Scales* (Vineland ABS; Sparrow, Balla, Cicchetti, & Doll, 1984) for children with ASD and intellectual and developmental disabilities (IDD) (Burack & Volkmar, 1992; Carpentieri & Morgan, 1996; Fenton et al., 2003; Liss et al., 2001).

While the existing literature suggests that there are profiles of relative strengths and weaknesses across developmental subdomains for children with DD, it remains difficult to ascertain whether or not these findings consistently reflect true differences in abilities due to the influence of measurement issues. For example, prior research has revealed discrepancies in the performances of children with DD on the same subdomain across different tests (Magiati & Howlin, 2001; Scattone, Raggio, & May, 2011). Plesa Skwerer et al. (2016) systematically compared multiple measures of receptive language ability in a sample of “minimally verbal” children and adolescents, five to 21 years old, with ASD and they found that receptive language performance varied significantly across the different assessment methods.

Assessment Challenges

Our ability to characterize profiles of developmental strengths and weaknesses in young children with DD is influenced by the quality of the measurement tools used (Ben-Sasson & Gill, 2014; Plesa Skwerer et al., 2016). Unfortunately, conducting reliable and valid assessments in young children with DD remains an ongoing challenge. Some of these challenges include the psychometric limitations of the tests available for young children, particularly floor effects, and the influence of deficits in one domain, such as motor skills, on a child’s performance in other domains, such as nonverbal cognition that requires motor skills to perform some test items

(Bradley-Johnson, 2001; Brady, Anderson, Hahn, Obermeier, & Kapa, 2014; DeVeney, Hoffman, & Cress, 2012). Test performances also can be impacted by problem behaviors, e.g. tantrums and non-compliance that are common in toddlers with DD (Baker, Blacher, Crnic, & Edelbrock, 2002; Hauser-Cram & Woodman, 2016; Keller & Fox, 2009; Krakow & Kopp, 1983). Prior research has documented difficulties in gaining and maintaining young children's attention and motivation throughout testing, and problems with comprehension of test instructions (Tager-Flusberg, 2000). These issues are amplified for children with limited spoken language skills and those under the age of three (Bradley-Johnson, 2001; Gerken, Eliason, & Arthur, 1994). Best practices for developmental assessment of young children birth through age three is not well represented in the current literature (Brito et al., 2019). It is pertinent that we investigate test performance patterns in young children to help inform best practices for developmental assessment in research and clinical work. One strategy that has been implemented to improve the overall accuracy of developmental assessments is a multi-method approach that provides information about specific skills from different sources (Joint Statement by the NAEYC and the NAECS/SDE, 2003). This method allows us to make hypotheses about test score inaccuracies due to extraneous factors or test limitations when there are performance discrepancies between sources, and alternatively, to feel more confident about score accuracy when performances are consistent across sources.

Types of Assessments

Two types of measures that can be part of a multi-method assessment are 1) direct observation measures that are administered by trained clinicians, and 2) caregiver report measures. Clinician-administered assessments provide a structured protocol in which trained professionals explicitly elicit skills and behaviors from a child. This form of assessment benefits from standardized administration and the interpretation of a trained professional. However, the validity of these measures are impacted by a child's attention and engagement during the testing session, severity of developmental delay, and discomfort with the unfamiliar testing environment or examiner (Plesa Skwerer et al., 2016). Alternatively, caregiver report measures can account for a child's performance across contexts and throughout daily life. However, these measures can be biased by inaccurate reporting, which may be particularly problematic when caregivers report about the comprehension skills of their young children (Charman, 2004; Scattone et al., 2011).

While some research findings suggest strong correlations between caregiver report and clinician-administered measures of language and adaptive skills in toddlers with DD (Björn, Kakkuri, & Leppänen, 2014; Dale, 1991; Scattone et al., 2011), the agreement between measurement types is sometimes inconsistent across different subdomains being measured. For example, Björn et al. (2014) found that in a sample of 12 to 18 month old children, there was a significant relation between the number of *words understood* on the *MacArthur-Bates Communicative Developmental Inventory* (MCDI; Fenson et al., 2007) and receptive language scores on the *Bayley Scales of Infant Development*, *3rd edition* (BSID-III; Bayley & Reuner, 2006), but Bayley receptive language scores were not significantly correlated with other receptive language variables from the MCDI, e.g., *first signs of understanding instructions and questions*. Other studies also have revealed discrepancies between caregiver report and clinician-administered measures for receptive versus expressive language ability, with stronger correspondence for the latter (Charman, 2004; Lyytinen, Laakso, Poikkeus, & Rita, 1999). Taken together, these findings indicate that the agreement between caregiver report and clinician-administered testing may not be consistent across different subdomains of abilities. Inconsistencies across measurement types can be informative for identifying areas of functioning that may need follow-up or more nuanced assessment, and thus makes valuable contributions to the process of establishing an accurate picture of a child's abilities. For example, if a child's parent-reported spoken language abilities were significantly greater than those observed by a clinician during a testing session, it may warrant follow-up from the clinician to assess setting variables during testing that could have impacted performance.

Limitations of the Current Literature

A major weakness in the current literature on ability profiles in toddlers with DD is that children with very limited spoken language ability are largely underrepresented. This is to a great extent because many standardized measures require children to use spoken language to communicate test responses, which often leads to the exclusion of young children who do not yet have spoken words (Brady et al., 2004). A reliance on spoken language to measure communication in such children is likely masking important individual differences that are present even in children who are not yet speaking (Cirrin & Rowland, 1985). Floor effects on many measurements also may contribute to this misconception by creating an inaccurate picture of little to no differentiation in the various abilities of these individuals (see Soorya, Leon, Trelles, & Thurm, 2018). Therefore, it is important that we attempt to broaden the current literature on developmental assessment performance in young children to include those with DD and limited spoken language.

The purpose of the current study is to examine performances across a range of developmental abilities on clinician-administered and caregiver report measures in a sample of toddlers with a general DD and less than 10 spoken words. Specifically, three questions are addressed: 1) what are the individual patterns of performance across developmental subdomains in this sample of toddlers? 2) what are the overall group patterns of performance across developmental subdomains? and 3) what is the level of agreement between types of assessment measures (i.e., caregiver report and clinician-administered)? Regarding individual patterns of performance across developmental subdomains, we hypothesized that participants would demonstrate at least some relative strengths and weaknesses across subdomains. Regarding group patterns of performance, we hypothesized that most performances across developmental subdomains would be significantly correlated and that there would be strong correspondence between measurement types (clinician, caregiver), but the agreement would be stronger for measures of expressive language and motor skills than for measures of receptive language.

Method

Participants

The current study included a sample of 129 children, mean chronological age = 29.77 months, *SD* = 5.04 months, range 21 to 48 months, with a general developmental delay and severe spoken language impairment, operationally defined as "a vocabulary of at most 10 intelligible spoken words and a score of less than 12 months on the Expressive Language Scale of the MSEL" (in Romski et al., 2010) and at least some beginning intentional communication ability, i.e. primitive vocalizations and gestures that refer to or request objects/events in the environment. We used 10 words and a score of less than 12 months on the MSEL because the children were very young, and we wanted to ensure that the children were not beginning to use speech as their primary means of communication. Spoken language was assessed through direct observation at the onset of the first testing session while the MSEL was being administered, and through discussion with parents about words used at home. Most children had no spoken words in their vocabulary prior to participating in the study. Some of the children had been exposed to manual signs, but none of the children had access to augmentative and alternative communication (AAC) devices prior to their participation. Children's receptive communication abilities varied with some performing in the "Very Low" range and others in the "Average" range according to the MSEL. All children's primary language was English. Medical diagnoses included seizures, genetic conditions (e.g., Down syndrome), cerebral palsy, and other unknown conditions. Children with primary etiologies of autism, hearing impairment/deafness or a speech and language delay without general developmental disability were excluded from the study. Vision and hearing acuity were within normal limits. Some children had a twin sibling with special needs who did not participate in the study.

