The Assessment of Phase of Preschool Language: Applying the language benchmarks framework to characterize language profiles and change in four- to five-year-olds with autism spectrum disorder

Helen E Flanagan and Isabel M Smith
Autism Research Centre, IWK Health Centre, Halifax, NS, Canada; Departments of Pediatrics and Psychology & Neuroscience, Dalhousie University, Halifax, NS, Canada

Fiona Davidson
Departments of Pediatrics and Psychology & Neuroscience, Dalhousie University, Halifax, NS, Canada

Abstract
Background and aims: We introduce the Assessment of Phase of Preschool Language (APPL), a rating form that characterizes children’s language according to a well-established framework recommended by Tager-Flusberg et al. (2009). The language benchmarks framework defines children’s language as falling at one of the Pre-verbal, First Words, Word Combinations, Sentences, or Complex Language phases for phonology, vocabulary, grammar, pragmatics, and overall language. The APPL is a flexible assessment tool that allows assessors to determine language phase using a range of assessment sources: natural language samples, standardized measures, and/or parent rating forms. Using the APPL, we examined language profiles in four- and five-year-olds with autism spectrum disorder and explored language development during a community-based Naturalistic Developmental Behavioral Intervention program.

Methods: Community speech-language pathologists completed the APPL with 95 four- and five-year-olds at the beginning of the treatment. The APPL was re-administered after a mean of 10 months of intervention (SD = 2 months) for 46 of these children. Children received treatment for up to 15 h per week in their homes and/or community childcare centers. Pivotal Response Treatment was the main form of intervention. The Picture Exchange Communication System or other augmentative and alternative communication systems were also used with many pre-verbal children.

Results: At the beginning of intervention, the most common language phase was Word Combinations (44%), followed by Pre-verbal (26%), Sentences (20%), and then First Words (10%). Only 24% of children had even profiles (i.e. phonology, vocabulary, grammar, and pragmatics skills at the same level). Phonology was a common area of relative strength, and pragmatics was a common area of relative weakness. Ten months of intervention was associated with gains in overall language phase for 37% of children. Approximately half gained at least one phase in Grammar (57%), Vocabulary (51%), and Phonology (46%), while Pragmatics improved for 33%. Gains varied based on initial language phase. Inclusion of skills using augmentative and alternative communication enhanced interpretation of change during intervention.

Conclusions: Four- and five-year-olds with autism spectrum disorder in this sample tended to have uneven skills across expressive language domains. Community-based Naturalistic Developmental Behavioral Intervention was associated with gains in language phase in older preschoolers with autism spectrum disorder. Gains varied across language domains and were influenced by initial language phase.

Implications: The Assessment of Phase of Preschool Language is a useful tool to support consistent application of the language benchmarks framework. It is important to consider all language domains when characterizing language skills and treatment impact in children with autism spectrum disorder.

Corresponding author:
Helen E Flanagan, IWK Health Centre, 5980 University Avenue, Halifax, NS B3K 6R8, Canada.
Email: helen.flanagan@iwk.nshealth.ca

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Children with autism spectrum disorder (ASD) present with varied cognitive, language, adaptive, and social skills. The presence of early functional speech has been identified as an important predictor of better long-term outcomes (Gillberg & Steffenburg, 1987; Howlin, Goode, Hutton, & Rutter, 2004; Mayo, Chlebowski, Fein, & Eigsti, 2013). Although early intervention programs for ASD differ, communication is often a key treatment goal (Rogers, 2006; Schreibman et al., 2015).

**Characterizing language skills**

Despite its importance as a predictor and treatment outcome, functional speech has been a challenging construct to define and measure. Many outcome studies use direct, standardized measures to characterize skills. Although there are important advantages to this assessment approach (established reliability and validity, well-defined norms), there are also disadvantages (Condouris, Meyer, & Tager-Flusberg, 2003). Children with ASD may be less comfortable in novel situations, such as testing environments. In addition, they may have difficulty with test-taking skills such as attending to instructions and focusing on task materials. Most standardized measures of expressive language have few items targeting very early developmental levels (e.g., <24 months), such that a small change in raw scores can have disproportionate effects on standard scores and age equivalencies (Tager-Flusberg et al., 2009). These variables can influence the utility of standardized measures when planning for intervention or assessing change over time.

