Summer 7-30-2021

Measuring Language Development in Children with Down Syndrome Who Use AAC

Ji Sun Park

Ji Sun Park
Candidate

Speech and Hearing Sciences
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Cathy A. Binger, Chairperson

Jessica D. Richardson

Cindy Gevarter

MEASURING LANGUAGE DEVELOPMENT

IN CHILDREN WITH DOWN SYNDROME WHO USE AAC

BY

JI SUN PARK

BACHELOR OF ARTS

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Speech-Language Pathology

The University of New Mexico
Albuquerque, New Mexico

July, 2021

Measuring Language Development in Children with Down Syndrome Who Use AAC

by

Ji Sun Park

B.A., Linguistics, Seoul National University, 2002

M.S., Speech-Language Pathology, University of New Mexico, 2021

Abstract

Purpose: This study examined the inter-observer agreement (IOA) and within-observer agreement as well as the clinical potential of newly proposed measures that are designed to monitor language progress of children with Down syndrome who use AAC. Measures were explored based on the Graphic Symbol Utterance and Sentence Development Framework.

Method: Participants included 8 preschoolers with Down syndrome. Four graduate student observers coded 13 measures across 57 intervention sessions. Each session was coded by two observers for IOA, and all sessions were recoded for within-observer agreement. Statistical analyses were completed on utterance level and session level.

Results: Across all observers and measures, an acceptable level of IOA and within-observer agreement was achieved, even though some measures demonstrated varied data.

Conclusion: Results provided initial evidence that the new measures can be used reliably. These findings are a first step in developing psychometrically sound ways to assess communication skills in children who use AAC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Measuring Language Development in Children with Down Syndrome Who Use AAC

Developing psychometrically sound measures to monitor the language progress of young children is important for a number of reasons, such as monitoring progress, identifying the existence of a speech and language issues, and characterizing the nature and severity of those issues. For similar reasons – especially monitoring progress and characterizing various aspects of language development – assessing and analyzing language abilities of children who use augmentative and alternative communication (AAC) is equally important (Binger et al., 2020).

For decades, AAC has been used to minimize language barriers and facilitate communication of people with speech and language disorders. AAC can take many forms, including the use of unaided AAC (e.g., manual signs, gestures), low-tech aided AAC (e.g., picture cards, communication boards), and high-tech AAC (e.g., speech-generating devices/SGDs, AAC software; American Speech-Language-Hearing Association, n.d.). AAC can be used effectively to improve the communication skills of children with communication disorders in various ways. Aided AAC interventions can promote peer interactions between children with communication disorders, with benefits including increased participation and increased frequency of communicative acts (Therrien et al., 2016). Further, AAC interventions have strong effects on improving communication skills and social functions as well as decreasing challenging behaviors in children with autism spectrum disorder (Ganz et al., 2012). In addition to the pragmatic goals, AAC also can be used to promote expressive syntax and grammar (Binger et al., 2011; Binger et al., 2017).

Early aided AAC systems were developed for children and adults with motor impairments such as cerebral palsy (Shane et al., 2012). Over time, both academic and clinical communities have realized the benefits of AAC for essentially all populations with severe speech disorders and complex communication needs, such as individuals with autism, Down syndrome (DS), and other severe disabilities (Beukelman & Light, 2020).

For children who are preliterate, one of the most common forms of aided AAC is the use of graphic symbols (i.e., line drawings and photographs). Surprisingly little is known about how children's expressive communication development unfolds *over time* when they use graphic symbols to communicate. Although researchers have developed a number of ways to monitor progress on specific goals and targets (e.g., , tracking the number of turns [Therrien & Light, 2016], presence of syntactic structures [Kovacs & Hill, 2015], or use of particular semantic relations [Binger et al., 2017]), to our knowledge, more comprehensive measures of graphic symbol development have not been systematically explored and validated to date. As an initial step toward this end, Binger and colleagues introduced the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020). Within this framework, these authors suggest a range of new measures that might be explored to track children's language progress over time. This framework applies a model of typical spoken language development to demonstrate how children who use graphic symbols might also proceed through the various stages of pragmatic, semantic, syntactic, and grammar development, with adjustments made for the unique features of graphic symbol communication. For example, the authors discussed the pros and cons of adapting spoken language measures to create measures such as mean length of utterances in symbols (MLUSym), which is akin to mean length of utterance in spoken language (MLU). Other measures such as percentage of relevant symbols (PRSym) are uniquely suited for graphic symbol measurement to resolve issues that do not exist in spoken language, such as unintentional selection of graphic symbols.

Communication Patterns of Children with DS

For a number of reasons, preschoolers with DS are prime candidates for aided AAC. They typically have cognitive abilities that exceed their language and speech skills (Martin et al., 2013), significantly low speech intelligibility (Chapman & Kay-Raining Bird, 2012) which may lower the estimations of their overall communication skills, and relatively strong social skills and interaction (Martin et al., 2009). The current investigation focuses on a range of aided AAC measures with emphasis on this population. Additional details of the language and speech skills associated with DS are discussed below.

Language Function

One prominent characteristic of children with DS is that they develop communication skills at a slower rate than typically developing children. For example, one relatively large study of children with DS found that only 23% of the participants produced 50 spoken words by age three (Berglund et al., 2001); in contrast, typically developing children reach this milestone at approximately 18 months (Paul et al., 2018). Relatedly, children with DS communicate using single words and gestures for a much longer period of time before they start combining words, compared with children who are typically developing (Martin et al., 2009). These communication delays often exceed their nonverbal cognitive delays. (Caselli et al., 2008).

Further, recent studies have demonstrated that in individuals with DS, nonverbal cognitive development and language skills do not always align, with non-verbal skills typically exceeding language skills (Chapman & Kay-Raining Bird, 2012). Clearly, spoken language abilities are among the biggest challenges for children with DS, and a wide range of variables exist within the spectrum of their language function (Abbeduto et al., 2003). In general, children with DS display more difficulty in expressive language compared with receptive language, at least during early childhood (Chapman & Kay-Raining Bird, 2012). Additional findings highlight the discrepancy between nonverbal IQ and expressive language. In a study of 71 school-aged children with DS with an average chronological age of 10.5 years, their average nonverbal age equivalence was 5;5 years, with a significantly lower average expressive syntax age equivalence of 3;5 years. Thus, the expressive syntax skills of children with DS fall below expectations for their cognitive level (Martin et al., 2013).

Among the language domains, vocabulary skills are usually relatively strong, with stronger receptive than expressive skills (Laws & Bishop, 2003). Expressive language delays are present not only on standardized tests but also in conversation, including decreased size and diversity of their expressive lexicons compared with their typically developing peers (Berglund et al., 2001; Chapman & Kay-Raining Bird, 2012). Even though vocabulary is a relative strength for children with DS compared with grammatical skills, children with DS demonstrate particular difficulty in learning abstract words, such as the words relating to emotions and mental states (Chapman & Kay-Raining Bird, 2012).

Children with DS experience significant challenges in both receptive and expressive syntax and morphology. For example, school-age children in one study (*M* = 16;6; SD = 3.1) demonstrated mean grammatical comprehension age-equivalency scores at a preschool/kindergarten level, with less of a discrepancy noted in their vocabulary comprehension (Abbeduto et al., 2003). Expressive grammar skills appear to be the most prominent area of delay, with children frequently omitting grammatical function words which results in lower MLU and simpler sentence structures (Chapman & Kay-Raining Bird, 2012).

Speech Intelligibility

Many factors likely contribute to the speech sound errors and poor intelligibility in children with DS, including the structural differences such as missing or additional muscles that characterize the distinctive facial structure (Martin et al., 2009). Macroglossia – a tongue that is disproportionately large compared with the oral cavity – is also thought to impact articulation of speech sounds and lower speech intelligibility (Kent & Vorperian, 2013). Additionally, hearing loss is more common in children with DS than typically developing children, which in turn can affect phonological development.

Certainly, poor speech intelligibility contributes to lower estimations of expressive language abilities; simply put, examiners cannot give a child credit for words and sentences that they do not understand. Differences in speech sound production are apparent in infancy, with fewer vocalizations and vegetative sounds (e.g., burping, crying) compared with typically developing infants as well as a 2-month delay in canonical babbling (Chapman & Kay-Raining Bird, 2012). Phonological errors are commonly found in children with DS during preschool and school-age years (Martin et al., 2009). Delays accelerate as children grow; children with DS experience predictable speech sound errors that do not adhere exclusively to typical developmental errors, and these errors do not necessarily resolve over time (Martin et al., 2009). According to a parental survey, the majority of children and adults with DS of all ages frequently experienced difficulties with intelligibility, with over 95% having at least some difficulty in being understood (Kumin, 1994).

Pragmatics

In contrast to language form, language use is often an area of strength for children with DS. Toddlers with DS participate in social interaction almost as frequently as their typically developing peers and display adequate topic maintenance skills during mother-child conversations (Chapman & Kay-Raining Bird, 2012). Strong social skills and the friendly nature of children with DS enable them to actively participate in verbal and nonverbal interaction activities using a variety of communicative functions (i.e., comments, answers, and protests; Martin et al., 2009). Put another way, children with DS largely have the social skills needed for effective communication, but they often lack the intelligibility and spoken language skills to maximize these strengths.

AAC as a Missing Link to Richer Communication

As discussed above, many children with DS have receptive language skills that exceed their spoken expressive language, and intelligibility also mediates expressive language abilities. Additionally, severe speech and language impairments often lower listeners’ expectations of the child’s cognitive ability, even though speech and language delay is not necessarily correlated with cognitive deficits (Cleland et al., 2010). AAC bypasses the need for spoken language and therefore has the potential to help children with DS achieve their true expressive language potential.

A growing body of evidence demonstrates that even young children with DS can readily learn to use aided AAC to increase their expressive language skills, with gains noted across all language domains (Allen et al., 2017; Barker et al., 2013; Binger & Light, 2007; Kent-Walsh et al., 2010). AAC provides a platform for children with complex communication needs to develop their language just as their typically developing peers would develop their spoken language. To this end, the developmental norms and trajectories of spoken language can be used as a guideline for intervention approaches and clinical decision making for aided language development (Binger et al., 2020).

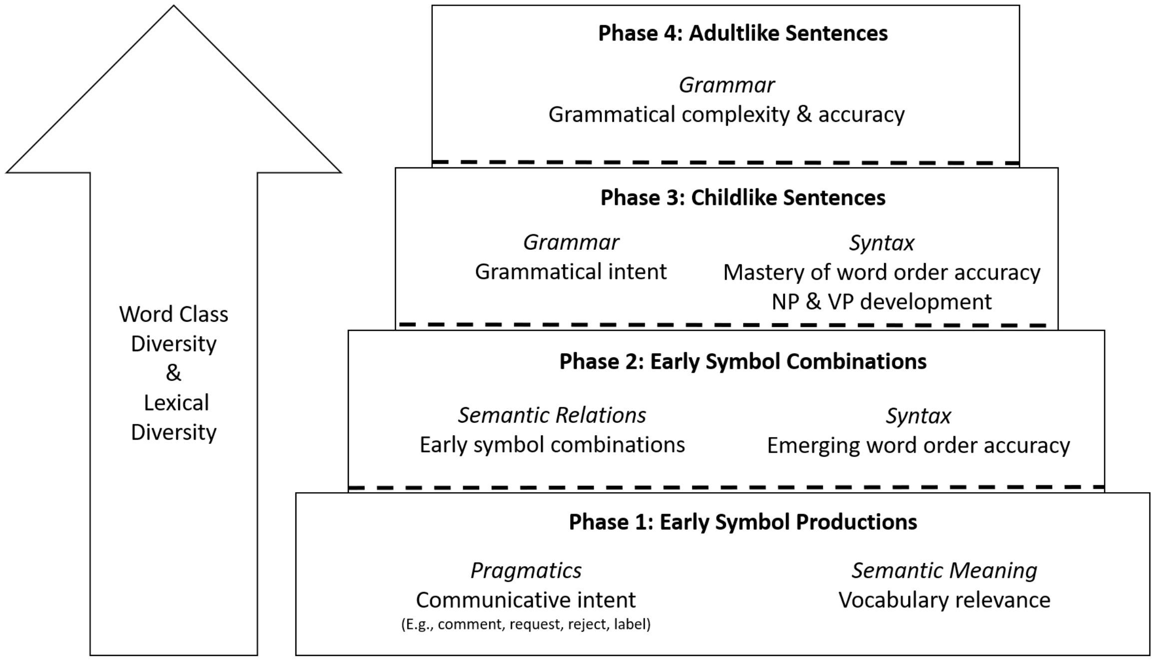
Measurement for Aided AAC

Even though aided AAC interventions have been highly successful with a wide range of populations (Ganz et al., 2012; Therrien & Light, 2016), determining the best ways to measure expressive language development progress has remained elusive. Assessing graphic symbol communication – particularly once children require more than single symbols and pre-programmed messages (e.g., one symbol selection results in the message "Hi, how are you?") – presents multiple challenges. Measuring pragmatic skills such as turn-taking rates or use of various communicative functions (requesting, rejecting, commenting, etc.) as well as tracking early semantic skills (e.g., vocabulary diversity) appears to be relatively straightforward. However, measuring sophisticated semantic skills and tracking grammar development is more complex. One approach is to borrow tools that are widely used to assess development of the spoken language, such as

MLU. However, for a variety of reasons, direct application of MLU to graphic symbol communication is problematic (Binger et al., 2020). For example, unlike spoken language in which the child typically only says what they intend to, when a child uses aided AAC, they may select symbols that are irrelevant or unintentional; this could inflate MLU. A related issue is when children search for a word they want on a communication device by selecting multiple graphic symbols until they find the one they want. For example, to say DOG IN BOX, the child might accurately select DOG, but not know which abstract line drawing for the prepositions represents IN. The child then selects each available preposition until he hears the word "in." This intended 3-symbol utterance could therefore be much longer (e.g., DOG ON UNDER ABOVE IN BOX), with only three symbols relevant to the intended utterance. In such a case, giving the child credit for a 6-symbol utterance when calculating MLU would give the child too much credit.

Binger and colleagues (2020) developed the Graphic Symbol Utterance and Sentence Development Framework to guide the development of graphic symbol utterance measurement using a model of typical language development. These authors suggested a range of measures that can be explored to track language development at various phases in development. *As an initial step toward determining the relative reliability and validity of these measures, the current investigation explores the inter-observer agreement (IOA) and within-observer agreement for a number of these measures,* as described below.

Figure 1
Graphic Symbol Utterance and Sentence Development Framework



Note: Dotted line in between layers indicates flexible boundaries.

Binger, C., Kent-Walsh, J., Harrington, N., & Hollerbach, Q. C. (2020). Tracking early sentence-building progress in graphic symbol communication. *Language, Speech, and Hearing Services in Schools*, *51*(2), 317-328.

Multi-Phase Measures

Across all phases of the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020), it is important to determine if aided utterances are completed independently or not, and also to determine if they are direct imitations of the communication partner. Reliably counting the number of symbols used in each utterance is essential to calculate other measures that are based on the number of symbols (e.g., Weighted Utterance Length). Word class diversity and lexical diversity are useful to monitor semantic and syntactic development of children who use AAC. Each of these multi-phase measures is discussed below.