Children were assessed as part of their initial inclusion in one of two sequential longitudinal studies investigating the effectiveness of caregiver-implemented early language interventions (Romski et al., 2010; Romski et al., *2020*). Participants for both studies were recruited from various local professionals that had experience working with children with DD within the metropolitan Atlanta area. The current sample consisted of 40 children of African American background (31%), 74 children of Caucasian background (57%), 10 children of Asian background (8%) and five children with a Multi-racial background (4%). Four caregivers identified their children as Hispanic (3%), 124 children were identified as Non-Hispanic (96%), and one child's caregiver did not report the child's ethnicity. The sample contained more male (N = 90, 70%) than female participants (N = 39, 30%).

For the studies, one caregiver for each child was selected to participate according to who could commit to participate in all of the study sessions. Caregivers who were selected to participate included 11 fathers and 118 mothers. Caregivers' mean age was 36 years, ranging from 21 to 45 years. Caregiver race included 37 individuals of African American background (29%), 74 of Caucasian background (57%), nine of Asian background (7%) and the remaining nine did not report their race. Additionally, 118 caregivers identified as Non-Hispanic (92%), three identified as Hispanic (2%), and eight did not report their ethnicity. Caregivers' education levels varied, one caregiver did not attend high school (0.8%), nine caregivers graduated from high school but did not attend college (7%), 18 caregivers completed at least some college (14%), 59 caregivers had a bachelor's degrees (46%), 39 caregivers had graduate or professional degrees (30%) and three caregivers did not report their education background. Caregivers' work statuses also varied and included professionals (e.g., lawyers, doctors, accountants) and full-time stay-at-home parents.

Procedures

All caregivers provided consent for themselves and their children to participate in these studies. A trained, certified speech-language pathologist with 5-13 years of experience and specific training in working with children with severe disabilities and limited speech administered a developmental assessment battery to each caregiver-child pair. The assessment included caregiver report and clinician-administered measures of general developmental level, adaptive behavior, communication, visual-spatial and motor skills. During testing, one clinician was present in the room throughout the session. For children who separated without distress, their caregiver was not present in the room during testing but was able to observe via a one-way viewing window from an adjacent room. For children who did not separate easily, their caregiver remained in the room. To maximize children's overall engagement, the length of testing sessions was limited (i.e., no more than one hour) and evaluations were completed over the course of five to six separate sessions. Frequent breaks and snacks were provided as needed, and the testing environment was enhanced to permit each child to demonstrate their full range of skills, e.g., if a child refused to remain seated, testing would be conducted on the floor. Time was allotted for each child to become familiarized with the testing room and caregivers were permitted to provide support throughout the testing sessions. Clinicians also utilized re-direction strategies including the use of test kit manipulatives to help maintain engagement during the testing sessions, as well as removing all other toys from the testing table that could be distracting. Testing was stopped if a child was not engaged, e.g. crying. Overall, clinicians indicated having little difficulty engaging children in testing with the use of re-direction strategies and test manipulatives, frequent breaks, and support as needed from caregivers. Therefore, although we do not have a direct standardized measure of child engagement, we believe that the test results in this study are a valid representation of the participant's abilities. Additionally, to maximize the accuracy of information provided by caregivers, a speech-language pathologist with expertise in child development and who directly observed each child asked follow-up questions as necessary while conducting caregiver interviews and reviewing caregiver questionnaires.

Measures

The measures used in this study were part of a larger assessment battery administered to all caregiver-child pairs and included the following: the MSEL (Mullen, 1995), the Vineland ABS or the *Vineland Adaptive Behavior Scales, Second Edition* (Vineland-II; Sparrow, Cicchetti, & Balla, 2005), the MCDI: *Words and Gestures* (Fenson et al., 2007), and the *Sequenced Inventory of Communication Development* (SICD; Hedrick, Prather, & Tobin, 2002). The MCDI: *Words and Gestures* form was completed by caregivers at home, and then reviewed in-person with a speech-language pathologist. See Table 1 for more information about each measure. We selected these measures because they have each been widely used in research and clinical assessment of young children with developmental disabilities to evaluate overall level of developmental functioning, as well as domain-specific areas of functioning like language (e.g., Brady, Marquis,

& McLean, 2004; Calandrella & Wilcox, 2000; Dale, 1991; Gleason & Coster, 2012; O'Toole & Fletcher, 2010; Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016; Swineford, Guthrie, & Thurm, 2015).

Two different versions of the Vineland survey interview form were used in this study because caregiver-child pairs participated in one of two longitudinal studies; the first of those studies began prior to the development of the Vineland-II and therefore those caregivers were administered the Vineland ABS, survey interview form. Caregivers participating in the second study were administered the Vineland-II, survey interview form. A speech-language pathologist introduced the Vineland ABS/II to each caregiver as an interview about day-to-day functioning across several domains. A semi-structured interview format using open-ended questions, as is described in the Vineland ABS and Vineland-II manuals was used to conduct the interviews. Primary modifications made to the Vineland-II included additional items used to address the following areas: ability to start and maintain conversations and spoken language skills in the Expressive and Receptive Language subdomains, independent living skills in the Daily Living Skills subdomain, and use and comprehension of nonverbal communication during social interactions in the Socialization subdomain. In the Vineland-II manual, large correlations were reported between subdomain standard scores for a sample of 24 children birth through age two, and 29 children three through six years between the Vineland ABS and Vineland-II. Mean differences between subdomain scores across the versions were small, except for the Daily Living Skills subdomain (Vineland ABS SS mean = 78.7, Vineland-II SS mean = 87.8). In general, mean subdomain scores from the Vineland-II were slightly higher than those from the Vineland ABS.

Results

Average performances of the toddlers in this sample fell at least two standard deviations below the mean on all five subdomains of the MSEL, and between one and a half and two standard deviations below the mean across all subdomains of adaptive functioning measured by the Vineland ABS/II. Additional descriptive statistics for the MSEL, Vineland ABS/II, SICD and MCDI are reported in Table 2.

Individual strengths and weaknesses across developmental domains

To investigate individual patterns of performances across developmental domains, the data were initially inspected for outliers. Boxplots were created to investigate potential ($z > 1.96$), probable ($z > 2.58$) and extreme ($z > 3.29$) outliers. 49 (38%) children demonstrated at least one score on a measure of cognition, language or adaptive functioning that was defined as an outlier compared to the performances of the other toddlers in this sample. The majority of these children demonstrated one outlier (N = 26, 20%), with only seven (5%) demonstrating outliers across three or more test subdomains (see Table 3).

The standard error of differences (SE_{diff}) was calculated (Coaley, 2014) to evaluate individual patterns of strengths and weaknesses across developmental subdomains. Because expressive language ability was restricted in this sample, i.e., all children produced no more than 10 spoken words for study inclusion, the SE_{diff} was calculated by comparing MSEL Receptive Language (RL), Fine Motor (FM), Gross Motor (GM) and Visual Reception (VR) scores to MSEL Expressive Language (EL) scores. Differences between individual participants' MSEL EL T-scores and T-scores on the other MSEL subdomains that were equal to two and three SE_{diff} units are reported in Table 4. There were relative weaknesses in EL compared to the other subdomains, except for three individuals, all with a diagnosis of Down syndrome (DS), who demonstrated stronger EL than GM ability (two with a difference of two SE_{diff} units, one with a difference of three SE_{diff} units).

Group patterns of performance across developmental domains

A scatterplot of overall developmental level (MSEL ELC) and overall adaptive functioning (Vineland ABS/II Adaptive Behavior Composite [ABC]) revealed a roughly positive linear relation, as shown in Figure 1. A simple linear regression revealed that performance on the MSEL ELC was significantly linearly related to performance on the Vineland ABS/II ABC, R^2 = .23, t = 14.86, p < .001. Visual inspection and a local regression estimation line (LOESS) of the scatterplot also revealed a potentially curvilinear relation between the MSEL ELC and Vineland ABS/II ABC. A hierarchical multiple regression revealed that while a significant linear relation was present between the MSEL ELC and Vineland ABS/II ABC, a significant amount of additional variance was explained by adding a quadratic parameter to the regression equation, ΔR^2 = .06, p = .001.

Visual inspection of the MSEL ELC and Vineland ABS/II LOESS (see Figure 1) revealed substantial variability in Vineland ABS/II ABC scores for children at different overall developmental levels as defined by their MSEL ELC standard scores. Children (N = 70; 54%) in the “Very Low” range according to the MSEL (i.e., ELC SS ≤ 55) demonstrated caregiver-reported adaptive skills on the Vineland ABS/II ranging from a standard score of 44 to 83. Children (N = 7; 5%) in the “Average” range according to the MSEL (i.e., ELC SS ≥ 85) also demonstrated a range of Vineland ABS/II ABC standard scores from 60 to 88.