Parent-report measures and natural language samples are important alternative sources of information to characterize functional language. Parent-report measures, such as the MacArthur-Bates Communicative Development Inventory (Fenson, Bates, Dale, Marchman, Reznick, & Thal, 2007), allow for the collection of information about everyday skills in young children and can be strongly correlated with other measurement approaches (Luyster, Kadel, Carter, & Tager-Flusberg, 2008). Natural language samples can be an especially valuable approach to characterizing early language development (Costanza-Smith, 2010). They provide opportunities to measure a wide range of expressive language abilities, including pragmatic skills, and are often collected in contexts that are familiar to the child. Importantly, natural language samples assess functional use of language: the degree to which the child uses varied language spontaneously, intelligibly, and across contexts, which may be especially valuable when planning for treatment and examining intervention impact.

**The language benchmarks framework**

In 2006–2007, the National Institute on Deafness and Other Communication Disorders convened a panel of researchers with expertise in the study of language development in young children with ASD to address this issue (Tager-Flusberg et al., 2009). Their goal was to explore strategies to encourage greater uniformity in the definition and measurement of spoken language in children with ASD. They recommended the use of a developmental framework with five phases of expressive language development: Phase 1: Pre-verbal; Phase 2: First Words; Phase 3: Word Combinations; Phase 4: Sentences; and Phase 5: Complex Language. They also recommended that all domains of expressive language be considered when characterizing language level: phonology, vocabulary, grammar, and pragmatics. The group provided detailed guidelines about the specific skills that a child must demonstrate to meet each phase in these four domains across a range of assessment sources.

In 2015, Ellawadi and Weismer used the language benchmarks framework to characterize communication profiles in 105 toddlers with ASD aged 23–39 months. They assessed phonology using a speech analysis, vocabulary and grammar using the Preschool Language Scale—Fourth Edition (Zimmerman, Steiner, & Pond, 2002), and pragmatics using the Early Social Communication Scales (Mundy et al., 2003). An acknowledged limitation of their assessment approach was that the highest level of pragmatics attainable was Phase 3: Word Combinations. Results indicated that communication profiles of young children with ASD were often uneven, with only 10% of the sample scoring at the same level across all domains. The two most common profiles were: (1) strength in
phonology, even scores in vocabulary/grammar, and weakness in pragmatics (present in 36% of the sample) and (2) weakness in pragmatics and even scores in other areas (present in 33% of the sample). Other profiles were rare (present in ≤10% of the sample).

In a related study, Kover, Davidson, Sindberg, and Weismer (2014) compared the impact of different language sampling contexts on developmental language phase in 63 three- to four-year-olds with ASD. Language phase was determined using 15-min natural language samples. Across sampling contexts, few children had even skills across domains: 16% had even profiles when the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2008) was used as a sampling context, 6% had even profiles when examiner–child play was used, and 14% had even profiles when parent–child play was used. Of note, this study only included children who produced at least 30 vocal utterances during the 15-min play sessions.

The language benchmarks framework also offers a valuable approach to characterizing treatment impact. For example, Paul, Campbell, Gilbert, and Tsiouri (2013) found that approximately half of a sample of 22 minimally verbal children with ASD aged 2–6 years assigned to receive either a discrete trial training intervention or a milieu communication training intervention attained benchmarks consistent with the First Words phase of language development. They used the expressive language domain of the Vineland Adaptive Behavior Scales—II (VABS-II; Sparrow, Balla, & Cicchetti, 2005), the Communication and Symbolic Behavior Scales—Behavioral Observation (Wetherby & Prizant, 2003), and the MacArthur-Bates Communicative Development Inventory (Fenson et al., 2007) to establish language phase. Importantly, Paul et al.’s criteria for attaining the First Words phase for pragmatics were less stringent than those recommended by Tager-Flusberg et al. (2009). Tager-Flusberg et al. recommended that children use words for a minimum of two communicative functions, including commenting (p. 650), whereas Paul et al. accepted any two pragmatic functions, without requiring commenting (e.g. requesting and social interaction).