Independence/ Co-Construction. How independently children who use AAC construct their utterances has long been discussed in the AAC discipline (Sutton et al., 2002), as using aided AAC presents unique challenges in this area. For example, aided communication is often grammatically incomplete, telegraphic, and missing grammatical morphemes (Savaldi-Harussi et al., 2019), and communication partners then guess or complete the intended message. This results in message co-construction; that is, the communication partner asks questions and expands upon the child's aided selections (Sutton et al., 2002), which is a unique aided AAC issue. This process obscures utterance boundaries, which causes problems when trying to analyze grammatical progress in aided communication. Each utterance, then, can be rated as "independent" (no substantial prompting from the clinician such as spoken directive or pointing to the symbol) or "co-constructed" (child receives direct prompting assistance while in the process of constructing an utterance).

Imitative/ Non-Imitative. Whether or not a child's utterance is imitative is similarly important. In general, children who rely on speech to communicate can imitate utterances that are slightly beyond their current expressive language functioning, and the frequency of imitation is a good predictor for the child's language development (Roulstone et al., 2002). This issue is all the more important when using aided AAC. When children use graphic symbols to communicate, one could argue that imitative aided utterances are not truly linguistic; they only require that the child imitate the "button pushing" of the partner. Therefore, differentiating between imitative and non-imitative utterances is important when working with children who use AAC.

MLU, MLUSym, and W-MLUSym. To measure language development over time, counting the number of symbols children use in each utterance is an obvious starting point. Conceptually, calculating the number of symbols per utterance is akin to calculating the number of words or morphemes per utterance in spoken language. In spoken language samples, the number of words or morphemes per utterance is used to calculate mean length of utterance (MLU). In the spoken child language literature, MLU is a widely used measure of grammatical development (Paul et al., 2018). Thus, one can calculate the mean length of utterance in symbols (MLUSym) using the number of symbols selected per utterance. However, as discussed above, applying the rules of spoken MLU to graphic symbol utterances presents significant issues, including the relatively high possibility of a child selecting the wrong symbols or using incorrect word order (Binger et al., 2019). One potential way to solve these issues is to develop a measure similar to MLU that takes both symbol relevance and word order into account; that is, a weighted MLU in symbols, or W-MLUSym. The present investigation includes a first attempt to explore this new W-MLUSym measure. Like the traditional MLU, W-MLUSym is based on an entire corpus of utterances.

To obtain W-MLUSym for a given corpus, a *weighted utterance length* is calculated for each utterance in a given sample. This weighted utterance length score is the product of the number of relevant symbols and word order score for each individual utterance. The number of relevant symbols is counted by determining whether or not each symbol is relevant to the context, and the word order is scored based on a 3-point scale (i.e., 1, .5, 0) to assess correctness of the order of symbols used in the utterance. The W-MLUSym for the corpus is the mean of the individual weighted utterance length scores. Note that in Table 1 below, the overall W-MLUSym is 2.0, even though in the sample, three of the four utterances contain more than two symbols. Thus, the W-MLUSym of 2.0 for this abbreviated sample reflects the relevance and word order issues present in some of the utterances. In other words, by multiplying the two measures, inflated MLUs are avoided by excluding irrelevant, inappropriate symbols, with the score (theoretically) shifting upward over time as the child improves not only the length of their utterances, but also the relevance and word order. By using this formula, longer utterances with low relevance or poor word order will earn less credit for the number of symbols compared with longer utterances with high relevance and accurate word order. The current study will examine the IOA and within-observer agreement of the weighted utterance length scores for each utterance.

Table 1

Example of Weighted Utterance Length Score for Individual Utterances and W-MLUSym for a Corpus

Utterance	Number of Symbols	Number of Relevant Symbols	Word Order Score	Weighted utterance length score
I EAT CAKE	3	3	1	3.0
I CAKE EAT	3	3	0.5	1.5
I EAT	2	2	1	2.0
I CAKE PUSH DRINK EAT	5	3	0.5	1.5
Mean Score	**MLUSym = 3.25**	**Mean no. relevant symbols = 2.75**	**Mean word order score = 0.75**	**W-MLUSym = 2.0**

Word Class and Lexical Diversity. All throughout early language development, children continue to develop their lexicons. This is depicted on the left side of Figure 1, with the arrow demonstrating the ongoing nature of semantic development. Measures of word class diversity and lexical diversity are both relevant to growth in this domain. Word class diversity refers to the child's use of various parts of speech (i.e., noun, pronoun, verb, adjective, adverb, preposition, conjunction, determiner). These data are used to track the number of different parts of speech used per session. For example, if a child had aided access to all eight parts of speech and used at least one noun, verb, pronoun, and preposition in a corpus but no other parts of speech, the word class diversity for this corpus would be 50% (i.e., 4 out of 8 parts of speech). To calculate this measure, then, the parts of speech used in each utterance must be tracked.

The other component in the arrow in Figure 1 is lexical diversity, which in spoken language is commonly measured by the number of different words (Watkins et al., 1995). In aided language, Binger and colleagues (2020) offer the corollary aided measure of number of different symbols (NDSym). NDSym was not included in the current study, but theoretically, it can be calculated in the same way as NDW; that is, by counting the number of non-repeated symbols within the sample.

Phase 1: Early Symbol Production

The first phase of the Graphic Symbol Utterance and Sentence Development Framework is the earliest phase of aided symbolic communication and is conceptualized as being akin to the first 50 words used expressively in spoken language development. One focal point of this phase is a pragmatic focus on *communicative intent*; that is, the child is learning to use graphic symbols to convey a meaningful message to another person using graphic symbols. For the current study, this is measured in a basic manner, simply indicating whether or not a clear communicative intent is present for each utterance. Communicative intent is present when the child selects graphic symbols with an intention to communicate something to the communication partner, rather than selecting random symbols with no communicative intent.

Percentage of relevant symbols (PRSym) is also likely to be most useful during Phase 1. PRSym is viewed as a measure of semantic meaning and is intended to account for the fact that when individuals use aided AAC, they have the opportunity to select symbols that are not relevant to the current context. A similar measure often used with adults who have aphasia is Correct Information Unit (CIU), which is used to measure the informativeness and efficiency of spoken language (Nicholas & Brookshire, 1993). Clinical expertise indicates that when some individuals are first learning to use graphic symbols to communicate, selecting irrelevant symbols may happen quite frequently, and this phenomenon is expected to decrease over time. This measurement, therefore, may provide clinicians with a useful way to measure early progress with graphic symbol communication; that is, the higher the PRSym for a given session, the more relevant the vocabulary is to a given context. To calculate PRSym for a given session, the percentage of relevant symbols must be calculated for each utterance. The relevance of each symbol is determined based on all available contextual information (e.g., child's object of attention, relatedness to prior topic of conversation, materials available in the environment, etc.; see Table 2).

Table 2
Example Calculations for Number of Symbols, Number of Relevant Symbols, and PRSym for the Target Utterance I EAT CAKE

Utterance	Number of Symbols	Number of Relevant Symbols	PRSym
I *PUSH CAKE	3	2	2/3 = 67%
I EAT *PLATE	3	2	2/3 = 67%
I CAKE	2	2	2/2 = 100%
TOTAL			**78%**

Phase 2: Early Symbol Combinations

During Phase 2, the child moves from simple single symbol productions into early symbol combinations, which is akin to the word combination phase for children who rely on spoken communication. Hence, children begin using early semantic relations (e.g., attribute-entity, agent-action). From a syntactic perspective, children are expected to begin to adhere to the rules of spoken word order, and in spoken language development, this is usually true (Tomasello, 2000). The main measure most relevant to this phase explored in the current study is the word order score. Just as children are apt to select graphic symbols that are not relevant to the given context, they also are far more prone to word order errors compared with children who are using spoken language (Binger et al., 2019). The word order score used in the current study is designed to account for this issue by assigning a score that indicates how accurately each utterance adheres to the rules of spoken English syntax. One of three scores is assigned to each utterance: 0 (no discernable word order is present), 0.5 (some word order is present, but at least one error exists), and 1.0 (word order is present with no definite errors; see examples in Table 3).

Table 3

Examples of Word Order Scores for the Target Utterance I EAT CAKE

Utterance	Word Order Score	Rationale
I EAT CAKE	1.0	Subject–verb–object order is present with no definite errors.
I CAKE EAT	0.5	Subject is in the correct position, but errors are present.
CAKE EAT I	0	No discernable word order is present.

Phase 3: Childlike Sentences

During this phase, the child is moving beyond early symbol combinations into true early sentences – that is, utterances that contain both a subject and a verb, however simplistic (Hadley, 2014). The measures that will be explored that originate from this phase include: (a) indicating whether or not a sentence structure is present for each utterance – that is, whether or not the utterance contains both a subject and a verb; (b) for the utterances that do qualify as sentences, determining whether or not a lexical verb is used, and (c) determining whether or not the grammatical intent is clear. The question of grammatical intent is an estimation of whether or not the child's intended message is obvious, even if the message is missing some grammatical elements. For example, COW BITE COOKIE contains a clear proposition – a cow bites a cookie – even though some of the grammatical elements are missing. In contrast, COW RED BITE COOKIE is not clear; a red cow could be biting a cookie, the cow may be biting a red cookie, or two propositions may be present – the cow is red, and biting a cookie is a separate proposition. Thus, COW BITE COOKIE has clear grammatical intent, but COW RED BITE COOKIE does not.

Further, each utterance that qualifies as a sentence will be coded to indicate the presence/absence of a unique subject-verb combination (i.e., USV; Hadley et al., 2018). To measure this, only the main subject and verb are listed. For example, for the sentences COW BITE COOKIE and COW BITE CAKE, only one USV is present: COW BITE. The number of USVs in a corpus reflects the child's ability to flexibly produce novel sentences based on grammatical rules instead of memorizing certain phrases and sentences (Hadley et al., 2018). In aided AAC as well, USV combinations are hypothesized to demonstrate the child's overall ability to use graphic symbols to produce flexible, rule-based sentences.

Finally, as children use graphic symbols to move from early symbol combinations (Phase 2) into early childlike sentences (Phase 3), they are expected to start using grammatical morphemes, just as children who are moving beyond early word combinations do in spoken language development. Simply tracking the presence of inflectional morphemes in each utterance allows for reporting the percentage of utterances containing grammatical morphemes for a corpus; this may prove to be a simple but useful way to indicate increased use of bound grammatical morphemes over time. In the current study, the presence/absence of grammatical morphemes will be indicated for each utterance.

Phase 4: Adultlike Sentences

During this phase, the child produces longer sentences with growing grammatical complexity and accuracy (Binger et al., 2020). All the measures introduced above can be used to measure the child's language at this phase, with the measures introduced in Phase 3 (i.e., presence of grammatical morphemes, grammatical intent, etc.), as well as the measures that are relevant to all phases, likely to be the most useful.

Reliability of the Measures

Above-mentioned measures were recently developed (Binger et al., 2020), but no formal reliability or validation procedures have been reported yet. Despite the logical nature of these measures, their reliability (as well as their validity and developmental sensitivity) must be established prior to widespread use in research and clinical practice, as no rational conclusions can be confidently drawn from unreliable data (Portney & Watkins, 2009). In terms of reliability, ensuring that scores can be objectively and reliably applied across observers is essential for any measure. Put another way, measures become much less useful when scores for the same child's behaviors vary across and/or within observers. The current investigation therefore focuses on two important components of measurement reliability: IOA and within-observer agreement.

Specific Aims

The first aim of the current investigation was to evaluate the IOA and within-observer agreement of data used to calculate a range of promising new measures to track early graphic symbol utterance development. All reliability data in the current study were coded at the utterance level and included the following: (a) measures applicable to all phases (i.e., independence, imitativeness, number of symbols, weighted utterance length score, and parts of speech [pronoun, noun, verb, etc.]); (b) Phase 1 measures (i.e., communicative intent clear/unclear, number of relevant symbols); (c) Phase 2 measure (i.e., word order scores); and (d) Phase 3 measures (i.e., SV presence/absence [present = SV present with correct word order; Hadley, 2014], grammatical intent clear/unclear, lexical verb present/absent, unique/repeated subject-verb combination, inflectional morphemes present/absent).

The second aim of this study was to explore the potential of the new measures to reflect the language development of children who use AAC during the early symbolic and early word combination phases. To accomplish this, the utterance level data collected during coding were further analyzed on a session level. Session level data were then compared with the characteristics of each phase of the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020) to determine whether the measures appeared to represent various aspects of aided language development of children with DS.

Method

Participants

Child Participants

Participants in the current study ranged in age from 3;4 to 5;9 (years; months). All participants were enrolled in a larger randomized controlled trial (RCT) designed to investigate the effects of an AAC intervention on the expressive language skills of preschoolers with DS. All data for the current study were collected prior to the start of this investigation. The main focus of the RCT was to increase aided utterance length and complexity. Participants in the larger study met the following inclusion criteria: (a) age 3;0 to 5;11 at the onset of the investigation; (b) English spoken as the primary language; (c) diagnosis of DS; (d) presence of a severe speech impairment, defined as less than 50% comprehensible language in the "with context" condition of the Index of Augmented Speech Comprehensibility in Children (I-ASCC; Dowden, 1997); (e) expressive vocabulary of at least 25 words/symbols on the MacArthur-Bates Communicative Development Inventories (CDI; Fenson, 2007) via any communication mode (speech, sign, aided AAC); and (f) parental report of functional vision and hearing for participation in study activities (See Table 4). Graduate and undergraduate students majoring in speech and hearing sciences who were unfamiliar with the participants judged comprehensibility for the I-ASCC; a different listener was used to score each sample to eliminate task familiarity influences.

Table 4
Participant Demographic Information

Child	CA (Sex)	I-ASCC		CDI	Prior aided AAC experience in months
		No context	With context		
1	5;4 (F)	9%	29%	124	0
2	5;3 (F)	0%	9%	214	0
3	4;6 (F)	29%	21%	106	1
4	4;4 (M)	0%	0%	245	0
5	5;8 (M)	13%	19%	260	0
6	4;2 (M)	3%	10%	78	6
7	3;4 (M)	0%	0%	74	0
8	5;9 (M)	45%	48%	347	0

Note. CA = chronological age in years;months; CDI = MacArthur-Bates Communicative Development Inventory (Fenson, 2007); I-ASCC = Index of Augmented Speech Comprehensibility in Children (Dowden, 1997)

Research Design

The current study is an IOA and within-observer agreement study. A total of 57 different intervention sessions from 8 child participants were rated by 4 observers. Informed consent was obtained from the parents of all participants prior to the larger study. Data were collected from all intervention phases (Month 1 – 4) in these analyses. Observers were masked to the intervention phase of each session.

Observers

Four graduate students majoring in speech and hearing sciences participated as observers. All students were paid as research lab employees. Observer 1 and Observer 2 had prior experience in coding using some of the measures with the Observer software for approximately 1 year. Observer 3 and Observer 4 were novice observers with no prior experience.