Pearson’s *r* correlations revealed moderate to large positive correlations between most cognitive, language and adaptive functioning performances between and within tests, except performances on motor subdomains (see Tables 5 and 6). Unexpectedly, the Expressive

Language (EL) subdomain of the MCDI was not significantly correlated with the Communication scale reported on the Vineland ABS/II.

Analysis of MCDI EL and Vineland ABS/II RL and EL by study. To further explore the non-significant correlation between MCDI EL and the Vineland ABS/II Communication scale, MCDI EL scores were correlated with Vineland ABS and Vineland II Receptive Language (RL) and Expressive Language (EL) subdomain raw scores, which are the two subdomains that make up the Communication scale. Because different versions of the Vineland were used across the two studies included in this analysis, and raw scores were not equally scaled across all participants for the EL and RL subdomains, correlations were run separately for participants in study 1 (N = 74) and study 2 (N = 55). This analysis revealed large, significant correlations between MCDI EL and Vineland-II RL and EL, $r = .48$, $p >.001$, and $r = .61$, $p >.001$, respectively for participants from the second study. However, correlational analysis for participants in the first study revealed a small, significant relation between MCDI EL and Vineland ABS RL, $r = .32$, $p = .007$, and a small, non-significant correlation between MCDI EL and Vineland ABS EL, $r = .22$, $p = .07$.

Relationship between caregiver report and clinician-administered measures by study

Paired samples t-tests were conducted to evaluate the correspondence between children's performances on a clinician-administered test (i.e., the MSEL) and a caregiver report interview (i.e., the Vineland ABS/II) across four developmental subdomains: gross motor (GM), fine motor (FM), receptive language (RL) and expressive language (EL). Because these subdomains are not comparable across the two Vineland versions, t-tests were run separately for each study. To allow for statistical comparisons across measures, z-scores were calculated. For both studies, there were no significant differences between caregiver-reported and clinician-assessed abilities, except within the GM subdomain, $t = -2.99$, $p = .004$ for study 1, and $t = 3.72$, $p < .001$ for study 2. For study 1, the average z-score for clinician-administered GM performance was lower than for caregiver-reported GM skills. For study 2, the reverse was true.

Effect sizes were generally large (Cohen, 1988, 1992) for the correlations between caregiver-reported (i.e., Vineland ABS/II) and clinician-evaluated (i.e., MSEL) GM, FM, RL and EL skills for study 1, $r = .84$, .72, .74, and .42, and study 2, $r = .87$, .69, .77, and .57. However, the effect sizes for EL were notably smaller than the effect sizes for RL and motor subdomains. To further investigate this finding, the sample was divided into two groups: children whose performance fell above or below the median MSEL EL raw score. For study 1, paired samples t-tests revealed a significant difference between caregiver report and clinician-administered assessment of EL in both groups; children who performed above the median (Group 1), $t = 4.91$, $p < .001$, and children who performed below the median (Group 2), $t = -2.01$, $p = .05$. For children who performed above the median (Group 1), average clinician-reported EL was greater than caregiver-reported EL, however EL performances were still moderately correlated, $r = .40$, $p = .007$. For children who performed below the median (Group 2), caregiver-reported EL was greater than clinician-assessed EL, and the correspondence was non-significant and weaker, $r = .24$, $p = .21$. For study 2, paired samples t-tests revealed a significant difference between caregiver-reported and clinician-administered assessment of EL in Group 2, $t = -2.74$, $p = .01$, but not for children in Group 1, $t = 0.89$, $p = .38$. Children who performed below the median (Group 2) demonstrated stronger caregiver-reported than clinician-administered EL abilities, which were not significantly correlated, $r = .18$, $p = .37$. Although not significantly different, children in Group 1 demonstrated slightly higher clinician-assessed than caregiver-reported EL abilities. Overall, for children in both studies with higher EL abilities, clinicians tended to report higher EL scores than caregivers, and for children with lower EL abilities, caregivers reported significantly higher EL abilities than clinicians.

Discussion

The purpose of this study was to investigate the performance profiles of toddlers with a general DD and limited spoken language across multiple subdomains of functioning (receptive language, fine motor, gross motor, visual reception, expressive language), and between types of measures (caregiver ratings, clinician-administered tests). Analysis of outliers revealed that approximately one third of the toddlers in this sample demonstrated an outlier in their performance on at least one developmental subdomain relative to the average group performance. Of these children, 14% demonstrated outliers across three or more subdomains relative to the mean sample performances. Approximately half of the toddlers in this sample demonstrated a significant difference between their EL ability and their performance in another developmental domain, i.e., most commonly a relative strength in RL or FM skills compared to EL (see Table 4). These findings highlight the presence of significant variability and strengths and weaknesses across different subdomains of functioning in very young children with DD and little to no spoken language. This finding is in line with previous research emphasizing the importance of considering variability in performance profiles across subdomains of children with DD, which is necessary for effective treatment planning (Scattone et al., 2011).

Unexpectedly, three toddlers in this sample exhibited significantly stronger EL than GM skills. This finding may be best understood by considering these children’s diagnosis of Down syndrome (DS). Some research suggests that early motor skill acquisition is delayed in individuals with DS (Palisano et al., 2001; Pereira, Basso, Lindquist, Silva, & Tudella, 2013). Regarding language, many individuals with DS demonstrate stronger RL than EL skills starting in early toddlerhood (Fidler et al., 2007; Miller & Miller, 1999). However, early language milestones, e.g. first single word acquisition, tends to be achieved at similar rates compared to mental-age matched peers without DS (Chapman, 1997; Fowler, 1990). The three toddlers in this study with DS and relatively stronger EL than GM abilities did demonstrate EL that was delayed but on par with their nonverbal mental ages. However, there was a total of 29 children in this sample with a diagnosis of DS, so 26 children who also had DS did not exhibit this profile of strengths and weaknesses, highlighting again the importance of considering individual profiles of abilities when evaluating young children with DD, even those with a shared genetic syndrome (Mervis & Robinson, 1999; Mervis, Robinson, Levy, & Schaeffer, 2003).

Regarding group performance patterns, average performances across developmental domains were highly related. Initial whole sample analysis of relations between overall developmental level and adaptive functioning suggested that performances in these domains are related in a generally linear fashion. This is in line with previous research findings demonstrating moderate correlations between performance on cognitive and adaptive functioning measures in children with DD (Frost, Hong, & Lord, 2017; Ray-Subramanian, Huai, & Ellis Weismer, 2011). Further analysis revealed a more complex, curvilinear relation between overall developmental level and adaptive functioning. For children with the lowest overall developmental level (as measured by the MSEL ELC), there was substantial variability in caregiver-reported levels of adaptive functioning. These findings could reflect true differences in adaptive functioning skills among children at similar developmental levels. However, differences this large in adaptive functioning would not be expected for toddlers performing this similarly with regard to overall developmental level. Previous research suggests a strong relation between level of neurodevelopmental impairment and adaptive functioning measured by the Vineland ABS

(Fidler, Hepburn, & Rogers, 2006; Ross & Weinberg, 2006). It is possible that the relation between developmental level and adaptive functioning in this sample was influenced by measurement challenges. Accurate caregiver report of adaptive functioning may be especially difficult to ascertain for children with significant disabilities whose independent adaptive skills are limited. Caregivers may work more to meet their children's needs in order to accomplish daily tasks, and thus have more difficulty, or less opportunities, to assess independent versus assisted abilities. Conversely, clinician-administered tests are limited to information collected during relatively short periods of time in an unfamiliar environment, which can be a particularly challenging setting for young children with DD (Baker et al., 2002; Hauser-Cram & Woodman, 2016), resulting in a possible underestimation of true abilities for some children. It is also possible that the floor effects of the MSEL contributed to this finding, such that actual variability in developmental skills between children scoring at or below the lowest limit of the test was truncated. This may have resulted in the appearance of a range of adaptive skills in a group of children with a similar MSEL ELC, but who actually demonstrate differences in developmental skills not captured by the MSEL.