**Development of the assessment of phase of preschool language**

In previous studies applying the language benchmarks framework, the same measures were used across all children (e.g. Ellawadi and Weismer (2015) used the Preschool Language Scale—Fourth Edition and the Early Social Communication Scales). We propose that the language benchmarks framework can also be meaningfully applied in contexts where a wide range of measures may be used (e.g. clinicians in different regions and multi-site research projects). Condouris et al. (2003) found that key language scores from natural language samples in children with ASD (mean length of utterance and number of different word roots) were significantly correlated with scores tapping similar constructs on standardized. As the criteria proposed by Tager-Flusberg et al. (2009) are stringent and robust, clinicians and researchers could, in principle, determine a child’s phase using varied measurement sources.

We further propose that there are advantages to including skills using augmentative and alternative communication (AAC) when characterizing language level in children with ASD. Tager-Flusberg et al. (2009) acknowledged that some children with ASD successfully acquire expressive language skills using AAC systems, but limited their recommendations about measures and benchmarks to spoken language. Recent estimates suggest that approximately 30% of children with ASD have minimal spoken language ability as they transition from preschool to school-age (see Tager-Flusberg & Kasari, 2013). Various AAC systems are taught to support non- or minimally verbal children (Mirenda, 2003), including sign language and the Picture Exchange Communication System (PECS; Frost & Bondy, 1994), and electronic tablet-based speech-generating devices are also increasingly used (van der Meer & Rispoli, 2010). Inclusion of skills with AAC when characterizing language level allows clinicians and researchers to capture important, functional communication skills in preschool children who nonetheless remain non-verbal; skills that are often the focus of intervention and can substantially improve quality of life.

We (authors HEF and IMS) developed the Assessment of Phase of Preschool Language (APPL) to support assessors in evaluating children using the language benchmarks framework. The APPL outlines possible sources of information that can be used to meet each language phase within each language domain (natural language samples, direct assessment measures, and/or parent rating forms; see Measures). APPL guidelines were derived primarily from the Tager-Flusberg et al. (2009) paper. For example, they include an instruction to omit echoic language during natural language samples, and the expectation that the natural language sample involve frequent presses from an adult, such as a clinician or parent, to communicate (pp. 645–646). In addition, examples of appropriate direct assessment measures tapping specific language domains are based
on those outlined on pages 648 and 649. A key adaptation from the Tager-Flusberg guidelines is the option to include skills demonstrated using AAC systems when characterizing language phase. For domains other than Phonology, APPL raters can indicate whether specific criteria were met using spoken language and also rate whether the same criteria were met using an AAC system. On the APPL, the same standard for children using an AAC system is applied as for children using spoken language. We acknowledge that this may be a high standard for AAC users that could be adapted in the future as developmental guidelines emerge for AAC use. A further difference from guidelines in Tager-Flusberg et al. (2009) involves recommended length of natural language samples. Tager-Flusberg et al. (2009) suggested that natural language samples “will typically be at least 30 minutes in length” (p. 645), whereas APPL guidelines suggest a minimum duration of at least 20 min (>30 if possible), with at least 50 spontaneous utterances if feasible based on the child’s developmental level. This is consistent with research showing that relatively short natural language samples can provide reliable information about key language skills (see Guo & Eisenberg, 2015) and increases efficiency of natural language sample collection and transcription when a clinician considers samples shorter than 30 min to be appropriate. Finally, the APPL includes more detailed guidelines regarding calculation of mean length of utterance than are provided by Tager-Flusberg et al. (2009). Guidelines were adapted from Lund and Duchan (1993) and included in the APPL to promote consistent scoring across raters.

**The current study**

In this study, we used the APPL to extend previous research exploring language skills and profiles in children with ASD using the language benchmarks framework. We also explored changes in language phase during intervention in a subset of our larger sample. Benchmark criteria were carefully applied, and no children were excluded due to comorbid diagnoses or low skill levels. We addressed the following specific research questions:

1. Prior to intervention:
   a. How is language phase characterized using the APPL in four- to five-year-old children with ASD assessed by community-based S-LPs?
   b. How does inclusion of AAC skills alter characterization of language phase?
   c. Is language phase similar or different across domains? Which domains are most often areas of strength or weakness?
2. In a sub-set of children re-assessed after intervention:
   a. How do APPL scores change during participation in a community-based Naturalistic Developmental Behavioral Intervention (NDBI) program providing Pivotal Response Treatment (PRT) and the PECS?
   b. How does inclusion of AAC skills alter characterization of change during treatment?