Materials

Intervention Sessions

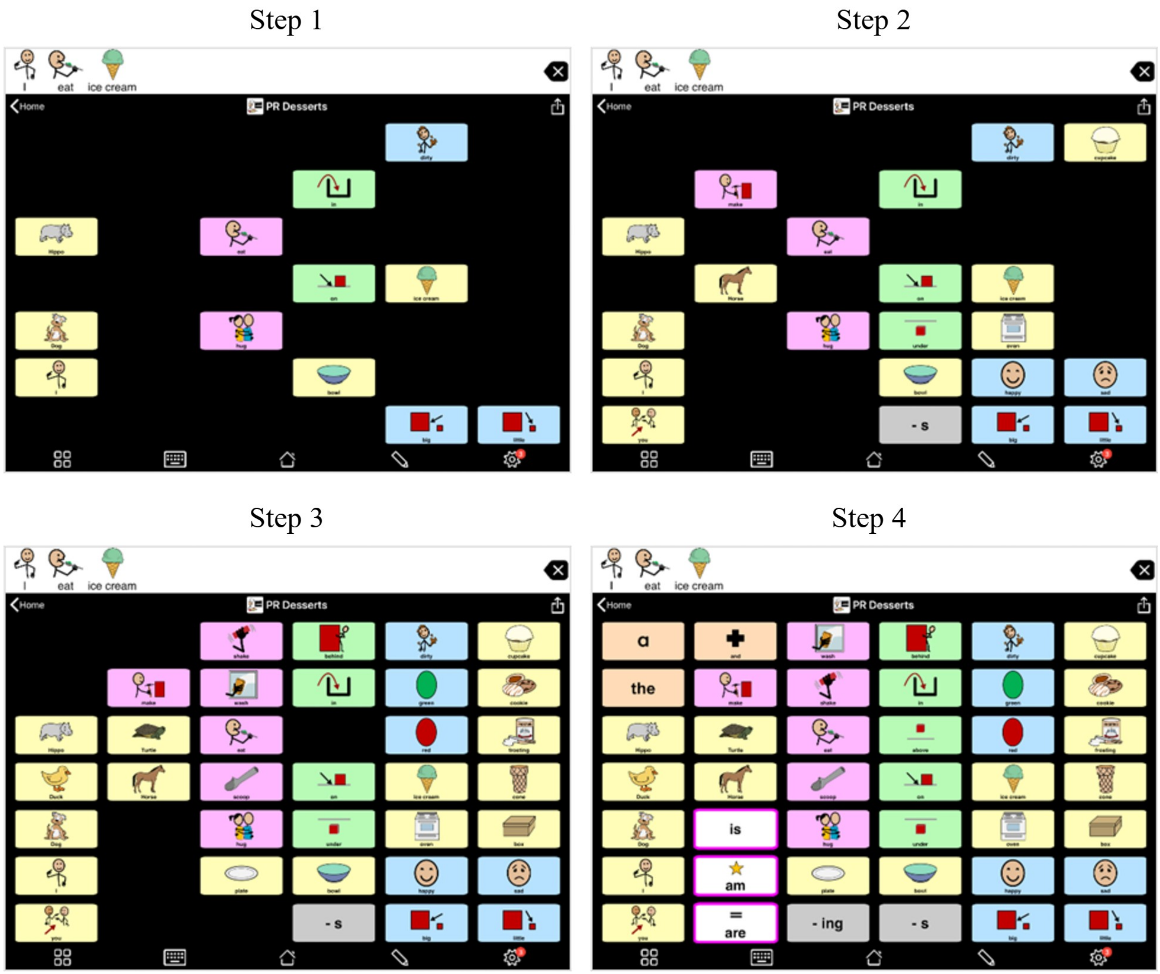
All participants used iPads equipped with an aided AAC app (ProLoQuo2Go) for the duration of the original RCT, including the sessions used in the current study. All participants used identical communication displays which were developed by the study team. Participants used one of nine different activity-based communication displays in each intervention session, with each display corresponding to various play routines (e.g., birthday party, farm, vehicles). The vocabulary on each display included various parts of speech (nouns, verbs, etc.), and Fitzgerald keys were used to construct all displays (McDonald & Schultz, 1973). All participants began the study using displays with a limited number of symbols, with symbols added systematically as the participants increased the number of different symbols they selected; that is, all participants began with Step 1 displays containing 12 symbols and progressed (as appropriate) up to Step 4 displays containing 42 symbols (see examples in Fig. 2). The steps increased according to pre-set criteria. For example, to progress from Step 1 to Step 2, participants had to use at least one symbol within four different word classes across two out of three consecutive sessions.

The participants selected for the current study originally were all randomly assigned to the intervention group within the larger RCT. Thus, these participants completed play-based intervention sessions twice weekly, in addition to initial assessment sessions and monthly measurement sessions. The play-based intervention sessions were used for the current study.

All sessions were video recorded using Noldus (Noldus Information Technology, Leesburg, VA) stationary lab equipment. Intervention was provided in the form of semi-structured play-based sessions using AAC-Generative Language Intervention (AAC-GLI). The intervention techniques used for AAC-GLI have been used in multiple prior AAC intervention studies (Binger, Kent-Walsh, King, & Mansfield, 2017; Binger, Kent-Walsh, King, Webb, & Buenviaje, 2017; Kent-Walsh et al., 2015). The intervention was delivered by speech-language pathologists (SLPs) and SLP graduate students. AAC-GLI includes three primary components: (a) intervention techniques including elicitation techniques (e.g., aided and spoken models, wait time, repetition with variety) and response techniques (e.g., expansions, extensions); (b) aided AAC technologies (described above); and (c) careful attention to contexts (which, in the current study, consisted of play-based contexts). Each intervention session included a review of the aided symbols that were to be used in the session and 25 minutes of play-based intervention.

Figure 2

Sample Displays for Dessert Play Routine



Each participant originally was scheduled to complete a total of 28 intervention sessions; that is, 7 play-based intervention sessions (plus a measurement session) each month over the course of 4 months. Due to the COVID-19 pandemic, however, only three participants completed all 4 months of the intervention (28 sessions; see Table 5).

Table 5
Number of Months of Intervention Completed by the Participants

Child	Month 1	Month 2	Month 3	Month 4
1	X	X	X	X
2	X	X		
3	X	X		
4	X			
5	X	X	X	X
6	X	X	X	X
7	X			
8	X			

Procedural fidelity (Schlosser, 2002) for the delivery of the intervention was monitored using fidelity checklists. Six coders completed procedural fidelity checks. SLP project managers at both study sites (University of New Mexico and University of Central Florida) taught the coders to complete the checklists using videos from a pilot investigation, with instruction continuing until the reviewers reached 95% compliance and agreement of coding on the fidelity checklists. A random sample of 20% of the sessions for each child was checked. Coders were masked to the phases of the investigation. Procedural fidelity was calculated by taking the number of steps followed correctly, divided by the total number of steps multiplied by 100. Mean fidelity adherence was 93.7% per child (range = 60% - 100%), indicating that the procedures were implemented consistently both across and within each participant.

Session Selection

Previously recorded videos of intervention sessions completed during the larger RCT were used for current study. Every session was recorded from two (synched) camera angles, with one camera focusing on the aided AAC display and the other focusing more widely on the overall interaction. Among all available intervention sessions, three out of seven sessions were selected for each participant within each month when possible. Sessions that contained fewer than 10 utterances were excluded; a minimum of 10 utterances were needed to calculate summary measures for Aim 2. Block randomization was used to select the eligible sessions, for a total of 57 sessions (see Table 6). This number of sessions was based on a power analysis completed before session selection indicating that at least 563 utterances were required to achieve the true kappa of .60 or higher with 50% agreement. Assuming each session contained at least 10 utterances, a total of 57 sessions were selected for coding.

Table 6
Selection of Intervention Sessions Using Block Randomization

Month	No. of participants	No. of sessions selected per participant	Total no. of sessions selected
1	8	6 participants: 3 1 participant: 2* 1 participant: 1	21
2	7	5 participants: 3 1 participant: 2* 1 participant: 1	18
3	3	3 participants: 3	9
4	3	3 participants: 3	9
Total			**57 sessions**

*For this participant, only two sessions were available during this month that contained at least 10 utterances

Development of the Coding Scheme

Noldus Observer XT-14 event coding software (Noldus Information Technology, Leesburg, VA) was used for all data analyses. The Observer software program synchronizes video recordings with behavioral coding. The coding scheme and operational definitions used in the current investigation were developed over the course of nine months across two separate research sites as part of a larger project. Data from a related but separate ongoing RCT involving different participants was used to develop the definitions and coding schemes, with modifications made for the current study as needed. The thesis student actively participated in this process over the course of all nine months.

Training

The thesis student served as one of the observers; she was heavily involved in developing the operational definitions and performed extensive coding prior to the start of the study. The three remaining observers completed training prior to coding, which included: (a) a two-hour training session with the thesis student and project manager, and (b) coding practice sessions. During the training session, observers reviewed the operational definitions of each measure (Appendix). The software program used to complete the coding (described below) was introduced, and at least three examples for each measure were provided. Observers completed trial scoring using related sessions that were not part of the current study. Training continued until observers achieved at least 80% accuracy for each measure across 2 consecutive sessions.

Procedures

Data Collection

For the current study, each of the four graduate students independently coded 28 to 29 intervention sessions for IOA and then recoded additional 7 to 8 sessions for within-observer agreement. Within the Observer software program, every symbol produced by the participants was transcribed. Unlimited repeated viewings were permitted. After transcribing each utterance, each utterance-level measure was then coded for all of the dependent measures (e.g., number of symbols, word order score, etc.; see Table 7). To ensure reliability of the raw data – that is, the aided utterance transcripts that served as the basis for all analyses – the thesis student compared the transcribed utterances across observers. If a discrepancy existed across the lists, the two observers collaborated to achieve consensus, and the behavioral data were recoded as needed.

Table 7
List of Measures

Measure	Description	Example	Measurement Type	Statistic
*All phases**				
Independence	Is the utterance independent or co-constructed?	Independent/ Co-constructed	Nominal	Gwet's AC1
Imitativeness	Is the utterance an imitation of the clinician's model or not?	Imitative/ Non-imitative	Nominal	Gwet's AC1
No. of symbols	How many symbols are used in each utterance?	1, 2, 3…	Ratio	ICC
Weighted utterance length score	What is the product of no. of symbols, PRSym, and word order score?	0, 0.6, 2.4	Ratio	ICC
Parts of speech	Which parts of speech are present in each utterance? (i.e., noun, pronoun, verb, adjective, adverb, preposition, conjunction, determiner)	Each one is listed	Nominal	Gwet's AC1
*Phase 1: Early symbol productions**				
Communicative intent	Is the communicative intent clear or unclear?	Clear/ Unclear	Nominal	Gwet's AC1
No. of relevant symbols	How many symbols are relevant to context?	0, 1, 2, 3…	Ratio	ICC
*Phase 2: Early symbol combinations**				
Word order score	How accurate is the word order is? (i.e., No errors = 1.0; At least one error = 0.5; No discernable word order = 0)	1.0, 0.5, 0	Ordinal	Weighted Kappa
*Phase 3: Childlike sentences**				
SV	Is subject + verb combination present or absent?	SV/ No SV	Nominal	Gwet's AC1
Grammatical intent	For SV sentences, is the intent of the message clear or unclear?	Clear/ Unclear	Nominal	Gwet's AC1
Lexical verb	Does the SV sentence include a lexical verb or a non-lexical verb (i.e., is, am, are)?	Lexical verb/ Non-lexical verb	Nominal	Gwet's AC1
USV	Is the SV combination unique or repeated?	USV/ Repeated SV	Nominal	Gwet's AC1
Inflectional morphemes	Is at least one inflectional morpheme present?	None/ At least one	Nominal	Gwet's AC1

Note. SV = subject + verb; USV = unique subject verb (Hadley et al., 2018); Gwet's AC1 = Gwet's Agreement Coefficient (Gwet, 2002; 2008); ICC = Interclass correlations coefficient (Koo & Li, 2016); Weighted kappa = Cohen's weighted kappa (Cohen, 1960)
*Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)

Data Analysis

Transcript Reliability. To ensure reliable data were in place, point-by-point (i.e., symbol by symbol) transcript reliability was calculated for 100% of the sessions (i.e., symbols that agreed/symbols that agreed + disagreed + missed). When each pair of observers met to compare transcripts, they counted the number of symbols agreed and disagreed and then mutually agreed on a master list of utterances for each session. Observers were permitted to watch the video again to resolve any discrepancies.

IOA and Within-Observer Agreement. Both IOA and within-observer agreement were calculated for each measure (See Table 7). Guidelines from Kottner and colleagues (Kottner et al., 2011) were followed for reporting IOA. IOA was first evaluated by session across observers. In this examination, each utterance was evaluated by two observers, and the IOA reflects the reliability of using the measures between the two observers. IOA also was evaluated across all sessions. In this examination, all utterances (N = 1507) were aggregated to estimate the IOA across all raters. Additionally, within-observer agreement was evaluated by each observer on two separate occasions.

For all nominal data (i.e., imitative/non-imitative, presence/absence of SV, clear/unclear grammatical intent, etc.), Gwet's Agreement Coefficient (Gwet's AC1) was used instead of Cohen's kappa, as Gwet's AC1 is less biased than Cohen's kappa in terms of the true agreement coefficient and overcomes criticisms of Cohen's kappa such as underestimation of true within-observer reliability (Gwet, 2002; 2008). Interpretations for Gwet's AC1 (which are based on kappa) are as follows: .01 or less indicate no agreement, .01-.20 as poor agreement, .21-.40 as fair agreement, .41-.60 as moderate agreement, .61-.80 as good agreement, and .81-1.00 as very good agreement (Landis &

Koch, 1977). For ordinal data (i.e., word order score), weighted kappa was used as it takes into account distances in ratings between observers, with squared weights of the differences computed to estimate IOA and within-observer agreement for ordinal measures (Cohen et al., 1960). Disagreements are weighted according to their squared distance from perfect agreement. Specifically, word order scores of 0 versus 0.5 and 0.5 versus 1.0 indicated better agreement than word order scores of 0 versus 1.0. For continuous data (i.e., number of symbols, number of relevant symbols, and weighted utterance length score), the intraclass correlation coefficient (ICC) was used to estimate the proportion of variation that is attributable to the participants relative to the total variation. Ratings in perfect agreement have no within-subject variation and thus no error variance, resulting in an ICC of 1.0. If there is little agreement between observers, the ICC will be closer to 0. Guidelines for reporting ICC for reliability estimates suggest that values below .50 are poor, between .50-.75 are moderate, between .75-.90 are good, and above .90 are excellent (Koo & Li, 2016).

Each observer transcribed and coded 28 or 29 different sessions (i.e., half of the 57 transcripts). Each session was coded a total of four times; that is, twice by each of the two observers assigned to a given session. Intervals between the two rounds coding was at least two weeks. Sessions were assigned using block randomization. Specifically, each pair of observers coded 9 or 10 of the same transcripts; for example, Observer 1 and Observer 2 coded 10 of the same transcripts, Observer 1 and Observer 3 coded nine of the same transcripts, etc. The assigned transcripts were balanced across participants and then randomized. In this manner, agreement for each dyad could be calculated to determine if there were coding differences across the dyads. *For the purposes of this* *study, ratings that had at least a "good" level of acceptability were considered to be adequately reliable (i.e., at least .61 for Gwet's AC1 and weighted kappa, and at least .75 for ICC).*

Secondary Analysis: Once utterance-level data were collected, further analyses were completed to assess aided language ability on corpus level (See Table 8). For the secondary analysis purpose, the data coded by the first observer of each dyad during the first round of coding were used. Utterance level data were exported to a Microsoft Excel spreadsheet for further analysis. Data were divided by sessions and then calculated for mean percentage and mean score for each measure on the spreadsheet.