Our findings also revealed strong, positive relations between most subdomains measured on cognitive, language and adaptive functioning measures. One deviation to this pattern was the presence of small, non-significant correlations between motor and language skills. It has been well-documented that early motor skills play an influential role in the development of language (Iverson, 2010). Several large scale reviews have confirmed that motor difficulties often co-occur in children with language impairments (Hill, 2001; Rechetnikov & Maitra, 2009). However, the relation between motor and language development is complex, and much remains to be understood about how these domains interact in children with developmental disabilities

(Hill, 2001). Our findings support the idea that the relation between these domains is not entirely straightforward and may not be linear across all children with DD, or at all stages of development.

Regarding the correspondence between caregiver report and clinician-administered measures, our findings demonstrated a strong overall correspondence between these two types of measures. This is in line with previous research indicating strong agreement between caregiver-reported and clinician-administered measures in children with DD (Björn et al., 2014; Dale, 1991; Scattone et al., 2011; Weismer et al., 2010). Our hypothesis that relations between caregiver-reported and clinician-administered abilities would be stronger for EL and motor subdomains, and weaker for RL was partially confirmed. Relations between measurement types for GM and FM skills were the strongest, but the relation between measurement types was weaker for EL than RL, which was true for children who received both versions of the Vineland (i.e. Vineland ABS and Vineland II). Further analysis between two subgroups of toddlers (i.e., relatively lower vs. higher EL as measured by the MSEL) revealed that caregiver report of EL abilities for children in the lower EL subgroup was significantly higher than clinicians' ratings. One possible explanation for this finding is the limited item range and sensitivity of the MSEL for children with very limited spoken language skills, i.e. many of the children in this sample were only administered a few items on the EL subdomain before they reached a ceiling. Therefore, the MSEL may not have captured the same quality of and variation in spoken language abilities that could be captured by parent report on the Vineland ABS/II. Alternatively, it is possible that caregivers of children with little spoken language have more difficulty describing their children's expressive language level because examples of these skills occur less frequently than for children with more spoken language, making it more challenging to accurately report. Further, because these skills occur less frequently, caregivers may observe them while clinicians do not, so caregiver ratings are influenced by these rare occurrences that evaluators do not observe. Average caregiver-reported and clinician-administered EL was also significantly different for the higher EL subgroup (statistically significant for study 1 only), with average caregiver report being lower than clinician ratings in both studies. However, in study 1 caregiver-reported and clinician-administered EL for the subgroup of children with higher EL were significantly, moderately correlated, unlike for the subgroup of children with lower EL. In study 2, caregiver-reported and clinician-administered measures were not significantly correlated for either EL subgroup, but for the subgroup of children with higher EL, the correlation was larger. Taken together, these results suggest that caregiver and clinician ratings of EL for toddlers with fewer EL skills tend to correspond less. It is important to acknowledge that some differences between measurement types is likely related to differences in the types of skills captured by each tool, i.e., adaptive day-to-day expressive language use across multiple contexts captured by the Vineland ABS/II compared to specific developmental expressive language skills elicited and observed within a structured setting on the MSEL. However, both tools are used clinically to describe EL abilities and therefore, regardless of the exact contribution of type of reporter versus skills captured by each tool, the important take home message is that significant differences can occur, particularly for children with the most limited EL skills. Consequently, these findings underscore the importance of conducting multi-method assessments of language; especially for children with very limited or no spoken words, as information from only one source is more likely to be an under- or over-estimate of a child’s abilities.

There are several limitations of this study. First, a comparison group, such as children with DD and 11 to 50 spoken words, may have been useful for drawing conclusions about the specificity of the developmental characteristics observed in the current sample of toddlers with DD and less than 10 spoken words. However, designing appropriate comparison groups for individuals with DD is challenging, especially for toddlers whose abilities are rapidly changing in nonlinear ways during early development (Mervis & Robinson, 1999; Mervis et al., 2003).

Second, because this study includes data from two time-separate projects, a different version of the Vineland was administered across studies. This resulted in some challenges for combining information across studies including different relations between the Vineland and other measures in study 1 compared to study 2. It is possible that these discrepancies reflect true differences in the samples of toddlers from the separate projects, or psychometric differences between the Vineland versions, e.g. additional items in the EL subscale on the Vineland II. It is important to take this limitation into consideration when interpreting Vineland findings.

Third, other psychometric limitations of the assessments used in this study should be acknowledged. Gleason and Coster (2012) reported that despite good overall correspondence with other levels of functioning indicators, results from measures like the Vineland-II should be interpreted cautiously in children with specific communication, sensory and/or motor deficits as they may impact scores across developmental subdomains in unique ways. The MSEL also has been critiqued for low test-retest reliability coefficients (< .80) for young children 25 - 56 months of age, and steep scoring gradients on items for children 20 months of age or performing at a 20-month level of ability (Bradley-Johnson, 2001). While we can't quantify the exact influence of possible measurement related errors on our findings, the excellent clinical experience of the speech-language pathologists who administered the tests enhanced our ability to obtain the most accurate descriptions of our participants as possible. We do, however, encourage readers to interpret our findings within the context of possible measurement error, particularly in those children with the most limited EL. Less consistent findings, such as the correspondence between clinician and caregivers on EL measures may reflect this influence of measurement error. Therefore, EL may require careful interpretation, and consideration of non-traditional assessment tools that may be more sensitive for this population (see Brio et al., 2019 for examples), once they are shown to improve such measurement characteristics. At the same time, there are robust patterns that we identified which can be interpreted more confidently, including the presence of significant, systematic variability in individual performances between developmental subdomains, strong correlations between subdomains for the whole group, and strong correspondence between parent and clinician report for most domains and subdomains.

Despite these limitations, findings from this study make several meaningful contributions to the field. Many of the toddlers who participated in this study demonstrated consistent, highly related levels of performance across developmental subdomains. A substantial number of toddlers, however, exhibited statistically and clinically significant strengths and weaknesses. This finding supports previous literature that found meaningful profiles of strengths and weaknesses in children with DD, such as ASD, and extends those findings to a group of children with DD and severe spoken language delays. The profiles of strengths and weaknesses we identified in this study have broader implications for clinical practice and research. Clinically, these profiles are important to consider because of the influence they may have on treatment planning and response to interventions. These findings underscore that practitioners should be sensitive to the unique needs of young children with minimal to no spoken words, who may vary significantly with regard to their strengths and weaknesses in other domains of functioning. With regard to research, these findings highlight the importance of careful selection of study samples when investigating young children with DD. The presence of significant strengths and weaknesses across important developmental domains should be considered as it may result in extremely heterogenous samples of young children, possibly complicating interpretation of results for specific research questions. Future studies should also investigate how individual differences across developmental abilities relate to response to interventions, and whether or not these early performance patterns in toddlers with DD remain consistent or change over time.

This study also underscores the importance of multi-method assessment for evaluation of toddlers with DD. Our findings suggest that for children with the lowest EL ability, caregiver-reported and clinician-assessed outcomes across cognitive, language and adaptive functioning subdomains may not be as strongly related as for children with more EL skills. Single-reporter assessment should be interpreted cautiously for children with very severe expressive language impairments. Based on the current findings, we recommend that clinicians working with young children with very limited spoken language administer multiple tests of developmental functioning, and look for patterns of strengths and weaknesses that are consistent across measures, while putting less weight in those scores that suggest a strength or weakness on only one measure. Researchers may also consider relying on patterns of scores rather than single scores to describe the developmental abilities of very young children with limited spoken language, and/or should consider the use of alternative measures (e.g., those described in Brio et al., 2019) when studying this population.

In conclusion, this study revealed meaningful differences in individual patterns of strengths and weaknesses across developmental subdomains, and between types of assessment measures in a group of children with limited spoken language and DD.

References

A Joint Position Statement of the National Association for the Education of Young Children and The National Association of Early Childhood Specialists in State Departments of Education (2003). *Early Childhood Curriculum, Assessment, and Program Evaluation Building an Effective, Accountable System in Programs for Children Birth through Age 8*. Retrieved September 22, 2020 from: https://www.naeyc.org/sites/default/files/globally-shared/downloads/PDFs/resources/position-statements/pscape.pdf

Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology*, *12*(4–5), 269–277.