Table 8

Utterance Level Measures and Corpus-Level Measures

Measures	Utterance Level Analysis	Corpus Level
*All phases**		
Independence	"Independent" vs. "Co-construction"	Percentage of independent utterances per session
Imitativeness	"Imitative" vs. "Non-imitative"	Percentage of non-imitative utterances per session
No. of symbols	Total no. of symbols	MLUSym
Weighted utterance length score	Weighted utterance length score	W-MLUSym
Parts of speech	Each part of speech listed	Mean no. of different parts of speech used per session
*Phase 1: Early symbol productions**		
Communicative intent	"Clear" vs. "Unclear"	Percentage of utterances with clear communicative intent
No. of relevant symbols	Total no. of relevant symbols	PRSym
*Phase 2: Early symbol combinations**		
Word order score	Word order score	Mean word order score
*Phase 3: Childlike sentences**		
SV	"SV" vs. "No SV"	Percentage of utterances with SV
Grammatical intent	"Clear" vs. "Unclear"	Percentage of SV utterances containing clear grammatical intent
Lexical verb	"Lexical verb" vs. "Non-lexical verb"	Percentage of SV utterances containing lexical verb
USV	"USV" vs. "Repeated SV"	No. of USV produced per session
Inflectional morpheme	"None" vs. "At least one"	Percentage of utterances containing at least one inflectional morpheme

Note. MLUSym = mean length of utterance in symbols; W-MLUSym = weighted mean length of utterance in symbols; PRSym = percentage of relevant symbols, SV = subject + verb; USV = unique subject verb (Hadley et al., 2018)
* Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)

Results

A total of 1,507 utterances in 57 sessions were coded for IOA and within-observer agreement. Transcript reliability was calculated before statistical analysis to assess how reliably observers transcribed aided utterances. Mean point-by-point (i.e., symbol by symbol) agreement for the transcripts was 92.7%, (*SD* = 6.55). All transcript discrepancies were resulted via consensus between the thesis student and observers.

Multi-Phase Measures

The overall IOA of the measures – that is, the IOA by session across observers – ranged from .74 to .99, with good agreement for imitativeness and weighted utterance length score, very good agreement for independence and parts of speech, and excellent agreement for number of symbols (See Table 9).

Table 9

Inter-Observer Agreement of Each Measure Across All Observers

Measures	Statistic	*M*	Range	*SD*	Interpretation
All phases*					
Independence[1]	.83	.83	.75-.90	.07	Very good
Imitativeness[1]	.74	.75	.62-.93	.12	Good
No. of symbols[2]	.93	.92	.81-.97	.07	Excellent
Weighted utterance length score[2]	.82	.81	.63-.92	.11	Good
Parts of speech[1]	.99	.99	.93-1.0	.03	Very good
*Phase 1: Early symbol productions**					
Communicative intent[1]	.98	.98	.97-.99	.01	Very good
No. of relevant symbols[2]	.92	.91	.80-.97	.07	Excellent
*Phase 2: Early symbol combinations**					
Word order score[3]	.66	.62	.39-.82	.16	Good
*Phase 3: Childlike sentences**					
SV[1]	.96	.96	.94-.99	.02	Very good
Grammatical intent[1]	.98	.98	.95-.99	.02	Very good
Lexical verb[1]	.98	.98	.96-.99	.01	Very good
USV[1]	.99	.99	.99-1.0	.01	Very good
Inflectional morpheme[1]	.98	.98	.96-.98	.01	Very good

Note. M = mean, SD = standard deviation, USV = unique subject verb (Hadley et al., 2018).
*Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)
[1]Gwet's Agreement Coefficient (Gwet's AC1; Gwet, 2002; 2008)
[2]Interclass correlations coefficient (ICC; Koo & Li, 2016)
[3] Cohen's weighted kappa (Cohen, 1960)

When more closely examining the reliability for each pair of observers, all dyads (e.g., Observers 1 vs. 2, 1 vs. 3, etc.) coded all measures reliably, with all IOA scores at acceptable levels (i.e., at least .61 for Gwet's AC1 and weighted kappa, and at least .75 for ICC) across measures and across dyads (See Table 10).

Table 10

Inter-Observer Agreement for Each Dyad

Measures	**1 vs. 2[a]**		**1 vs. 3[a]**		**1 vs. 4[a]**		**2 vs. 3[a]**		**2 vs. 4[a]**		**3 vs. 4[a]**	
	Stat *(SE)*	**95% CI**	**Stat *(SE)***	**95% CI**	**Stat *(SE)***	**95% CI**	**Stat *(SE)***	**95% CI**	**Stat *(SE)***	**95% CI**	**Stat *(SE)***	**95% CI**
All phases[b]												
Independence[1]	.89 (.03)	[.83,.94]	.75 (.04)	[.67,.84]	.90 (.03)	[.83,95]	.79 (.04)	[.72,.87]	.88 (.03)	[.81,.94]	.76 (.04)	[.68,.84]
Imitativeness[1]	.72 (.03)	[.65,.78]	.66 (.05)	[.56,.75]	.85 (.03)	[.80,.90]	.75 (.04)	[.67,.82]	.93 (.02)	[.89,.97]	.62 (.05)	[.52,.72]
Number of symbols[2]	.95	[.94,.96]	.97	[.97,.98]	.98	[.98,.99]	.81	[.77,.84]	.96	[.95,.97]	.86	[.83,.88]
Weighted utterance length score[2]	.92	[.90,.93]	.91	[.89,.93]	.84	[.81,.87]	.63	[.56,.69]	.77	[.72,.81]	.80	[.76,.83]
Parts of speech[1]	.99 (.01)	[.98,1.0]	1.0	-	1.0	-	.93 (.02)	[.89,.98]	.99 (.01)	[.99,1.0]	1.0	-
Phase 1: Early symbol productions[b]												
Communicative intent[1]	.97 (.01)	[.94,.99]	.99 (.01)	[.97,1.0]	.98 (.01)	[.97,.99]	.98 (.01)	[.96,.99]	.98 (.01)	[.96,1.0]	.97 (.01)	[.95,.99]
Number of relevant symbols[2]	.95	[.94,.96]	.97	[.96,.98]	.95	[.94,.96]	.80	[.76,.83]	.95	[.94,.96]	.84	[.81,.87]
Phase 2: Early symbol combinations[b]												
Word order score[3]	.82	-	.76	-	.60	-	.55	-	.60	-	.39	-
Phase 3: Childlike sentences[b]												
Sentence structure[1]	.96 (.01)	[.93,.99]	.96 (.01)	[.94,.99]	.97 (.01)	[.94,.99]	.96 (.01)	[.93,.99]	.99 (.01)	[.97,1.0]	.94 (.02)	[.91,.98]
Grammatical intent[1]	.99 (.01)	[.97,1.0]	.99 (.01)	[.97,1.0]	.96 (.01)	[.94,.99]	.99 (.01)	[.98,1.0]	.99 (.01)	[.98,1.0]	.95 (.02)	[.92,.98]
Lexical verb[1]	.99 (.01)	[.98,1.0]	.99 (.01)	[.98,1.0]	.98 (.01)	[.96,1.0]	.99 (.01)	[.97,1.0]	.99 (.01)	[.98,1.0]	.96 (.01)	[.93,.99]
USV[1]	.99 (.01)	[.98,1.0]	.99 (.01)	[.99,1.0]	.99 (.01)	[.99,1.0]	.99 (.01)	[.99,1.0]	.99 (.01)	[.98,1.0]	.99 (.01)	[.99,1.0]
Inflectional morpheme[1]	.96 (.01)	[.94,.99]	.98 (.01)	[.96,.99]	.98 (.01)	[.96,.99]	.98 (.01)	[.97,1.0]	.98 (.01)	[.95,1.0]	.98 (.01)	[.97,1.0]

Note. Stat = statistic; SE = standard error; CI = confidence interval; USV = unique subject verb (Hadley et al., 2018)

[a]Observers

[b]Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)

[1]Gwet's Agreement Coefficient (Gwet's AC1; Gwet, 2002; 2008)

[2]Interclass correlations coefficient (ICC; Koo & Li, 2016)

[3] Cohen's weighted kappa (Cohen, 1960)

The overall within-observer agreement for the measures considered pertinent across phases was above .80, with good agreement for weighted utterance length score and very good/excellent agreement for independence, imitativeness, number of symbols, and parts of speech (See Table 11).

Table 11
Within-Observer Agreement Across All Observers

Measures	*M*	*SD*	Interpretation
*All phases**			
Independence[1]	.87	.05	Very good
Imitativeness[1]	.85	.10	Very good
Number of symbols[2]	.95	.06	Excellent
Weighted utterance length score[2]	.88	.09	Good
Parts of speech[1]	.99	.01	Very good
*Phase 1: Early symbol productions**			
Communicative intent[1]	.98	.01	Very good
Number of relevant symbols[2]	.94	.06	Excellent
*Phase 2: Early symbol combinations**			
Word order score[3]	.72	.20	Good
*Phase 3: Childlike sentences**			
SV[1]	.97	.02	Very good
Grammatical intent[1]	.99	.01	Very good
Lexical verb[1]	.99	.01	Very good
USV[1]	.99	.01	Very good
Inflectional morpheme[1]	.98	.01	Very good

Note. M = mean, SD = standard deviation, USV = unique subject verb (Hadley et al., 2018).
*Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)
[1]Gwet's Agreement Coefficient (Gwet's AC1; Gwet, 2002; 2008)
[2]Interclass correlations coefficient (ICC; Koo & Li, 2016)
[3] Cohen's weighted kappa (Cohen, 1960)

The observers reliably recoded transcripts for within-observer agreement, with all measures above acceptable levels. Parts of speech was recoded most consistently by the observers, with the average coefficient close to 1.0. Independence and number of symbols ratings were at or above .80 for all observers. Imitativeness reliability scores ranged from .70 (Observer 3) to .92 (Observer 4; seeTable 12).

Table 12

Within-Observer Agreement for Each Observer

Measures	Observer 1		Observer 2		Observer 3		Observer 4	
	Stat (*SE*)	95% CI	Stat (*SE*)	95% CI	Stat (*SE*)	95% CI	Stat (*SE*)	95% CI
*All phases**								
Independence[1]	.90 (.02)	[.87,.93]	.89 (.02)	[.86,.92]	.80 (.02)	[.76,.84]	.91 (.01)	[.88,.94]
Imitativeness[1]	.88 (.02)	[.85,.92]	.90 (.01)	[.87,.93]	.70 (.03)	[.65,.75]	.92 (.01)	[.90,.95]
Number of symbols[2]	.99	-	.96	-	.87	-	.98	-
Weighted utterance length score[2]	.97	-	.90	-	.75	-	.91	-
Parts of speech[1]	.99 (.01)	[.98,.99]	.99 (.01)	[.99,1.0]	1.0 (-)	-	1.0 (-)	-
*Phase 1: Early symbol productions**								
Communicative intent[1]	.99 (.01)	[.98,.99]	.98 (.01)	[.97,.99]	.96 (.01)	[.95,.98]	.99 (.01)	[.98,.1.0]
Number of relevant symbols[2]	.98	-	.96	-	.86	-	.96	-
*Phase 2: Early symbol combinations**								
Word order score[3]	.88	-	.79	-	.44	-	.76	-
*Phase 3: Childlike sentences**								
SV[1]	.98 (.01)	[.97,.99]	.97 (.01)	[.95,.98]	.94 (.01)	[.92,.96]	.99 (.01)	[.98,1.0]
Grammatical intent[1]	.99 (.01)	[.99,1.0]	.98 (.01)	[.97,.99]	.98 (.01)	[.96,.99]	.99 (.01)	[.98,1.0]
Lexical verb[1]	1.0 (NA)	NA	.99 (.01)	[.98,1.0]	.97 (.01)	[.96,.98]	.99 (.01)	[.99,1.0]
USV[1]	.99 (.01)	[.99,1.0]	.99 (.01)	[.98,.99]	.99 (.01)	[.99,1.0]	.99 (.01)	[.99,1.0]
Inflectional morpheme[1]	.99 (.01)	[.98,.99]	.98 (.01)	[.97,.99]	.96 (.01)	[.94,.97]	.99 (.01)	[.98,.1.0]

Note. Stat = statistic; SE = standard error; CI = confidence interval; USV = unique subject verb (Hadley et al., 2018)

*Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)

[1]Gwet’s Agreement Coefficient (Gwet’s AC1; Gwet, 2002; 2008)

[2]Interclass correlations coefficient (ICC; Koo & Li, 2016)

[3] Cohen’s weighted kappa (Cohen, 1960)

Phase 1 Measures

For IOA, Phase 1 measures had very good agreement (communicative intent) to excellent agreement (no. of relevant symbols; Table 9). The IOA for all observers for all Phase 1 measures was at least .80, indicating an acceptable level of agreement (Table 10). It should be noted, however, that the 95% confidence intervals fell below .80 for Observers 2 versus 3 for the number of relevant symbols.

Within-observer agreement for these measures were very good (communicative intent) to excellent (no. of relevant symbols; Table 11). Observers recoded their transcripts consistently for all Phase 1 measures; all agreement levels were above .80, with most being above .90 (Table 12).

Phase 2 Measures

The only Phase 2 measure was word order scores, with a weighted kappa of 0.66, which is considered "good" agreement. Notably, the word order score IOA ranked the lowest across all measures (Table 9). Variability was noted in the IOA coefficients across dyads, with a high of 0.82 for Observers 1 and 2 to a low of 0.39 for Observers 3 versus 4 (Table 10).

Within-observer agreement of word order score was also considered "good" agreement (0.72; Table 11). Relatively high variability was noted, ranging from 0.88 (Observer 1) to 0.44 (Observer 3; Table 12).

Phase 3 Measures

Phase 3 measures included presence or absence of SV, grammatical intent, lexical verbs, USV, and inflectional morphemes and were analyzed using Gwet's AC1. IOA of all Phase 3 measures ranged from .96 to .99 indicating very good agreement (Table 9). All dyads coded the utterances consistently, and all 95% confidence interval agreement levels exceeded .90 (Table 10).

Within-observer agreement of these measures was also above .90 corresponding to very good agreement (Table 11). All observers recoded these measures reliably and demonstrated agreement level above .90 for all measures (Table 12).

Secondary Analysis

Based on the utterance-level data collected during coding, session-level measures were calculated (Table 13).

Table 13

Secondary Analysis Results for Session-Level Measures

Measures	M	SD
*All phases**		
Percentage of independent utterances per session	64.1	15.2
Percentage of non-imitative utterances per session	74.0	14.1
MLUSym	2.0	1.1
W-MLUSym	1.8	1.0
Mean no. of different parts of speech used per session	4.5	0.6
*Phase 1: Early symbol productions**		
Percentage of utterances with clear communicative intent	98.5	3.3
PRSym	99.1	2.3
*Phase 2: Early symbol combinations**		
Mean word order score	0.9	0.06
*Phase 3: Childlike sentences**		
Percentage of utterances containing SV	9.9	9.1
Percentage of SV utterances containing clear grammatical intent	66.7	43.1
Percentage of SV utterances containing lexical verb	71.2	42.0
No. of USV produced per session	1.6	1.5
Percentage of utterances containing $\geq$ one inflectional morpheme	2.6	3.6

Note. M = mean; SD = standard deviation; MLUSym = mean length of utterance in symbols; W-MLUSym = weighted mean length of utterance in symbols; PRSym = percentage of relevant symbols, SV = subject + verb; USV = unique subject verb (Hadley et al., 2018)

* Phase numbers refer to phases on the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020)

Discussion

This study revealed encouraging findings about the newly proposed measures. The results provide initial evidence that a range of measures designed to monitor the progress of graphic symbol utterances can be reliably coded. Statistical analyses indicate that most of the measures that were studied met at least minimal standards (i.e., “good agreement”), and nearly all measures were coded with high levels of agreement. This was true when the observers were compared to each other (IOA) and with themselves (within-observer agreement). Secondary analysis results also provide promising corpus-level data that are consistent with the perspective of the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020).

Multi-Phase Measures

Results from the first two multi-phase measures, independence/co-construction and imitative/non-imitative, had acceptable levels of agreement (good agreement or higher) both within and across observers. These measures are important to characterize how dependent communicators are on their communication partners when they construct the utterances within naturalistic situations. The secondary analyses indicated that participants produced more than half of their utterances independently (63%) and non-imitatively (74%). This is particularly encouraging given that the participants were preschoolers with DS – all with receptive language delays – who had little to no past aided AAC experience.

The number of symbols and weighted utterance length score are hypothesized to be useful for describing aided language growth over time. Specifically, these data can be used to calculate MLUSym and W-MLUSym, which are derived from the commonly used spoken language equivalent, MLU (Rice et al., 2006). Given the known issues with MLUSym explored in the introduction, the fact that the weighted utterance scores (i.e., the basis for W-MLUSym) reached acceptable levels of IOA and within-observer agreement is encouraging; most dyads of observers and all individual observers scored in "excellent" to "good agreement" ranges for weighted utterance scores. However, weighted utterance length scores and W-MLUSym are based in part on word order scores, which were the least reliably coded scores included in this investigation. This is potentially problematic for the new W-MLUSym; if word order is to be included as part of W-MLUSym, improving the reliability for this score is needed (discussed further below).

Due to the language development characteristics and minimal prior experience of AAC in the child participants of the current study, their aided language productions were expected to be characteristic of the beginning phases of early expressive language development. The results were consistent with this expectation, with an overall MLUSym (derived from the "Number of symbols" measure) of 2.0 and W-MLUSym of 1.8. This means that the participants were regularly combining symbols, and theoretically functioning within Phases 1to 2 and emerging into Phase 3 of the Graphic Symbol Utterance and Sentence Development Framework (Binger et al., 2020). Additional work is needed to establish coding reliability of these measures with children who are functioning at higher levels.

One of the measures with the highest agreement rate was parts of speech, with a coefficient close to 1.0 across all dyads and all individual observers. Several factors may have enhanced the ease of coding for this measure. Displays for each activity scene were carefully organized to maintain balance across parts of speech. Further, symbols were grouped together and had color-coded backgrounds according to the parts of speech (see Fig. 2). In addition, observers were provided with the list of vocabulary for each display. Secondary analysis of the parts of speech revealed that the child participants used approximately four to five different parts of speech at least once during the session on average (Table 12). Notably, only four parts of speech (noun, verb, adjective, preposition) were included in the Step 1 to Step 3 displays. This symbol diversity may be reflective of the nature of these play intervention sessions, with clinicians providing many models and other supports to encourage the participants to use all available vocabulary. Comparisons with controls are needed to verify this.

Measures Focused on Phases 1 and 2

As discussed above, the participants demonstrated language skills largely commensurate with Phases 1 and 2, and the findings provide strong support for the reliability of the measures that were explored for children functioning within these earlier phases. Within and across all observers, near perfect agreement was achieved for determining both communicative intent and symbol relevance. From a developmental perspective, demonstrating clear communicative intent is a crucial component of early communication development (Wetherby & Rodriguez, 1992). The relevance of symbol selection – an issue unique to aided AAC – is also fundamental to successful aided communication (Binger et al., 2020). The findings for these measures, however, should be viewed as preliminary. Little variability was apparent, with the secondary data analyses indicating that virtually all utterances were coded as having clear communicative intent (98.5%) and were relevant to the context (mean PRSym 99%). More variability is needed to demonstrate that these measures are differentiating between distinct behaviors; that is, a more substantial number need to lack clear communicative intent. Notably, the intervention itself likely led to this lack of differentiation in these two measures. The current sessions consisted of carefully crafted play routines accompanied by activity scene displays, with all vocabulary directly relevant to the play. Thus, clinicians could infer communicative intent and assume relevance for most participant turns. Additional research that includes children functioning within other phases, in other less structured contexts, and with access to a wider array of vocabulary is required to further establish the reliability of these measures.

Although the overall agreement of all Phases 1-2 measures was within acceptable levels (i.e., >.60 for Gwet's AC1 and weighted kappa; >.75 for ICC), word order scores proved to have the lowest level of agreement across all measures, with agreement falling below acceptable levels for individual observers. The IOA for the two least experienced observers was rated as fair (.39), and the within-observer reliability for one of these same observers was rated as moderate (.44). Notably, the participants were functioning at a level when struggles with word order may be most likely to appear. That is, the data indicate that participants on average were in Phase 2 (early symbol combinations), when children are in the early stages of expressive syntactic development. In spoken language development, this corresponds to the period when children are sorting out how to combine words, usually at 24-32 months in typically developing children (Hadley, 2014). Many children with DS are still learning to combine words in the preschool years due to the delayed development of expressive language (Chapman & Bird, 2011). Unlike communicative intent and relevance of vocabulary, the current data set appears to have focused on the period of development when word order issues are most likely to occur, thus challenging the observers and presenting a strong test of this measure. The data provide preliminary evidence that this measure presents coding challenges, particularly for inexperienced observers. Improved training methods may assist observers in achieving higher levels of reliability. Observers of the current study reported that various examples presented during the training and the practice coding were highly useful to learn to use the measures. Increasing hands-on activities during the training may improve the training outcome. At the same time, changes in training efforts must be balanced with the ultimate clinical feasibility of this measure; that is, measures that require a high level of training to achieve reliability are less clinically feasible. Additional research is needed to determine how reliably this measure can be coded for children functioning across all language phases and how clinically feasible this measure ultimately proves to be.

Measures Focused on Phase 3

All of the measures hypothesized to pertain to Phase 3 – that is, utterances that qualify as childlike sentences, because they contain both a subject and a verb – demonstrated excellent reliability with all measures within the highest agreement category both across and within observers. However, due to the limited nature of the data set, the results should be interpreted with a caution. The first measure – the presence/absence of SV – included all of the same data as the prior measures (i.e., the 1,486 utterances that had clear communicative intent). From there, however, only utterances that contained SVs were included in subsequent analyses, and far less data were available for these measures. Only 9% (136 utterances) of the total utterances were sentences and coded for these measures, which included clarity of grammatical intent, presence or absence of a lexical verb, and USV versus repeated SV. Notably, however, all participants produced at least one SV utterance, so this measure did apply to all participants.

Interestingly, the vast majority of these 136 utterances (i.e., 93%) were deemed to have clear grammatical intent. This was defined as an estimation of whether or not the intended sentence was evident, regardless of the completeness of grammatical elements. For example, DOG EAT CAKE is an utterance with a clear grammatical intent. In contrast, DOG EAT PUSH CAKE does not have a clear grammatical intent, despite the presence of SV. The utterance may be interpreted as “Dog eats cake,” “Dog pushes cake,” or perhaps a combination; therefore, the listener cannot clearly determine the message the child is trying to convey. Further, all but two of these same 136 utterances contained lexical verbs. As with communicative intent and relevance of symbols, more variability in the data is needed to determine if observers can truly reliably discern the difference between clear versus unclear grammatical intent, presence versus absence of a lexical verb, and USV versus repeated SV. Two features of the data are likely affecting the results. First, if all 1,486 utterances had been coded for grammatical intent (rather than just the 136 SV utterances), more variability likely would have occurred. Estimating the underlying grammatical intention of utterances – whether they are spoken or aided – that lack a subject or verb is a known challenge. This can be seen, for example, in studies that focus on building early semantic relations (Binger et al., 2019). The high rate of lexical verb productions was likely affected by the nature of the intervention. Participants did not have access to non-lexical verbs (which were *IS*, *AM*, and *ARE* in the current study) until they reached the Step 4 displays (Fig. 2). Out of the eight participants, only two progressed to the Step 4 displays. For USV versus repeated SV, more variability was noted (i.e., 91 USV productions vs. 43 repeated SV productions), which provides stronger evidence for the reliability of this measure.

Agreement Across Observers

Patterns of differences across observers were noticed both in IOA and within-observer agreement. The observers of the current study included two experienced observers (Observers 1 and 2) and two novice observers (Observers 3 and 4). Observer 1 was the thesis student, who had been heavily involved in developing operational definitions for the measures. Both Observers 1 and 2 participated in many pilot sessions prior to the start of the study. Novice observers were truly novice: they only received the two-hour training and two practice coding sessions. The differences in coding were most obvious in the word order score. IOA of Observers 1 versus 2 was in the "Very good agreement" category while IOA of Observers 3 versus 4 was categorized as "Fair agreement." Within-observer agreement of word order score also reflected this pattern, with Observers 1 and 2 scoring higher agreement than the Observers 3 and 4. These findings are consistent with previous research indicating that the amount of experience, participation in training sessions, and opportunities to practice improve reliability when SLPs use clinical measures (John & Enderby, 2000). Future research needs to determine clinical feasibility of the measures and investigate ways to increase the reliability for novice observers such as additional training sessions, modified training materials (e.g., operational definition, video presentation, etc.), and perhaps simplifying the measures themselves. Measures can only be applied to clinical settings when clinicians can readily learn to use them reliably.

Clinical Implications

Results of the current study suggest that many of the newly proposed measures are relatively easy to learn to code reliably. This is promising for clinical applications, although future work, as discussed above, is warranted. Secondary analysis of the data focusing on the session level measures (e.g., MLUSym, W-MLUSym, etc.) also provided initial evidence that these measures appear to be meaningful to the language development of the children with DS (Table 12). As previously mentioned, most of the child participants were largely new to aided AAC. However, the majority (63%) of their utterances were independently constructed without the communication partner's prompt, and nearly three-quarters of the utterances were non-imitative; that is, most utterances contained at least one novel symbol that was not included in the clinician's previous utterance. Further, virtually all utterances were both intentional (99%) and relevant to the context (≥99%). Taken together, these data indicate that at least in supported play-based sessions, the eight preschoolers with DS in the current study were able to independently produce intentional, relevant, aided utterances that were not mere imitations of the clinician. Of course, additional work is required to determine which measures truly monitor progress over time; that is, to ensure that measures such as W-MLUSym increase in a similar developmental trajectory as MLU does in spoken language development.

It must be noted that the high levels of communicative intent and relevance likely were affected by the technical choices and aided AAC setup choice made as part of the intervention. As previously mentioned, the available vocabulary was carefully selected and placed within event schemas for each play routine, so most symbols selections necessarily were relevant to the context. Further, a limited number of symbols were available to the participants, particularly during the earliest sessions (see Fig. 2), which further ensured the chances that a selection would be relevant. These approaches were designed to help participants build early success and confidence, and to allow them to focus on building their syntactic and grammatical skills as quickly as possible. Such learning patterns have been identified in previous, related intervention studies (e.g., Tönsing et al., 2014; Kent-Walsh et al., 2015). Using these approaches have resulted in rapid gains in the production of rule-based, multi-symbol aided utterances.

Relatedly, most utterances also were coded as having correct word order, with a mean word order score of .95 with relatively little variation (i.e., *SD* = .18). This indicates that most utterances had no discernable word order issues (which is scored as 1.0), compared with utterances that have at least one error (.5) or no discernable word order in the utterance (0). However, for several reasons, caution is warranted in interpreting the word order findings. First, all single symbol utterances were awarded a word order score of 1.0, which accounted for a substantial percentage (40%) of the utterances and therefore likely inflated this score. One solution would be to eliminate single symbol utterances from word order analyses in future projects. Also, past findings indicate that when young children begin to use graphic symbols to communicate, they *do* exhibit word order issues. For example, Binger and colleagues (2019) found that word order issues were particularly prevalent for utterances containing reversible agents and objects in SVO sentences such as DOG DROP COW and COW DROP DOG (Binger et al., 2019). Two key differences between the study by Binger and colleagues (2019) and the current investigation include: (a) the levels of support offered by the clinicians, with no supports offered in probes in the previous study, compared with high levels of clinician support in the current study, and (b) little to no expectation of children producing utterances known to be prone to word order issues (i.e., reversible SVO utterances) in the current study. Further, including additional data from children who are functioning at higher language levels will assist with determining the developmental trajectory and sensitivity of this (and the other) scores.

Limitations and Future Directions

One of the limitations of the current study is the focus on children who are in the early stages of learning to use aided AAC. This was further compounded by the greater number of available sessions for participants who, because of the COVID-19 pandemic, had only completed the first month or two of a four-month intervention. As a result, approximately one-third of the sessions were selected from Months 3 and 4. This imbalance should be taken into account when interpreting the data, and future research needs to include children with more advanced aided language skills who produce longer, more sophisticated utterances as well as more complex AAC displays with a greater variety of vocabulary. This will allow for a fuller exploration of the reliability for the measures designed to reflect the growing grammatical skills of children who use aided language to communicate. One notable measure of interest is word order scoring, given the reliability issues noted with the population in the current study. This measure has particular importance, given known word order issues in graphic symbol communication (Binger et al., 2019; Binger & Light, 2008) and its inclusion in the proposed formula for computing W-MLUSym. Future reliability research also would benefit from explorations of additional contexts. The current study focused on play sessions that were supported by clinicians, with aided AAC displays that included only relevant vocabulary and required no navigation to locate vocabulary. Additional work is needed to see how reliably the studied measures can be coded in different conditions.

Replicating this study with populations other than children with DS is recommended as well. For example, measuring the number of relevant symbols may be useful for children with autism spectrum disorder who have discrimination issues and tend to repetitively press irrelevant symbols while using graphic symbols. Investigating how measures differ across participants will further specify the usefulness of these measures.

Another limitation is the fact that the secondary analysis data do not measure growth or learning over time. At the time of this writing, data from the larger RCT have not yet been completed, so the full intervention effects are unknown. Future research is required to examine growth patterns once the intervention effects are determined.

Additionally, establishing the IOA and within-observer reliability of the measures is only a first step to establishing strong psychometrics. Future studies need to evaluate additional aspects of reliability as well as validity, developmental sensitivity, and clinical feasibility. In terms of reliability, another possibility is split-half reliability, which is known to be calculable from language samples (Heilmann et al., 2010). For validity, a range of measures are of interest, including construct validity, convergent and divergent criterion validity (e.g., how closely the measures do and do not correlate with related constructs such as receptive language levels and not with unrelated constructs such as nonverbal IQ), and social validity (Castilla-Earls & Fulcher-Rood, 2018; Messick, 1996; Schlosser, 1999). Various aspects of clinical feasibility also need to be explored, including the time and effort it takes to collect and analyze the data, the length and format of training, possible modifications to the procedures that increase efficiency (such as knowing the minimal number of utterances that need to be collected and still maintain validity), and clinical acceptability (which can be determined using surveys and qualitative methods such as focus groups). Clinical feasibility is a critical factor for implementation of a new method, therefore, various perspectives such as facilitators and barriers need to be examined closely (Hickey et al., 2019)

Conclusion

This study provides preliminary evidence that aided utterances produced by preschoolers with Down syndrome who use AAC can be reliably measured and quantified. Although not all the measures were equally easy to rate, an acceptable level of agreement was established for most measures, both across and within observers. Additionally, the findings of this study demonstrated that the newly proposed measures show promise for reflecting the aided language functioning of children who use AAC – at least for children who are functioning in the early symbolic and early word combination phases of development. This foundation is an important accomplishment for developing psychometrically sound ways to facilitate systematic assessment and treatment, effectively monitor progress, and expand the application of AAC for those who need AAC to communicate.