Baker, B. L., Blacher, J., Crnic, K. A., & Edelbrock, C. (2002). Behavior Problems and Caregivering Stress in Families of Three-Year-Old Children With and Without Developmental Delays. *American Journal on Mental Retardation*, *107*(6), 433. http://doi.org/10.1352/0895-8017(2002)107<0433:BPAPSI>2.0.CO;2

Bayley, N., & Reuner, G. (2006). *Bayley scales of infant and toddler development: Bayley-III* (Vol. 7). Harcourt Assessment, Psych. Corporation San Antonio, Tex, USA.

Ben-Sasson, A., & Gill, S. V. (2014). Motor and language abilities from early to late toddlerhood: using formalized assessments to capture continuity and discontinuity in development. *Research in Developmental Disabilities*, *35*(7), 1425–1432.

Björn, P. M., Kakkuri, I., & Leppänen, P. H. T. (2014). The covariation between caregiveral and expert evaluations of early language skills. *Early Child Development and Care*, *184*(5), 706–718.

Bradley-Johnson, S. (2001). Cognitive assessment for the youngest children: A critical review of tests. *Journal of Psychoeducational Assessment*, *19*(1), 19–44.

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 (Baltimore, Md. : 1985)*, *30*(2), 147–59. http://doi.org/10.3109/07434618.2014.904923

Brady, N. C., Marquis, J., Fleming, K., & McLean, L. (2004). Prelinguistic predictors of language growth in children with developmental disabilities. *Journal of Speech, Language, and Hearing Research*, *47*(3), 663–677.

Brito, N. H., Fifer, W. P., Amso, D., Barr, R., Bell, M. A., Calkins, S., Flynn, A., Montgomery-Downs, H. E., Oakes, L. M., & Richards, J. E. (2019). Beyond the Bayley: neurocognitive assessments of development during infancy and toddlerhood. *Developmental Neuropsychology*, *44*(2), 220–247.

Bruckner, C., Yoder, P., Stone, W., & Saylor, M. (2007). Construct validity of the MCDI-I receptive vocabulary scale can be improved: Differential item functioning between toddlers with autism spectrum disorders and typically developing infants. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1631–1638.

Burack, J. A., & Volkmar, F. R. (1992). Development of low- and high-functioning autistic children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *33*(3), 607–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1374418

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: a comparison of statistical risk models. *Developmental Psychology*, *36*(6), 793.

Calandrella, A. M., & Wilcox, M. J. (2000). Predicting language outcomes for young prelinguistic children with developmental delay. *Journal of Speech, Language, and Hearing*

Research, *43*(5), 1061–1071.

Carpentieri, S., & Morgan, S. B. (1996). Adaptive and intellectual functioning in autistic and nonautistic retarded children. *Journal of Autism and Developmental Disorders*, *26*(6), 611–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8986847

Caselli, M. C., Vicari, S., Longobardi, E., Lami, L., Pizzoli, C., & Stella, G. (1998). Gestures and words in early development of children with Down syndrome. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1125–1135.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proceedings of the National Academy of Sciences*, *100*(9), 5567–5572.

Chapman, R. S. (1997). Language development in children and adolescents with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, *3*(4), 307–312. http://doi.org/10.1002/(SICI)1098-2779(1997)3:4<307::AID-MRDD5>3.0.CO;2-K

Chapman, R. S., & Hesketh, L. J. (2000). l. *Mental Retardation and Developmental Disabilities Research Reviews*, *6*(2), 84–95.

Charman, T. (2004). Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders*, *34*(1), 59–64.

Cirrin, F. M., & Rowland, C. M. (1985). Communicative assessment of nonverbal youths with severe/profound mental retardation. *Mental Retardation*, *23*(2), 52–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3990553

Coaley, K. (2014). *An introduction to psychological assessment and psychometrics*. Sage.

Dale, P. S. (1991). The validity of a caregiver report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research, 34*(3), 565–571.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay I: Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language and Hearing Research, 46*(June 2003), 544–560.

DeVeney, S. L., Hoffman, L., & Cress, C. J. (2012). Communication-based assessment of developmental age for young children with developmental disabilities. *Journal of Speech, Language, and Hearing Research : JSLHR*, *55*(3), 695–709. http://doi.org/10.1044/1092-4388(2011/10-0148)

Fenson, L. (2007). *MacArthur-Bates Communicative Development Inventories: Words and Gestures, Desktop Scannable*. Brookes.

Fenton, G., D'ardia, C., Valente, D., Del Vecchio, I., Fabrizi, A., & Bernabei, P. (2003). Vineland adaptive behavior profiles in children with autism and moderate to severe developmental delay. *Autism, 7*(3), 269–287.

Fidler, D. J., Hepburn, S., & Rogers, S. (2006). Early learning and adaptive behaviour in toddlers with Down syndrome: evidence for an emerging behavioural phenotype? *Down's Syndrome, Research and Practice : The Journal of the Sarah Duffen Centre, 9*(3), 37–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16869373

Fidler, D. J., Philofsky, A., & Hepburn, S. L. (2007). Language phenotypes and intervention planning: Bridging research and practice. *Mental Retardation & Developmental Disabilities Research Reviews, 13*(1), 47. http://doi.org/10.1002/mrdd.20132

Fodstad, J. C., Matson, J. L., Hess, J., & Neal, D. (2009). Social and communication behaviours in infants and toddlers with autism and pervasive developmental disorder-not otherwise specified. *Dev Neurorehabil*, *12*(3), 152–157. http://doi.org/10.1080/17518420902936748

Fowler, A. E. (1990). Language abilities in children with Down syndrome: evidence for a specific syntactic delay. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective* (pp. 302–328). Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511581786.010

Frost, K. M., Hong, N., & Lord, C. (2017). Correlates of Adaptive Functioning in Minimally Verbal Children With Autism Spectrum Disorder. *American Journal on Intellectual and Developmental Disabilities*, *122*(1), 1–10. http://doi.org/10.1352/1944-7558-122.1.1

Gerken, K. C., Eliason, M. J., & Arthur, C. R. (1994). The assessment of at-risk infants and toddlers with the Bayley Mental Scale and the Battelle Developmental Inventory: Beyond the data. *Psychology in the Schools, 31*(3), 181–187. http://doi.org/10.1002/1520-6807(199407)31:3<181::AID-PITS2310310302>3.0.CO;2-Q

Gleason, K., & Coster, W. (2012). An ICF-CY-based content analysis of the Vineland Adaptive Behavior Scales-II. *Journal of Intellectual and Developmental Disability*, *37*(4), 285–293.

Hauser-Cram, P., & Woodman, A. C. (2016). Trajectories of internalizing and externalizing behavior problems in children with developmental disabilities. *Journal of Abnormal Child Psychology*, *44*(4), 811–821.

Hedrick, D. L., Prather, E. M., & Tobin, A. R. (2002). *Sequence Inventory of Comminication Development* (Revisd Edi). Los Angeles, CA: Western Psychological Services.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11344592

Hill, E. L. (2010). Motor difficulties in specific language impairment: evidence for the Iverson account? - a commentary on Iverson's "Developing language in a developing body: the relationship between motor development and language development". *Journal of Child Language*, *37*(2), 287–92. http://doi.org/10.1017/S0305000909990444

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261.

Keller, K. M., & Fox, R. A. (2009). Toddlers with developmental delays and challenging behaviors. *Early Child Development and Care*, *179*(1), 87–92. http://doi.org/10.1080/03004430600987126

Krakow, J. B., & Kopp, C. B. (1983). The Effects of Developmental Delay on Sustained Attention in Young Children. *Child Development*, *54*(5), 1143. http://doi.org/10.2307/1129670

Laws, G., & Hall, A. (2014). Early hearing loss and language abilities in children with Down syndrome. *International Journal of Language & Communication Disorders*, *49*(3), 333–42. http://doi.org/10.1111/1460-6984.12077

Liss, M., Harel, B., Fein, D., Allen, D., Dunn, M., Feinstein, C., … Rapin, I. (2001). Predictors and correlates of adaptive functioning in children with developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 219–230.

LORD, C., & PICKLES, A. (1996). Language Level and Nonverbal Social-Communicative Behaviors in Autistic and Language-Delayed Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(11), 1542–1550. http://doi.org/10.1097/00004583-

199611000-00024

Luyster, R. J., Seery, A., Talbott, M. R., & Helen, T.-F. (2011). IDENTIFYING EARLY-RISK MARKERS AND DEVELOPMENTAL TRAJECTORIES FOR LANGUAGE IMPAIRMENT IN NEURODEVELOPMENTAL DISORDERS. *Developmental Disabilities Research Reviews*, *17*(2), 151–159. http://doi.org/10.1002/ddrr.1109

Lyytinen, P., Laakso, M., Poikkeus, A., & Rita, N. (1999). The development and predictive relations of play and language across the second year. *Scandinavian Journal of Psychology*, *40*(3), 177–186.

Magiati, I., & Howlin, P. (2001). Monitoring the progress of preschool children with autism enrolled in early intervention programmes: problems in cognitive assessment. *Autism : The International Journal of Research and Practice*, *5*(4), 399–406. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11777256

Mervis, C. B., & Robinson, B. F. (1999). Methodological Issues in Cross-Syndrome Comparisons: Matching Procedures, Sensitivity (Se), and Specificity (Sp). *Monographs of the Society for Research in Child Development*, *64*(1), 115–130. http://doi.org/10.1111/1540-5834.00011

Mervis, C. B., Robinson, B. F., Levy, Y., & Schaeffer, J. (2003). Methodological issues in cross-group comparisons of language and cognitive development. *Language Competence across Populations: Toward a Definition of Specific Language Impairment*, 233–258.

Miller, J. F., & Miller, J. F. (1999). Profiles of language development in children with Down syndrome. *Improving the Communication of People with Down Syndrome*, 11–39.

Milne, S., & McDonald, J. (2015). Assessing adaptive functioning in preschoolers referred for diagnosis of developmental disabilities. *Infants & Young Children*, *28*(3), 248–261.

Mullen, E. M. (1995). *Mullen scales of early learning*. AGS Circle Pines, MN.

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.

O'Toole, C., & Fletcher, P. (2010). Validity of a parent report instrument for Irish-speaking toddlers. *First Language*, *30*(2), 199–217.

Palisano, R. J., Walter, S. D., Russell, D. J., Rosenbaum, P. L., Gémus, M., Galuppi, B. E., & Cunningham, L. (2001). Gross motor function of children with down syndrome: Creation of motor growth curves. *Archives of Physical Medicine and Rehabilitation*, *82*(4), 494–500. http://doi.org/10.1053/APMR.2001.21956

Pereira, K., Basso, R. P., Lindquist, A. R. R., Silva, L. G. P. da, & Tudella, E. (2013). Infants with Down syndrome: Percentage and age for acquisition of gross motor skills. *Research in Developmental Disabilities*, *34*(3), 894–901. http://doi.org/10.1016/J.RIDD.2012.11.021

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591–604. http://doi.org/10.1177/1362361315600146

Ray-Subramanian, C. E., Huai, N., & Ellis Weismer, S. (2011). Brief report: adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*, *41*(5), 679–84. http://doi.org/10.1007/s10803-010-1083-y

Rechetnikov, R. P., & Maitra, K. (2009). Motor impairments in children associated with impairments of speech or language: a meta-analytic review of research literature. *The American Journal of Occupational Therapy : Official Publication of the American*

Occupational Therapy Association, *63*(3), 255–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19522134

Romski, M. A., Sevcik, R. A., Hyatt, A. M., & Cheslock, M. (2002). A continuum of AAC language intervention strategies for beginning communicators. *Exemplary practices for beginning communicators: Implications for AAC*, 1-23.

Romski, M., Sevcik, R. A., Adamson, L. B., Cheslock, M., Smith, A., Barker, R. M., Bakeman, R. (2010). Randomized comparison of augmented and nonaugmented language interventions for toddlers with developmental delays and their parents. *Journal of Speech, Language, and Hearing Research*, *53*(2), 350–364.

Romski, M. A., Sevcik, R. A., Adamson, L. B., Barton-Hulsey, A., Smith, A., Barker, M., Whitmore, A., & Bakeman, R. (2020). *Parent-Coached Early Augmented Language Interventions for Toddlers with Developmental Delays: A Randomized Comparison*. Manuscript in preparation.

Ross, G., & Weinberg, S. (2006). Is there a relationship between language delays and behavior and socialization problems in toddlers? *Journal of Early Childhood and Infant Psychology*, *2*, 101–116. Retrieved from http://ezproxy.gsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-14172-008&site=ehost-live&scope=site

Scattone, D., Raggio, D. J., & May, W. (2011). Comparison of the Vineland Adaptive Behavior Scales, Second Edition, and the Bayley Scales of Infant and Toddler Development, Third Edition. *Psychological Reports*, *109*(2), 626–34. http://doi.org/10.2466/03.10.PR0.109.5.626-634

Shevell, M., Majnemer, A., Platt, R. W., Webster, R., & Birnbaum, R. (2005). Developmental and Functional Outcomes at School Age of Preschool Children With Global Developmental Delay. *Journal of Child Neurology*, *20*(8), 648–654. http://doi.org/10.1177/08830738050200080301

Singer Harris, N. G., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with williams and down syndromes. *Developmental Neuropsychology*, *13*(3), 345–370. http://doi.org/10.1080/87565649709540683

Soorya, L., Leon, J., Trelles, M. P., & Thurm, A. (2018). Framework for assessing individuals with rare genetic disorders associated with profound intellectual and multiple disabilities (PIMD): the example of Phelan McDermid Syndrome. *The Clinical Neuropsychologist*, *32*(7), 1226–1255. http://doi.org/10.1080/13854046.2017.1413211

Sparrow, S. S., Balla, D. A., Cicchetti, D. V, & Doll, E. A. (1984). *Vineland adaptive behavior scales: Survey form manual*. American Guidance Service.

Sparrow, S. S., Cicchetti, D. V, & Balla, D. A. (2005). *Vineland adaptive behavior scales* (2nd editio). Circle Pines, MN: American Guidance Service.

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

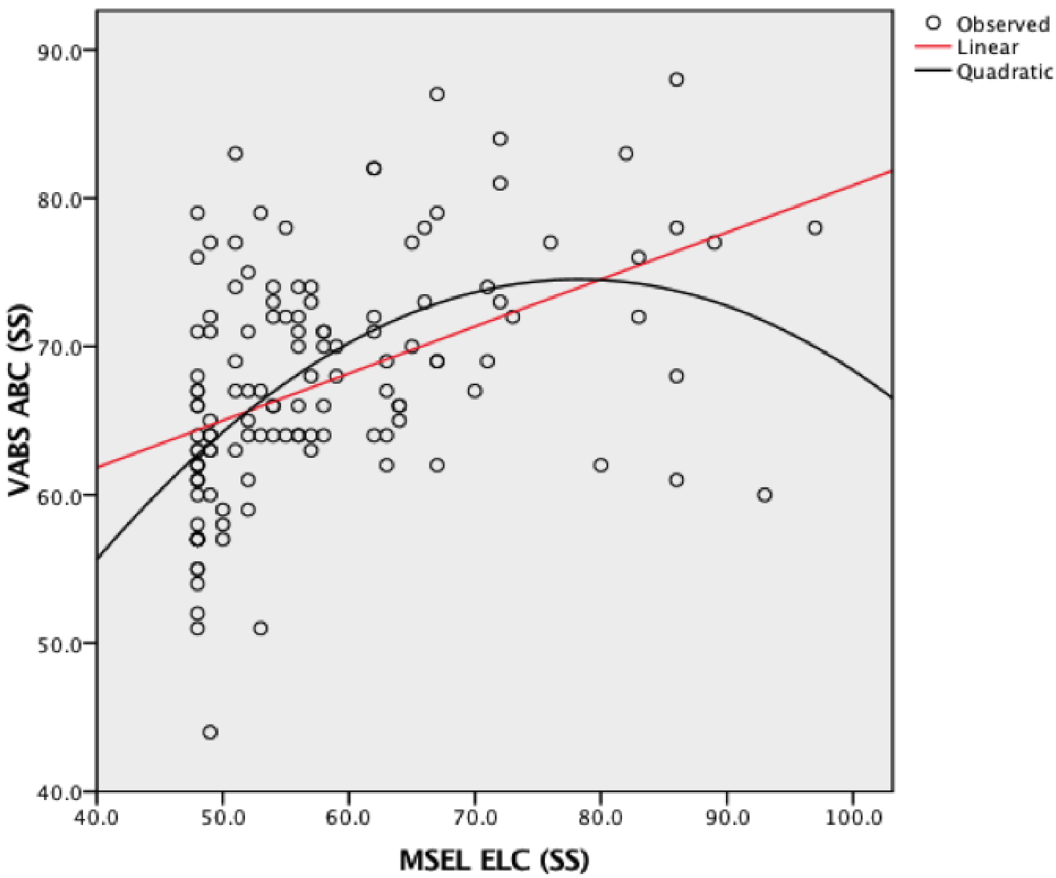
Tager-Flusberg, H. (2000). The challenge of studying language development in children with autism. In L. Menn, N. Bernstein Ratner, L. (Ed) Menn, & N. (Ed) Bernstein Ratner (Eds.), *Methods for studying language production.* (pp. 313–332). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. Retrieved from http://ezproxy.gsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1999-16036-014&site=ehost-live&scope=site