Appendix

Measurement Reliability
Word by Word Project
Operational Definitions for All Measures

Table of Contents

Goals

- Operationally define range of measures used to assess progress in "Word by Word" studies
- Establish reliability for newly created measures used to track graphic symbol utterance progress

Utterance Boundaries & Transcription Conventions

All data

- Transcribe aided utterances using CAPITAL LETTERS.
  - Insert a space between symbols. *Symbols that contain more than one word should be transcribed as a single word*.
    - Examples
      - RED DOG
      - I EAT ICECREAM
  - Use square brackets to transcribe symbols selected by the examiner.
    - Example
      - Child: DOG
      - Examiner: EAT
      - ➔ Transcribe as DOG [EAT]
  - Use curly brackets to transcribe unconnected symbols selected by the child during a co-constructed utterance.
    - Example
      - Child: DOG
      - Examiner: EAT
      - Child: BIG
      - Examiner: CAKE
      - ➔ Transcribe as DOG [EAT] {BIG} [CAKE]
- Inflectional morphemes: Put a space then a hyphen before each inflectional morpheme. Regardless of the context for –S (plural, possessive, 3rd person singular, contracted copula/auxiliary), follow this same convention.
  - Examples
    - I AM EAT -ING
    - DOG -S PLATE
    - DOG –S EAT -ING
- Transcribe the child's final production on Observer
  - ***Video Probe Data***
    - Code the child's final production for each target*. The final production may or may not be played back (by the child or researcher selecting the message bar).
      *Exception: If child produces more than one utterance for a target, code the utterance that is most reflective of the child's linguistic ability, rather than behavior or attending skills.
  - ***Play Intervention/Play Measurement/Storybook/Activity Scene Data***

- *Measures are designed to capture linguistic abilities of the children; not operational skills such as selecting or clearing the message bar.*
- Utterance boundaries are determined by several contextual factors
  - Child selects symbol(s) in either an independent or co-constructed utterance and it is clear that the utterance is completed.
    - *Note*: For co-constructed utterances, the utterance may be partly completed by the examiner.
  - Child selects symbol(s) & message bar → No one erases the message → Child *adds to* message with no turns in between
    - This all counts as one utterance = count entire message once
    - Exception: It's obvious that the child intended to produce two separate messages and simply forgot to erase prior production
  - If the previous message was not erased before the new utterance began:
    - Code as a *separate utterance* if it is clearly a mistake (e.g., examiner indicates that they forgot to erase the previous message, there is a clear change in context or clearly a new turn, with new utterance being produced or elicited).
      - Example
        - Child: RED DOG
        - [Message bar]
        - Child: adds BLUE DOG
        - Message that is played is RED DOG BLUE DOG
        - Examiner indicates that they should've erased the previous message.

          → Code RED DOG and BLUE DOG as two separate utterances.
    - Code as one utterance if the child indicates they need to add more symbols (either verbally or by gestures) or adds symbols in response to the examiner's prompt.
      - Example
        - Child: RED DOG
        - [Message bar]
        - Examiner: What is the red dog doing?
        - Child: adds EAT
        - Message that is played is RED DOG EAT

          → Code as one utterance of three symbols

Utterance boundaries of co-constructed utterances

- Sometimes the examiner deletes the symbol immediately after the child selects it to elicit a correct utterance. Transcribe the final utterance excluding the symbols selected by the child and deleted by the examiner.
- Example:
  - Examiner is trying to elicit RED DOG. She provides spoken directive and pointing.
  - Examiner: Says "red" and points to the RED symbol.

- Child: RED
    - Examiner: Says "dog" and points to the DOG symbol.
    - Child: HORSE
    - Examiner: Deletes HORSE, says "dog," and points to the DOG symbol.
    - Child: HIPPO
    - Examiner: Deletes HIPPO, says "dog," and points to the DOG symbol.
    - Child: DOG
    - Utterance played back: RED DOG
    - Transcript: RED DOG
- If the clinician verbally responds to the child's utterance and deletes it, we transcribe both the first utterance and the second utterance.
- Example:
    - Clinician is trying to elicit RED DOG.
    - Child: RED HIPPO
    - Clinician: "Is this a red hippo?" (Spoken open-ended prompt)
    - Child: No.
    - Clinician: Deletes the utterance.
    - Child: RED DOG
    - Transcript: RED HIPPO is one utterance & RED DOG is a separate utterance
    - If the clinician deletes only HIPPO and child selects DOG to complete RED DOG, Transcribe the final RED DOG

Utterance boundary of self-corrected utterances

- If a child deletes a symbol before the clinician says anything, do not transcribe it as an utterance.
    - Example:
    - Child: RED DOG, deletes dog, HIPPO
    - Transcript: RED HIPPO
- If a child deletes a symbol following the clinician's prompt, transcribe it as two utterances.
    - Example:
    - Child: RED DOG
    - Clinician: "Do you want a red dog?"
    - Child: Deletes RED DOG and selects RED HIPPO
    - Transcript: RED DOG & RED HIPPO
- If the child deletes only DOG and then selects HIPPO to complete RED HIPPO, give the child credit for both RED DOG and RED HIPPO here, as this was all child selections.

Measures of Interest

The following measures are behaviors and modifiers to be coded in Observer (See Observer SOP & Video Walkthrough)

- Imitative/ Non-imitative
- Independent vs. co-constructed aided utterances
- Communicative intent present/ absent
- No. of symbols
- No. of relevant symbols
- Word order score (0, .5, 1)
- SV structure (i.e., is there a subject-verb present?)
  - Grammatical intent
  - Lexical verb
  - Add subject-verb combination to comments
  - USV (Unique Subject-Verb)
  - Lexical/non-lexical verb
- Grammatical Intent
- Inflectional morphemes present/not present
- Part of speech
  - Noun
  - Pronoun
  - Verb
  - Preposition
  - Adjective
  - Determiner
  - Conjunction

The following measures will be calculated automatically based on the measures above

- Percentage of relevant symbols (PRSym)
  - Number of Symbols and Number of Relevant Symbols are used
- MLU in symbols (MLUSym)
  - Number of Symbols per utterance are used
- Weighted MLU in symbols (W-MLUSym)
  - No. relevant symbols * Word order score = W-MLUSym
  - *NOTE: This is how we are initially calculating W-MLUSym. We will explore various methods to see what best captures positive shifts in grammar. For example, we will try adding in a measure that captures the number of different parts of speech.*

Operational Definitions

Imitativeness

Definitions	**Purpose**	**How to code**
Non-imitative utterance: Child's aided production contains at least one novel concept compared with the examiner's prior aided/spoken utterance. *Imitative utterance*: Child's aided production contains no novel symbols compared with examiner's immediately prior aided/spoken utterance. The intention is to capture immediate, not delayed imitations.	To differentiate imitative utterances from non-imitative utterances. Utterances that are imitative demonstrate a lower level of linguistic sophistication and internalization.	Code each utterance to indicate if it is *non-imitative* or *imitative*

Details

Video Probe Data

This code is not relevant for Video Probes

Play Intervention/Play Measurement/Storybook/Activity Scene Data

Code each utterance as either *independent* or *co-constructed*.

- Possible codes include (a) *non-imitative,* (b) *imitative – spoken + aided model,* (c) *imitative – spoken model,* or (d) *imitative-aided model.*
  - Non-imitative
    - The aided utterance contains **at least one novel symbol** compared with examiner's immediately prior spoken or aided utterance.
    - If examiner provides a brief spoken prompt telling the child to produce an utterance (e.g., "Now you tell me)", compare the child's utterance with the examiner's model.
    - Example
      - Examiner: Oh, I'm eating cake. What are you eating?
      - Child: EAT CAKE [Non-imitative]

      Rationale: To respond to the question "What are you eating?," the child's linguistic task is to respond appropriately to the question, which requires EAT CAKE. Thus, this is not considered an imitation.
    - Example
      - Examiner: Oh, I'm eating cake. I EAT CAKE. Now you tell me.
      - Child: EAT COOKIE [Non-imitative]

        Rationale: The child added the novel symbol COOKIE.
  - Imitative – Spoken + aided model
    - All symbols are in the examiner's immediately prior *spoken and aided* utterance.

        - If examiner then provides a brief spoken prompt telling the child to produce the utterance, use this same code.
        - Example:
            - Examiner: I'm eating cake. I EAT CAKE. Now you tell me.
            - Child: EAT CAKE [Imitative – Spoken + aided model]
              Rationale: The child's utterance contains ALL symbols from the examiner's immediately prior spoken + aided utterance. The spoken prompt "You tell me" does not count as the Examiner's immediately prior utterance.

    - Imitative – Spoken model
        - All symbols are in the examiner's immediately prior *spoken* utterance.
        - If examiner then provides a brief spoken prompt telling the child to produce the utterance, use this same code.
        - Example:
            - Examiner: Oh, I'm eating cake. Now you tell me.
            - Child: EAT CAKE [Imitative – spoken model]
              Rationale: Same as above
    - Imitative – Aided model
        - All symbols are in the examiner's immediately prior AIDED utterance.
        - If examiner then provides a brief spoken prompt telling the child to produce the utterance, use this same code.
        - Example:
            - Examiner: I EAT CAKE. Now you tell me.
            - Child: EAT CAKE [Imitative – Aided model]
              Rationale: Same as above

    - If the examiner offers *binary choices* or *multiple options* via spoken or aided modalities, consider the child's production non-imitative.
        - Example:
            - Examiner: Do you want red horse or blue horse? RED HORSE BLUE HORSE
            - Child: BLUE HORSE [non-imitative]
            - Child: RED HORSE [non-imitative]
            - Examiner: We have a yellow, train, blue train, green train, and dirty train.
            - Child: DIRTY TRAIN [non-imitative]

Independent vs. Co-Constructed Aided Utterances

Definition	Purpose	How to code
Independent utterance: Child's aided production is produced without substantial prompting from the Examiner. *Co-constructed utterance*: Child received prompting assistance while in the process of constructing an utterance.	To differentiate utterances the child produces without any assistance from utterances that are supported by the examiner while the child is in the act of producing the utterance.	Code each utterance to indicate if it is *independent* or *co-constructed*

Details

Video Probe Data

This code is not relevant for Video Probes

Play Intervention/Play Measurement/Storybook/Activity Scene Data

Code each utterance as either *independent* or *co-constructed*.

- Independent
  - While constructing the utterance, the child receives none of the supports listed below from the examiner.
  - It's ok for the examiner to provide purely *operational* supports (i.e., App/iPad functioning; e.g., to select the message bar, erase a selection, etc.). If the examiner provides any *linguistic* suggestions (i.e., selecting symbols to construct the message; e.g., "That's not quite right, etc.") then it's co-constructed.
- Co-constructed
  - Child receives at least one of the following types of assistance from the examiner *while in the act of constructing an utterance*:
    - Spoken open-ended prompt
      - Examiner's prompt is general; does not tell the child which specific symbol to select
      - Example:
        - Child: I EAT
        - Examiner: What are you eating?
        - Child: adds CAKE. Plays back I EAT CAKE.
    - Spoken directive prompt
      - Examiner tells the participant to produce at least one particular symbol
      - Example:
        - Child: I EAT
        - Examiner: You're eating **cake**.
        - Child: adds CAKE. Plays back I EAT CAKE.
    - Pointing directly to the symbol(s)

        - Examiner points *directly* to at least one symbol as the participant constructs an aided utterance
        - Pointing generally to the device or to an area of the display does not count; utterances receiving this kind of assistance are considered independent

    - Selects the symbol(s)
        - Examiner selects at least one symbol during the child's aided utterance.
            - Code the child's initial selection only. Examiners selections are in square brackets [ ]. Child's second selection is in { } brackets.
            - Example:
                - Utterance: [DOG] EAT [CAKE]
                    - ➔ Only EAT will receive codes.
                - Utterance: [DOG] EAT [BIG] {CAKE}
                    - ➔ Only EAT will receive codes.

Please refer to utterance boundaries and transcriptions section on page 3.

Communicative Intent: Present/Not Present

Definition	Purpose	How to code
When communicative intent is present, the child is selecting graphic symbols with an intention to communicate something to someone else	To differentiate aided utterances in which symbols are chosen intentionally versus selection of random symbols with no clear communicative purpose	Code each utterance to indicate if communicative intent is *present* or *not present.*

Details

- Code communicative intent as either *present* or *not present*
- Code communicative intent "Present"
  - When the child is producing a meaningful utterance of any length with intent to communicate
  - Code as "present" if any part of the message has clear communicative intent
- Code communicative intent "Not Present"
  - When the child is just "messing" or randomly selecting symbols across an entire utterance
  - If the child randomly selects numerous symbols and is obviously messing around, do not take the time to try to transcribe all of the selected symbols
  - *Utterances for which communicate intent is "Not Present" receive no additional codes*

Number of Symbols

Definition	Purpose	How to calculate
Total number of symbols that the child selects in final message.	This is used to calculate MLU in symbols (MLUSym) and Weighted MLU in symbols (W-MLUSym).	Count the number of symbols, including symbols representing both free and bound morphemes.

Details

Indicate the total number of symbols in each utterance

Adjacent duplications

- Count adjacent duplications only once
- Child: RED COW COW = 2 symbols; do not consider the 2^{nd} COW in any analysis
- Child: RED COW RED = 3 symbols; no two identical adjacent symbols; consider all symbols in all analyses

Number of Relevant Symbols

Definition	**Purpose**	**How to calculate**
Total number of symbols that are relevant to the target/context.	This is designed as a *semantic* measure. The results are used to calculate the percentage of relevant symbols (PRSym)	Count the total number of both free and bound morphemes that are relevant to the target/context

Details

Indicate the total number of relevant symbols in each utterance

Video Probe Data

General rules

- If a symbol is part of the VP target or the list of relevant symbols, it's relevant.
- Both targeted and additional symbols that are counted as relevant for the video probes are included in these Video Probe Accurate Additions spreadsheets.
  - These lists were carefully constructed and agreed upon by multiple study team members.

Bound morphemes

- Bound morphemes are counted as relevant if they are part of the original target or if they contribute logical semantic information.
  - *Exception*: Bound morphemes used as a single utterance are not counted as relevant, even if they appear to be targeted
- Symbols representing free-standing morphemes
  - These are considered "relevant" if the word logically fits in the context of the video, regardless of the target. Examples:
    - HAPPY is relevant in all videos except for those in which the only animal to appear is explicitly sad.
      - Otherwise, the animals can all reasonably be judged to be happy.
    - BLUE is relevant in all videos in which WASH is part of the target, as the washcloth being used is blue.
- Bound morphemes (-S and -ING)
  - Must come immediately <u>after</u> an appropriate free morpheme to be considered relevant
    - *Exception*: If bound morphemes are part of the original target, count the exact number that appear in the target as relevant no matter where they appear in the child's utterance.
    - Examples
      - Target: MONKEY -S RED CAR
      - Child: MONKEY RED -S CAR
        - Relevance: -S counts as relevant

        - Rationale: -S is part of the target.
        - Word order score: 0.5
      - Child: MONKEY RED -S CAR -S
        - Relevance: Only one -S counts as relevant.
        - Rationale: There is only one -S in the target. Coders cannot count both -S as relevant because the first -S is not located immediately after an appropriate free morpheme, and there is only one car in the video.
- If the bound morpheme contributes logical semantic information, count it as relevant.
  - Example 1
    - Target: DIRTY MONKEY IS IN THE GREEN CAR
    - Child: MONKEY -S CAR
      - Relevance: -S ***does*** count as relevant.
      - Rationale: It's logical to think that the car could belong to Monkey
    - Child: MONKEY CAR -S
      - Relevance: -S does ***not*** count as relevant.
      - Rationale: There are not two cars in the video.
  - Example 2
    - Target: PIG IS HUGGING THE RED CAR
    - Child: PIG -S BEHIND CAR
      - Relevance: -S ***does*** count as relevant.
      - Rationale: The -S here can logically be construed as a contraction for 'is.'
  - Example 3
    - Target: DIRTY COW IS HUG –ING THE DIRTY BED
    - Child: COW -S DIRTY -S BED -S DIRTY
      - Relevance: All are relevant except for the middle -S
      - Rationale: Interpreted as: COW IS DIRTY -S (contractible copula, + 1 irrelevant -s); BED IS DIRTY (contractible copula)
      - Word order score: 0.5

Determiners (A and THE)

- Must be before an appropriate free morpheme to be considered relevant
  - *Exception*: If articles are part of the original target, count the exact number that appear in the target as relevant no matter where they appear in the child's utterance.
- Examples
  - Target: MONKEY IS IN THE RED CAR
  - Child: MONKEY RED THE CAR, *or* MONKEY RED CAR THE
    - Relevance: THE counts as relevant.
    - Rationale: THE is part of the target.
    - Word order score: 0.5
  - Child: MONKEY RED THE CAR THE
    - Relevance: Only one THE counts as relevant.