Waterhouse, L., & Fein, D. (1982). Language skills in developmentally disabled children. *Brain and Language*, *15*(2), 307–333. http://doi.org/10.1016/0093-934X(82)90062-1

Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders, 40*(10), 1259–1273.

Wilkins, J., & Matson, J. L. (2009). A comparison of social skills profiles in intellectually disabled adults with and without ASD. *Behav Modif*, *33*(2), 143–155. http://doi.org/10.1177/0145445508321880

Figure 1. Scatterplot of MSEL ELC and Vineland ABS/II ABC



Note. The scores plotted above are standard scores (M=100, SD=15). ELC = Early Learning Composite, ABC = Adaptive Behavior Composite. Linear regression line estimation is illustrated in red, $F = 37.0$, $p < .001$ quadratic regression estimation is illustrated in black, $F = 25.56$, $p < .001$.

Table 1. Measures used in this study

Measure	Domains assessed	Administration	Scores produced	Reliability and validity
MSEL	1. Overall developmental level 2. Gross motor 3. Fine motor 4. Visual reception 5. Expressive language 6. Receptive language	Clinician-administered	1. Raw scores 2. T-scores (M=50, SD=10)	Mullen (1995) reported median Cronbach's alphas for Mullen subtests between .75 and .85, test-retest correlation coefficients between .76 and .83, interscorer correlation coefficients between .91 and .99 and strong construct and concurrent validity. The MSEL has also demonstrated strong construct validity in children with ASD and other developmental delays (Swineford et al., 2015).
Vineland ABS/II	1. Overall adaptive skills 2. Motor skills 3. Daily living skills 4. Socialization 5. Communication	Semi-structured parent interview	1. Raw scores 2. *v*-scale scores 3. Standard scores (M=100, SD=15) 4. Age equivalents	Sparrow et al. (1984) reported split-half reliability coefficients ranging from .83 to .94 across all domains and slightly lower reliability coefficients for subdomains ranging from .69 to .84, test-retest reliability coefficients for children from birth through 4 years, 11 months ranging from .78 to .92, and interrrater reliability coefficients ranging from .62 to .78. They also reported strong validity illustrated by moderate correlations with other adaptive behavior measures. Similarly, Floyd et al. (2015) reported Cronbach's alphas above .80 for the Vineland-II, test-retest correlation coefficients greater than .90 and adequate interrater reliability and validity.
MCDI	1. Words understood 2. Words produced	Parent report	1. Raw scores 2. Percentiles	Fenson et al. (2007) reported Chronbach's alphas of .95 and .96 for Words Understood and Words Produced, respectively. The MCDI has also demonstrated strong concurrent validity with other language assessments (Nordahl-Hansen et al., 2014) including strong correlations with clinician-administered direct assessments of language skills, such as the Bayley Scales (Bayley & Reuner, 2006) in young children 12 to 18 months of age (Björn et al., 2014) and laboratory observations of vocabulary use (Thal, O'Hanlon, Clemmons, & Fralin, 1999).

SICD	1. Receptive language 2. Expressive language	Clinician-administered and parent report items	1. Age equivalents	Hedrick et al. (1984) reported sufficient interrater reliability on a subset of the normative sample; the average agreement between raters for whether individual items should be rated as “pass” or “fail” was 96% (range = 90% to 100%). Test-retest reliability was also sufficient; the average percentage agreement between administrations, which were approximately one week apart, was 93% (range = 88 to 99%). Correlation coefficients between the SICD and other measures of language (e.g. the PPVT) ranged from .75 to .80.

Table 2. Participant performances across assessment measures

	N	Min	Max	M	SD
MSEL					
Gross motor[1]	108	19	66	29.1	11.0
Visual reception	129	19	73	30.3	12.4
Fine motor	129	19	54	24.9	8.2
Receptive	129	19	59	26.8	11.1
Expressive	129	19	39	21.3	4.0
ELC	129	48	97	57.6	11.2
SICD					
Receptive	128	4	40	17.0	6.6
Expressive	129	3	24	12.6	4.9
MCDI					
Receptive	125	2	395	118.7	97.7
Expressive	125	0	75	8.6	12.7
Vineland ABS					
Communication SS	74	55	82	67.1	6.5
Daily Living SS	74	48	92	69.5	8.5
Socialization SS	74	55	92	72.7	8.9
Motor skills SS	74	30	111	72.5	13.0
ABC	74	44	88	65.0	7.3
Vineland II					
Communication SS	55	6	86	67.5	12.8
Daily Living SS	55	29	96	74.7	12.8
Socialization SS	55	61	93	75.9	7.9
Motor skills SS	55	52	96	77.9	10.2
ABC	55	51	87	70.9	7.5

Note. ELC = Early Learning Composite (SS); ABC = Adaptive Behavior Composite (SS). T-scores are reported for each MSEL subscale (M = 50, SD = 10). Age equivalents are reported for both SICD subscales. Raw scores are reported for MCDI Receptive = number of words understood, and MCDI Expressive = number of spoken words.

[1]20 children do not have MSEL gross motor SS's because they were 34 months of age or older and normative data for this subscale is not available for children over 33 months of age, one child's MSEL did not include this subscale because it was administered by an outside psychologist.

Table 3. Number of types of outliers across measures

Measure	Potential (%)	Probable (%)	Extreme (%)
MSEL			
Gross Motor	5 (3.9)	0 (0)	1 (0.8)
Fine Motor	7 (5.4)	1 (0.8)	1 (0.8)
Visual Reception	4 (3.1)	0 (0)	2 (1.6)
Receptive Language	4 (3.1)	4 (3.1)	0 (0)
Expressive Language	4 (3.1)	3 (2.4)	2 (1.6)
Vineland ABS/II			
Communication	1 (0.8)	1 (0.8)	0 (0)
Daily Living Skills	4 (3.1)	2 (1.6)	1 (0.8)
Socialization	6 (4.7)	0 (0)	0 (0)
Motor Skills	0 (0)	2 (1.6)	1 (0.8)
MCDI			
Receptive	4 (3.1)	3 (2.3)	0 (0)
Expressive	2 (1.6)	2 (1.6)	3 (2.3)
SICD			
Receptive	4 (3.1)	2 (1.6)	1 (0.8)
Expressive	5 (3.9)	0 (0)	0 (0)

Note. Potential, Probable and Extreme refer to the levels of outliers. Percentages represent the percentage of outliers present from all children who were administered each measure.

Table 4. Number of participants demonstrating relative differences between MSEL Expressive Language performance and performances on other subdomains

MSEL subdomain	Two SE_{diff} units	Three SE_{diff} units
Gross Motor	38	25
Fine Motor	25	21
Visual Reception	48	31
Receptive Language	34	18

Note. SE_{diff} = standard error of differences. SE_{diff} scores were calculated using the following equation: square root $[(SE_{m1})^2 + (SE_{m2})^2]$, where SE_{m1} and SE_{m2} are the standard errors of measurement for score 1 (e.g., MSEL Expressive Language) and score 2 (e.g., MSEL Receptive Language).

Table 5. Correlations across assessment measures

	MSEL					SICD	
	GM	VR	FM	RL	EL	RL	EL
MSEL GM	1.00						
MSEL VR	0.27**	1.00					
MSEL FM	0.66**	0.57**	1.00				
MSEL RL	0.13	0.65**	0.38**	1.00			
MSEL EL	0.27**	0.23**	0.36**	0.37**	1.00		
SICD RL	0.11	0.61**	0.37**	0.87**	0.38**	1.00	
SICD EL	0.31**	0.30**	0.36**	0.44**	0.68**	0.44**	1.00
MCDI RL	0.14	0.43**	0.32**	0.62**	0.28**	0.64**	0.36**
MCDI EL	0.21*	0.18*	0.29**	0.30**	0.55**	0.28**	0.45**
VABS Com	-0.06	0.27**	0.08	0.46**	0.22*	0.41**	0.36**
VABS DLS	0.32**	0.29**	0.35**	0.21*	0.24**	0.20*	0.29**
VABS Soc	0.07	0.40**	0.18*	0.40**	0.25**	0.39**	0.32**
VABS MS	0.66**	0.18*	0.56**	-0.03	0.15	-0.02	0.17

Note. Raw scores were used for the MSEL and MCDI; age equivalents were used for the SICD; standard scores were used for the VABS = Vineland ABS/II domains. GM = Gross Motor, VR = Visual Reception, FM = Fine Motor, RL = Receptive Language, EL = Expressive Language. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Table 6. Correlations across assessment measures cont'd

	MCDI		Vineland			
	RL	EL	Comm	DLS	Soc	MS
MCDI RL	1.00					
MCDI EL	0.44**	1.00				
VABS Com	0.32**	0.14	1.00			
VABS DLS	0.19*	0.16	0.50**	1.00		
VABS Soc	0.24**	0.09	0.65**	0.57**	1.00	
VABS MS	0.03	0.13	0.23**	0.52**	0.30**	1.00

Note. Raw scores were used for the MCDI; standard scores were used for the VABS = Vineland ABS/II domains. RL = Receptive Language, EL = Expressive Language, Comm = Communication, DLS = Daily Living Skills, Soc = Socialization, MS = Motor Skills. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Comments	Response
Reviewer 1	
Introduction & Abstract	
Please clarify in the abstract which subgroup of toddlers is being referred to when noting "the relation between caregiver report and clinician ratings was weaker for children with limited expressive language". Since toddlers were required to have fewer than 10 spoken words to be included, it could be said that the entire sample has limited expressive language.	We edited this sentence in the abstract for clarity.
The final sentence referring to "clinically meaningful strengths and weaknesses in other developmental domains…" is not well represented in the results and discussion. This does not seem to be take home message of the paper. However, if it is, this should be clarified in the results and discussion sections.	We have edited the abstract so that it better reflects the results and discussion sections. We have also more explicitly referred to clinically and statistically meaningful strengths and weaknesses across developmental domains throughout the results and discussion sections, as this is one, but not the only take-home message of this study.
The introduction could be streamlined and then enhanced by a clearer argument for the necessity of the research questions posed in this study.	We have reorganized the introduction and made edits to provide a clearer argument for the necessity of the research question.
How will looking at agreement amongst measures address the assessment challenges and profiles of performance that are highlighted as significant barriers in the introduction?	We have described how agreement amongst measures can be used to address the assessment challenges we describe.
Methods	
Participants: Please provide justification for why 10 words was the cut point for inclusion. There is a growing literature regarding the term "minimally verbal" (typically referring to age 5 and up but can be extended downward when exploring children who are "pre verbal") which sets the bar at less than	We have added a justification for the use of 10 words as the cut point on p. 7. We reviewed the Tager-Flusberg & Kasari (2013) article and there is a range of number of words that can be used as a cut point from 0 to 20 in the studies cited in that article, which includes the Romski et al. (2010) paper.

20 spontaneous functional spoken words (e.g., Tager-Flusberg & Kasari, 2013).	
How was the benchmark of less than 10 words measured? The authors mentioned "observed functional use"- how is functional use defined and how was the observation conducted?	We have modified and expanded on our operational definition of "observed functional use" of 10 spoken words and how it was observed.
Please provide more information about baseline communication characteristics for the children.	We have provided additional information about communication characteristics.
Please clarify the diagnoses of the included children. The discussion speaks to three children with Down Syndrome. Are there only three children with DS? Hard to understand this portion of the discussion without greater context about the sample.	We added additional information about the diagnoses of the included children. We also clarified the number of children with Down syndrome (N = 29, p. 17)
Assuming that the choice of measures was made based on those available from the prior two studies, however, can the authors provide some rationale as to why examination of these specific measures is important to advance research and clinical assessment?	Yes, our choice of measures was made largely based on what was available from the prior studies. However, we have added information in the measures section about why investigating these measures makes a significant contribution to clinical assessment and research.
Please provide details on the training and experience of the assessors. To the authors' point, this population can be particularly tricky to engage in standardized assessments. Understanding who the assessors are, their training with this population, and tools they may have used to support the children's engagement would help to increase confidence in the accuracy of the scores.	We provided additional information about the assessors.
Please also describe how the VABS and the MCDI were conducted. What information/ instruction was provided to the caregivers?	We have edited the measures section to include how the VABS and MCDI were conducted.
Was the VABS conducted as an interview? Questionnaire?	The VABS was conducted as a semi-structured interview, we have added this information.

Please clarify which MCDI form was administrated? If different forms were administered across participants please clarify how that was dealt with.	Only the *Words and Gestures* form was used, we have added this to the method section.
Are there measures that would speak to the child's ability to engage in the assessments? In the introduction the authors speak to the presence of challenging behavior, difficulties sustaining attention, etc. as potential barriers to test administration. Is there any information to characterize these variables in this sample?	We did not administer any formal, standardized measures of engagement during the testing sessions. However, we have added additional information about engagement strategies utilized by clinicians to maintain engagement during testing, and informal reports by clinicians that overall participants were able to be engaged well during the testing sessions.
What else is known about the demographics of these families. Are there other variables that might influence a caregivers' accuracy of reporting? For those who have had other children with special needs and used these measures before or participated in psychoeducational or behavioral skills training interventions, it's very possible that they may a different view on their children's skills than first time parents or parents who have not had access to these types of services.	We added caregiver age, ethnicity, work status, and information about siblings with special needs. We also commented on procedures to maximize the accuracy of caregiver-report.
Results & Discussion	
The results section on the top of page 11 implies subgroup analyses, please clearly indicate (perhaps with subheadings) where subgroup analyses are being conducted and how those groups were defined.	The analysis on the top of page 11 was run for participants for study 1 and study 2 separately due to different Vineland versions that were administered in each study. We have added a subheading and additional description of how the analysis was conducted for clarification.
Considering the limitations of application of standardized cognitive and language measures with young children who have developmental delays/disorders and who have very limited or no spoken language, how is the reader to interpret these findings? How heavily are the findings potentially influenced by measurement error and all the factors that the authors mention in the introduction in terms	We have added further discussion of the influence of measurement error on our findings, and implications for interpretation.

of the difficulties engaging this population in order to understand if we have an accurate read of the children's skills or not?	
It's unclear how the reader is to interpret the significance of the discrepancy between clinician administered tests and caregiver report. If we remove the notion of reporter, is it expected that for example, expressive vocabulary on the MCDI would be higher correlated to the much wider range of communication skills assessed in the communication subscale of the VABS-II or expressive language subscale of the MSEL? How do we understand how much of these findings is related to type of reporter vs. the skills captured by the respective tools?	We have added an explanation of how to interpret the findings regarding discrepancies between clinician administered and caregiver report tests on p. 20.
What are the broader clinical implications of this work? What is the take away for the reader in clinical practice and what is the take away for the reader working in a research setting using these types of measures?	We have more clearly described the broader clinical and research-related implications of our findings on p. 22.
When you look to the raw scores that a clinician would use, how different are they? Do they provide different clinical pictures of this child? If yes, what do you suggest the clinician do with that information since the recommendation is to move forward with multiple modes of assessment?	It was difficult for us to answer this question because we are unclear what scores the reviewer is talking about when they say "raw scores that a clinician would use," however we have added recommendations for clinicians when scores between reporters is inconsistent.