        - Rationale: There is only 1 THE in the target, and the second THE is not located before an appropriate free morpheme.
- *"To be"* verbs (IS, AM, ARE)
    - Must come after the relevant free morpheme
    - Exception
        - If the 'to be' verbs are part of the original target, count the exact number that appear in the target as relevant no matter where they appear in the child's utterance.
        - Examples:
            - Target: MONKEY IS IN THE RED CAR
            - Child: MONKEY RED IS CAR
                - Relevance: IS counts as relevant
                - Rationale: IS is part of the target.
                - Word order score: 0.5
            - Child: MONKEY RED IS CAR IS
                - Relevance: Only one IS counts as relevant.
                - Rationale: There is only one IS in the target, and the second IS is not located after the relevant free morpheme.

Play Intervention/Play Measurement/Storybook/Activity Scene Data

General rules

- If a selected symbol is in any way relevant to the topic of communication, count the symbol as relevant.
    - In general, give the child the benefit of the doubt
- Obvious errors are not relevant
    - Partial utterances that contain some symbols where the child is just "messing around"
        - Give the child credit for the relevant symbols, and no credit for the irrelevant ones
- If child selects multiple symbols successively from the same category (e.g., prepositions, adjectives), the context determines how many of the symbols are relevant.
    - Example
        - Child: IN ON UNDER BARN
            - It's likely that the child is searching for the correct word, and just one of these prepositions is relevant, so this most likely has 2 relevant symbols
        - Child: BIG BLUE RED DOG
            - If the dog is both big and blue, both are relevant. If the dog is not red and no red dog is relevant in the broader context, then red is not relevant.

Nouns

- *Characters* and *objects* are relevant if they are
    - Mentioned in a recent conversational turn
    - Within view of the child
    - Being requested by or commented upon by the child

Adjectives

- *Adjectives* are relevant if they are
  - Relevant to the current topic of conversation
  - A characteristic of a character or object the child is referring to

Prepositions

- *Prepositions* are relevant if they
  - Describe the location, even if they select the wrong preposition (e.g., IN instead of ON)

Articles

- *Articles* are relevant as long as
  - The utterance contains at least one adjective or noun
  - Examples
    - Child: BLUE THE
      - Relevance: THE is relevant.
      - Rationale: It could be a word order issue (Which one do you want? THE BLUE)
    - Child: IN THE
      - Relevance: THE is ***not*** relevant.
      - Rationale: There is nothing here for the article to be describing

"*To be*" verbs

- *IS, AM*, and *ARE* are
  - Relevant if they are attached to a subject
  - Examples
    - Child: COW IS
      - Relevance: IS is relevant, regardless of the relevance of COW
      - Rationale: IS agrees with the subject.
    - Child: COW ARE
      - Relevance: ARE is relevant, regardless of the relevance of COW
      - Rationale: This utterance lacks SV agreement but the child is moving in the right direction
  - Not relevant if they
    - Seem to appear randomly, with no logical connection to the rest of the utterance
    - Are selected in isolation; e.g., an utterance that is just the symbol *IS*

Bound morphemes

- -ING is relevant if
  - Any lexical verb (that is, any verb other than IS, AM, or ARE) is present in the utterance, even if in the wrong word order.
  - Examples:
    - Child: COW –ING WASH
      - Relevance: -ING is relevant.
      - Rationale: There is a lexical verb WASH even though the word order is wrong.
    - Child: COW –ING or COW IS -ING
      - Relevance: -ING is not relevant.
      - Rationale: There is no lexical verb.

- -S is relevant if
    - The utterance contains a noun, pronoun, or a lexical verb even if in the wrong order
    - Remember that –S can function as a
        - plural (BLUE COW –S)
        - plural (BLUE –S COW)
        - contracted copula (COW -'S WASH –ING THE CAR)
        - 3rd person singular marker (COW WASH –(E)S THE CAR)
        - possessive (COW –'S PLATE)

Adjacent repetitions

- If the child repeats any symbols adjacently, count the symbol only once.

Non-adjacent repetitions

- All non-adjacent repetitions of a relevant symbol count as relevant.
    - Target: RED MONKEY IN CAR.
    - Child: CAR MONKEY CAR IN CAR
        - Relevance: All five symbols are relevant (PRSym = 100%).

Relevance vs. word order

- Do not attend to word order. If a symbol is relevant, count it.

Word Order Score*

Definition	Purpose	How to calculate
Rating of how accurate the word order is. *Accurate vs. complete:* "Correct" word order does not necessarily mean "complete." An utterance can have a score of 1 even if elements are missing.	This is designed to be a *syntactic* measure.	Each utterance receives a score of 0, .5, or 1. • 0 = No discernable word order is apparent. • .5 = Some word order is apparent but is not entirely accurate or clear • 1.0 = Word order has no discernable errors

*Note: This is our initial approach to determining word order. We likely will explore additional avenues.

Details

Single symbol utterances

- Single symbol utterances = 1.0
  - *Exception*: Bound morphemes used as single symbol utterance = 0

General rules:

- If any part of the word order is *inaccurate*, it cannot be scored as a 1.
- If any part of an utterance with flawed word order is *accurate*, such as putting the subject first but the rest is a mess, score this as a 0.5.

Labeling and Listing Within Same Part(s) of Speech

- If the child provides a list of nouns or adjectives by themselves, score this as a 1.
  - SAD HAPPY
  - COW RED BOX BLUE
  - GREEN DOG YELLOW MONKEY (Also see *Adjectives and Subject-Verb-Complement* section)
  - PUSH SHAKE
  - SHAKE PUSH
  - IN ON
  - I YOU

Contextual information

- Take all contextual information into account. If it's clear from the context that there are word order errors, do not give full credit.

***Video Probe Data** Examples*

- Target: BIG PIG IS PUSH -ING THE GREEN BATHTUB
- Child: GREEN PIG BIG HAPPY BATHTUB
  - Word order score: .5

        - Rationale: Pig & bathtub are in the right order; also the bathtub can be considered big
    - Child: GREEN PIG HAPPY BATHTUB
        - Word order score: 0
        - Rationale: No elements are in the right order.
    - Target: I AM ABOVE COW 'S BATHTUB
    - Child: BATHTUB ABOVE
        - Word order score: 0
        - Rationale: No elements are in the right order.

    - Target: ELEPHANT IS IN PIG'S AIRPLANE
    - CHILD: AIRPLANE IN ELEPHANT
        - Word order score: 0
        - Rationale: Elephant and Airplane are in the wrong order.

Relevance vs. word order

- Examine word order without regard to relevance

***Video Probe Data** Examples*

- Target: YOU ARE UNDER THE BLUE BOX
- Child: IS MONKEY BOX BLUE
    - Word order score: .5
    - Rationale: Subject appears before object, so it's not a 0. However, it's also possible that the child meant to produce "Is Monkey's box blue?" However, we are not giving the children credit for possibly trying to create questions during video probes (see below).

Subject-Verb-Object (SVO) vs. Object-Verb-Subject (OVS)

- If the child produces an SVO utterance as an OVS, this is a 0, unless other elements of the utterance have a correct word order.

***Video Probe Data** Examples*

- Target (SVO): THE RED COW IS WASH –ING THE BLUE CAR
- Child: CAR WASH COW
    - Word order score: 0
    - Rationale: OVS production
- Child: CAR WASH RED COW
    - Word order score: .5
    - Rationale: Partial credit for Adj + Noun (RED COW)

Multiple adjectives

- No particular word order is required for adjacent adjectives
    - SAD BLUE COW (1)
    - BLUE SAD COW (1)

Adjectives and Subject-Verb-Complement (SVC)

- If an adjective + noun go together, give credit if it's presented as an S(V)C* utterance:

***Video Probe Data** Examples*

  - Target: RED COW IS ON THE BLUE BOX
  - Child: RED COW
    - Word order score: 1
    - Rationale: Adjective + noun
  - Child: COW IS RED
    - Word order score: 1
    - Rationale: SVC
  - Child: COW RED
    - Word order score: 1
    - Rationale: S(V)C
  - Child: COW BOX LITTLE
    - Word order score: 1
    - Rationale: S(V)C
  - Child: COW RED BOX BLUE
    - Word order score: 1
    - Rationale: 2 S(V)C utterances; e.g. Cow is red, and the box is blue.
  - Child: COW RED BLUE BOX
    - Word order score: .5
    - Rationale: You cannot easily make this into one grammatical utterance
  - *Exception*: If there is more to the utterance than S(V)C, this likely is a word order error; example:
    - Target: RED COW IS ON THE BLUE BOX
    - Child: COW RED ON BOX
      - Word order score: .5
      - Rationale: These are two separate propositions (Cow is red; Cow is on the box); you can't easily make this a grammatically complete utterance that a preschooler would say. ON BOX is correct.
  - Target: COW -'S RED BOX
  - Child: BOX RED COW
    - Word order score: 0
    - Rationale: Same as above; these are two separate propositions (The box is red and/or there's a red cow).
- Brush up on subject complements online as needed.

Same order as target

- If two symbols are in the same order as the target – even if other symbols appear in between – give the child (some) credit.

  ***Video Probe Data** Example:*
  - Target: YOU ARE PUSHING LION -'S BATHTUB
  - Child: LION BOX AIRPLANE BATHTUB
    - Word order score: .5
    - Rationale: LION + BATHTUB are in the correct order

Use of AND

- Score as 1, as "and" implies no particular word order
- Note: The child may be using "and" as part of a listing function
    - HUG AND AIRPLANE
    - AIRPLANE AND HUG
    - PIG AND SAD
    - SAD AND PIG

Clear syntax with no relevance

- In the rare case of an irrelevant utterance with accurate, clear syntax, give the child credit.
- ***Video Probe Data*** *Example:*
    - Target: RED COW IS WASH-ING THE BLUE CAR
    - Child: LION IN BOX
        - Word order score: 1
        - <u>Rationale</u>: Accurate and clear syntax

Question forms

- **Video Probes**: We are assuming the children are not attempting to create questions.
- *Rationale*: To give them credit for this is likely to inflate their scores; it's more likely that they have word order issues vs. trying to form a question.
    - Target: YOU ARE UNDER THE BLUE BOX
    - Child: ARE YOU UNDER BOX
        - Word order score: .5
        - <u>Rationale</u>: Inversion of YOU and ARE

Sometimes, this is just hard & we won't get 100% agreement all the time!

Video Probe Data *Example*

- Target: MONKEY –S DIRTY CAR
- Child: BEHIND MONKEY CAR (UCF4 Mo1 #2)
    - Word order score: Can argue this as a .5 or a 1.0

Subject-Verb Present/ Not Present

Definition	Purpose	How to code
Determine which utterances contain both a subject and verb. Childlike sentences contain both a subject and a verb.	The presence of a subject and verb determines if the utterance can be classed as a sentence or not.	• Code SV "present" vs. "not present" for each utterance

Details

- "SV not present"
    - No SV adjacent to each other, or
    - SV are not in the correct order
        - Child: EAT DOG [VS instead of SV] Note: DOG could also be an object here – that is, someone is eating the dog – in which case, this would be an (S)VO utterance, and SV is still not present.)
- "SV present"
    - Subject-verb MUST be adjacent and in the correct order
        - *Note*: The child does not have to demonstrate SV agreement.
        - Child: DOG DRIVE TRACTOR
            - SV present
        - Child: TRACTOR DOG DRIVE
            - SV present (SV is present and adjacent and in the correct order, even though the object is not in the correct order)
        - Child: DOG CAKE EAT
            - SV not present; DOG (S) and EAT (V) are separated by the object (O)
        - Child: MONKEY –S ON BED
            - This counts as a sentence. The –S could be a contracted copula.
        - Child: PIG 'S IN THE BATHTUB
            - This counts as a sentence. The –S could be a contracted copula.

Modifiers

- *Grammatical intent clear/ Grammatical intent unclear*
    - Details are in the next section; whether an utterance is "clear" or "unclear," complete the modifier re: lexical verbs.
    - *Lexical verb/ Non-lexical verb*
        - Lexical verbs include all verbs (in the current study) other than IS, AM, and ARE
        - Remember: IS, AM, and ARE can function as an auxiliary verb or a main verb.
            - I AM DIRTY = AM is a main V
                - This utterance does not contain a lexical verb
            - I AM WASH CAR = AM is an Aux V, and WASH is the main V
                - This utterance does contain a lexical verb (WASH)

- *Unique Subject-Verb (USV) coding*
    - Each utter is coded as either "USV" or "Repeated SV combination"
    - If USV, transcribe the USV in the comment section.
    - To code USVs, *only analyze the subject and verb*. Ignore the rest of the utterance.
    - USVs contain unique subject + *lexical* verb combinations
    - If the utterance is coded as "non-lexical verb," do not code USV. Examples:
        - MONKEY WASH CAR
            - WASH = Lexical Verb
            - Code the USV: MONKEY WASH
        - MONKEY IS WASHING CAR
            - WASH = Lexical Verb:
            - Code the USV: MONKEY WASH
        - MONKEY IS SAD
            - IS = Main V, which is a non-Lexical Verb
            - Do not Code USV
        - DOG WASH CAR: USV = DOG WASH
        - DOG WASH AIRPLANE: Repeated SV combination of DOG WASH
            - By definition, this is not a USV; it's a repeated SV combination. The core subject-verb combination is the same.
        - Child: PIG 'S WASH IN BATHTUB
            - This counts as a sentence. The –S could be a contracted auxiliary.
            - SV = PIG WASH
        - Child: I –S WASH –S BATHTUB
            - This counts as a sentence. The child could be trying to attach a 'to be' verb to the pronoun "I" (i.e., I IS WASH BATHTUB)
            - SV = I WASH
- *Subject-Verb Agreement*
    - Only use this code for utterances that contain at least one 'to be' verb (IS, AM, ARE)
    - Code modifier "Subject-verb agreement (is, am, are) present" if the subject and 'to be' verb agree. Examples:
        - DOG IS
        - DOG IS EAT
        - DOG IS EAT -ING
        - HIPPO IS BIG
        - I AM HAPPY
        - YOU ARE LITTLE
        - YOU ARE HIDE
        - YOU ARE HIDE -ING

- Code modifier "Subject-verb agreement (is, am, are) absent" if the subject and 'to be' verb do *not* agree. Examples:
    - DOG ARE
    - DOG ARE EAT
    - DOG ARE EAT -ING
    - HIPPO AM BIG
    - I ARE HAPPY
    - I IS BIG
    - YOU AM HIDE
    - YOU AM HIDE –ING
    - YOU IS LITTLE

Grammatical Intent

Definition	Purpose	How to code
Enough grammatical information is present to confidently determine what the child is trying say (i.e., grammatical intent is clear), even if some grammatical elements are missing	This is used to determine whether a sentence is still at Phase 1/Phase 2 vs. moving into the realm of Phase 3/Phase 4; that is, moving into childlike and adultlike sentences	Code grammatical intent as "clear" or "unclear"

Details

- Only use this code for utterances coded as "SV present" (i.e., sentences)
  - Codes: Clear vs. Unclear
- The basic question
  - Is there enough grammatical information there that we know what the child is trying say, and just is missing some grammatical elements? Or are we unsure of what the thrust of the utterance is?
  - The utterance can be a stripped down, childlike sentence and still be clear. Examples:
    - DOG DRIVE CAR
    - DOG DRIVE
  - If there is doubt about what the intended sentence, it's unclear.
- Do not refer back to the target
  - Look at the sentence as its own proposition
    - The idea is to rate the grammatical intent, not the semantics
    - Even if it's not readily apparent what the child is referring to from the video, play context, etc., if the grammatical intent is clear, rate it as "clear."

Tense/Agreement

- Tense markers, such as present progressive –ing vs. past tense –ed, are not required for the sentence to be 'clear'

Word order

- Subject and verb must be in the correct order to be rated as "clear"
- It's possible for sentences to have a word order of .5 and still have clear grammatical intent. This may be rare, but it's possible.
  - ***Video Probe Example***
    - Target: BIG MONKEY IS WASH -ING THE AIRPLANE
    - Child: MONKEY WASH AIRPLANE BIG
      - SV = MONKEY WASH
      - Word order score = .5
      - Grammatical intent = Clear
    - Child: BIG MONKEY AIRPLANE WASH
      - SV = No SV, so do not rate this one
        - SV are not adjacent
      - Word order score = .5

Parts of Speech

Definition	Purpose	How to code
Indicate the presence of the following parts of speech: • Noun • Pronoun • Verb • Preposition • Adjective • Determiner • Conjunction	This measure is used to determine word class diversity of the child's utterances	Select each of the part of speech that is present in the utterance

Details

For each utterance

- Present
  - At least one part of speech is used in the utterance
    - The only time this won't happen is if the utterance consists only of bound morphemes
  - Modifiers: Select each part of speech that is present
    - Only select a part of speech once per utterance, even if it occurs multiple times. Example:
      - DOG DRIVE CAR
        - Noun
        - Verb
      - THE HIPPO -S IN BATHTUB
        - Determiner
        - Noun
        - Verb (Note: the –S is considered to be a contracted copula)
        - Preposition
- Absent
  - No parts of speech are used; i.e., the utterance consists only of bound morpheme(s).

Inflectional Morphemes: Present/ Not Present

Definition	**Purpose**	**How to code**
The available inflectional morphemes include -S and -ING. This measure indicates if either of these are appropriately used in a given utterance.	This is a very broad measure designed to simply capture the child's use of bound grammatical morphemes – any bound grammatical morphemes – over time.	Code each utterance to indicate if any grammatical morphemes are *present and appropriately used* or *not present.*

Details

- An inflectional morpheme is Present if used correctly in the utterance
    - COW HIDE-ING = present
    - DOG -S AIRPLANE = present
    - DOG –S = present
    - HIPPO-S WASH -ING CAR = present
- Inflectional morpheme is Not Present if used incorrectly
    - -S HIPPO BOX = not present
    - DOG -ING CAR = not present
- This is only used once per utterance, regardless of the number of inflectional morphemes.

References

Abbeduto, L., Murphy, M. M., Cawthon, S. W., Richmond, E. K., Weissman, M. D., Karadottir, S., & O'brien, A. (2003). Receptive Language Skills of Adolescents and Young Adults With Down or Fragile X Syndrome. In *American Association on Mental Retardation* (Vol. 149). http://meridian.allenpress.com/ajidd/article-pdf/108/3/149/2028276/0895-8017

Allen, A. A., Schlosser, R. W., Brock, K. L., & Shane, H. C. (2017). The effectiveness of aided augmented input techniques for persons with developmental disabilities: a systematic review. *AAC: Augmentative and Alternative Communication*, *33*(3), 149–159. https://doi.org/10.1080/07434618.2017.1338752

American Speech-Language-Hearing Association (n.d.). *Augmentative and Alternative Communication* (Practice Portal). Retrieved 12, 15, 202, from http:www.asha.org/Practice-Portal/Professional-Issues/Augmentative-and-Alternative-Communication/.

Barker, R. M., Akaba, S., Brady, N. C., & Thiemann-Bourque, K. (2013). Support for AAC use in preschool, and growth in language skills, for young children with developmental disabilities. *AAC: Augmentative and Alternative Communication*, *29*(4), 334–346. https://doi.org/10.3109/07434618.2013.848933

Berglund, E., Eriksson, M., & Johansson, I. (2001). *Parental Reports of Spoken Language Skills in Children With Down Syndrome*.

Beukelman, D. R., & Light, J. C. (2020). *Augmentative & Alternative Communication: Supporting Children and Adults with Complex Communication Needs* (Fifth Edition). Brookes Publishing.

Binger, C., Kent-Walsh, J., Harrington, N., & Hollerbach, Q. C. (2020). Tracking early sentence-building progress in graphic symbol communication. *Language, Speech, and Hearing Services in Schools*, *51*(2), 317–328. https://doi.org/10.1044/2019_LSHSS-19-00065

Binger, C., Kent-Walsh, J., King, M., & Mansfield, L. (2017). Early sentence productions of 3-and 4-year-old children who use augmentative and alternative communication. *Journal of Speech, Language, and Hearing Research*, 60(7), 1930-1945.

Binger, C., Kent-Walsh, J., King, M., Webb, E., & Buenviaje, E. (2017). Early Sentence Productions of 5-Year-Old Children Who Use Augmentative and Alternative Communication. *Communication Disorders Quarterly*, *38*(3), 131–142. https://doi.org/10.1177/1525740116655804

Binger, C., & Light, J. (2007). The effect of aided AAC modeling on the expression of multi-symbol messages by preschoolers who use AAC. *AAC: Augmentative and Alternative Communication*, *23*(1), 30–43. https://doi.org/10.1080/07434610600807470

Binger, C., & Light, J. (2008). The morphology and syntax of individuals who use AAC: Research review and implications for effective practice. *Augmentative and Alternative Communication*, 24(2), 123-138.

Binger, C., Maguire-Marshall, M., & Kent-Walsh, J. (2011). Using aided AAC models, recasts, and contrastive targets to teach grammatical morphemes to children who use AAC. *Journal of Speech, Language, and Hearing Research*, *54*(1), 160–176. https://doi.org/10.1044/1092-4388(2010/09-0163)

Binger, C., Richter, K., Taylor, A., Williams, E. K., & Willman, A. (2019). Error patterns and revisions in the graphic symbol utterances of 3- and 4-year-old children who need augmentative and alternative communication. *AAC: Augmentative and Alternative Communication*, *35*(2), 95–108. https://doi.org/10.1080/07434618.2019.1576224

Caselli, M. C., Monaco, L., Trasciani, M., & Vicari, S. (2008). Language in Italian Children With Down Syndrome and With Specific Language Impairment. *Neuropsychology*, *22*(1), 27–35. https://doi.org/10.1037/0894-4105.22.1.27

Castilla-Earls, A., & Fulcher-Rood, K. (2018). Convergent and divergent validity of the grammaticality and utterance length instrument. In *Journal of Speech, Language, and Hearing Research* (Vol. 61, Issue 1, pp. 120–129). American Speech-Language-Hearing Association. https://doi.org/10.1044/2017_JSLHR-L-17-0152

Chapman, R. S., & Kay-Raining Bird, E. (2012). *Language Development in Childhood, Adlescence, and Young Adulthood in Persons with Down Syndrome*.

Cleland, J., Wood, S., Hardcastle, W., Wishart, J., & Timmins, C. (2010). Relationship between speech, oromotor, language and cognitive abilities in children with Down's syndrome. *International journal of language & communication disorders*, 45(1), 83-95.

Cohen, J., Spohn, H., Solomon, L., & Steinman, A. (1960). *A COEFFICIENT OF AGREEMENT FOR NOMINAL SCALES 1*.

Dowden, P. A. (1997). Augmentative and alternative communication decision making for children with severely unintelligible speech. *AAC: Augmentative and Alternative Communication*, *13*(1), 48–59. https://doi.org/10.1080/07434619712331277838

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.

Ganz, J. B., Davis, J. L., Lund, E. M., Goodwyn, F. D., & Simpson, R. L. (2012). Meta-analysis of PECS with individuals with ASD: Investigation of targeted versus non-targeted outcomes, participant characteristics, and implementation phase. *Research in Developmental Disabilities*, *33*(2), 406–418. https://doi.org/10.1016/j.ridd.2011.09.023

Gwet, K. (2002). *Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogeneity*.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. https://doi.org/10.1348/000711006X126600

Hadley, P. A. (2014). Approaching early grammatical intervention from a sentence-focused framework. *Language, Speech, and Hearing Services in Schools*, *45*(2), 110–116. https://doi.org/10.1044/2014_LSHSS-14-0017

Hadley, P. A., Mckenna, M. M., & Rispoli, M. (2018). *Clinical Focus Sentence Diversity in Early Language Development: Recommendations for Target Selection and Progress Monitoring*. https://doi.org/10.23641/asha.5895976

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). *Language Sampling: Does the Length of the Transcript Matter?*

Hickey, J., Shrubsole, K., Worrall, L., & Power, E. (2019). Implementing aphasia recommendations in the acute setting: speech-language pathologists' perspectives of a behaviour change intervention. *APHASIOLOGY*, *33*(5), 606–628. https://doi.org/10.1080/02687038.2018.1561416?needAccess=true

John, A., & Enderby, P. (2000). Reliability of speech and language therapists using therapy outcome measures. *International Journal of Language and Communication Disorders*, *35*(2), 287–302. https://doi.org/10.1080/136828200247197

Kent, R. D., & Vorperian, H. K. (2013). Speech impairment in down syndrome: A review. In *Journal of Speech, Language, and Hearing Research* (Vol. 56, Issue 1, pp. 178–210). https://doi.org/10.1044/1092-4388(2012/12-0148)

Kent-Walsh, J., Binger, C., & Hasham, Z. (2010). Effects of parent instruction on the symbolic communication of children using augmentative and alternative communication during storybook reading. *American Journal of Speech-Language Pathology*, *19*(2), 97–107. https://doi.org/10.1044/1058-0360(2010/09-0014)

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kottner, J., Audige, L., Brorson, S., & Donner, A. (2011). Kottner et al 2011 Reliability study guidelines. *Journal of Clinical Epidermiology*, *64*, 96–106.

Kovacs, T., & Hill, K. (2015). A tutorial on reliability testing in AAC language sample transcription and analysis. *AAC: Augmentative and Alternative Communication*, *31*(2), 148–158. https://doi.org/10.3109/07434618.2015.1036118

Kumin, L. (1994). SYNDROME I N NATURAL SETTINGS: PARENTS' PERSPECTIVE '. In *O Perceptual and Motor Skills* (Vol. 78).

Landis, J. R., & Koch, G. G. (1977). *The Measurement of Observer Agreement for Categorical Data* (Vol. 33, Issue 1). https://about.jstor.org/terms

Laws, G., & Bishop, D. V. M. (2003). A Comparison of Language Abilities in Adolescents With Down Syndrome and Children With Specific Language Impairment. In *Journal of Speech, Language, and Hearing ResearchVol* (Vol. 46).

Martin, G. E., Klusek, J., Estigarribia, B., & Roberts, J. E. (2009). Language Characteristics of Individuals With Down Syndrome. In *Top Lang Disorders* (Vol. 29, Issue 2). http://journals.lww.com/topicsinlanguagedisorders

Martin, G. E., Losh, M., Estigarribia, B., Sideris, J., & Roberts, J. (2013). Longitudinal profiles of expressive vocabulary, syntax and pragmatic language in boys with fragile X syndrome or Down syndrome. *International Journal of Language and Communication Disorders*, *48*(4), 432–443. https://doi.org/10.1111/1460-6984.12019

McDonald, E. T., & Schultz, A. R. (1973). Communication boards for cerebral-palsied children. *Journal of Speech and Hearing Disorders*, 38(1), 73-88.

Messick, S. (1996). *Validity and Washback in Language Testing*.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 36(2), 338-350.

Paul, R., Norbury, C., & Gosse, C. (2018). *Language disorders from infancy through adolescence: listening, speaking, reading, writing, and communicating*. St. Louis: Elsevier.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice* (Vol. 892). Pearson/Prentice Hall.

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). *Mean Length of Utterance in Children With Specific Language Impairment and in Younger Control Children Shows Concurrent Validity and Stable and Parallel Growth Trajectories*.

Roulstone, S., Loader, S., Northstone, K., & Beveridge, M. (2002). The speech and language of children aged 25 months: Descriptive data from the Avon longitudinal study of parents and children. *Early Child Development and Care*, *172*(3), 259–268. https://doi.org/10.1080/03004430212126

Savaldi-Harussi, G., Lustigman, L., & Soto, G. (2019). *The emergence of clause construction in children who use speech generating devices*.

Schlosser, R. (1999). Social validation of interventions in augmentative and alternative communication. *Augmentative and Alternative Communication*, 15(4), 234-247.

Schlosser, R. (2002). On the importance of being earnest about treatment integrity. *Augmentative and Alternative Communication*, 18(1), 36-44.

Shane, H. C., Blackstone, S., Vanderheiden, G., Williams, M., & Deruyter, F. (2012). Using AAC technology to access the world. *Assistive Technology*, *24*(1), 3–13. https://doi.org/10.1080/10400435.2011.648716

Sutton, A., Soto, G., & Blockberger, S. (2002). Grammatical issues in graphic symbol communication. *AAC: Augmentative and Alternative Communication*, *18*(3), 192–204. https://doi.org/10.1080/07434610212331281271

Therrien, M. C. S., & Light, J. (2016). Using the iPad to facilitate interaction between preschool children who use AAC and their peers. *AAC: Augmentative and*

Alternative Communication, *32*(3), 163–174. https://doi.org/10.1080/07434618.2016.1205133

Therrien, M. C. S., Light, J., & Pope, L. (2016). Systematic Review of the Effects of Interventions to Promote Peer Interactions for Children who use Aided AAC. *AAC: Augmentative and Alternative Communication*, *32*(2), 81–93. https://doi.org/10.3109/07434618.2016.1146331

Tomasello, M. (2000). *Do young children have adult syntactic competence?* www.elsevier.com/locate/cognit

Watkins, R. v., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*, *38*(6), 1349–1355. https://doi.org/10.1044/jshr.3806.1349

Wetherby, A. M., & Rodriguez, G. P. (1992). Measurement of communicative intentions in normally developing children during structured and unstructured contexts. *Journal of Speech and Hearing Research*, *35*(1), 130–138. https://doi.org/10.1044/jshr.3501.130