**Method**

**Participants**

This study was carried out within the context of a larger program of research exploring the comparative effectiveness and costs of publicly funded early intervention programs in eastern Canada (the Preschool Autism Treatment Impact (PATI) Study; Smith et al., 2019). The children in the current study were a convenience sample of children from the larger study whose Speech-language pathologists (S-LPs) chose to participate in the evaluation of the APPL. Participants were 95 preschoolers with ASD, all aged four or five years (87% boys), with a subset of 46 children included in change-over-time analyses (see Table 1 for demographic information). The APPL was only available at the end of intervention for some children because of differences in clinical practice across provincial regions and over time. Children whose data were included in change-over-time analyses did not differ significantly from those with only start-of-intervention data \( p \geq .47 \); see Table 1). Some S-LPs completing the APPL at the end of intervention had access to scores at treatment start. In 58% of cases, the same S-LP completed the APPL at both time points.

All participants were beginning the Nova Scotia Early Intensive Behavioral Intervention program (NS EIBI; Bryson et al., 2007) in one of six regions of a small province. Approval was granted by the IWK Health Centre Research Ethics Board and by REBs in each of the provincial health authorities that existed in Nova Scotia at that time. The NS EIBI program is a publicly provided program open to all preschool children with ASD, regardless of symptom severity or the presence of comorbid conditions. As shown in Table 1, adaptive functioning ranged from the Delayed to Average range, and total scores on the Social Responsiveness Scale, second edition (SRS-2) ranged from the Severe to the Normal range. To be eligible for the EIBI program, children must have an ASD diagnosis by a specialist clinician, based on clinical judgment informed by the ADOS and a detailed developmental history (typically the Autism Diagnostic Interview—Revised; Rutter, LeCouteur, & Lord, 2003). Analyses in Smith et al. (2019) suggest that demographic characteristics for families participating in research on the NS EIBI program are similar to census figures.
Measures

**VABS-II (Sparrow, Balla, & Cicchetti, 2005).** The VABS-II is a parent-report measure assessing adaptive functioning in the following four areas: Communication, Socialization, Daily Living Skills, and Motor Skills. In this study, the VABS-II was completed as a caregiver rating form at the beginning of intervention, as part of the larger PATI study (Smith et al., 2019). The Adaptive Behavior Composite (overall score) was used as a demographic variable, and the age-equivalent score from the Expressive language sub-domain was used to assess APPL validity.

**SRS-2 (Constantino & Gruber, 2012).** The SRS-2 is a parent rating scale that was designed to be a continuous measure of ASD symptoms. Recent research (Hus, Bishop, Gotham, Huerta, & Lord, 2013) suggests that scores are better interpreted as indicating general levels of impairment, as they are influenced by variables such as cognitive level and challenging behavior. The SRS-2 was completed at the beginning of intervention as part of the larger PATI study (Smith et al., 2019). The total T score was included in this study as a demographic variable.

**APPL**

**Description.** We (authors HEF and IMS) created the APPL to condense guidelines from the language benchmarks framework (Tager-Flusberg et al., 2009) into a clinical rating form for use in clinical or research contexts. The goal was to provide a platform through which the benchmarks framework could be easily understood and applied by those assessing individual children with ASD. The APPL allows assessors to determine language phase for individual children using any of several available assessment approaches that are consistent with APPL guidelines. All domains of the APPL at all phases can be completed by collecting and analyzing a natural language sample. Standardized child testing and/or standardized parent report can be used as an alternative or supplementary assessment source for all language domains at most phases (e.g. a parent rating form assessing pragmatics can be used to determine whether a child has attained phase 3, 4, or 5 of the Pragmatics domain; see Table 2). Aside from a requirement that early phonological skills be assessed using a natural language sample, no particular type of measure is obligatory at any stage.

A page of the APPL is dedicated to each expressive language domain: phonology, vocabulary, grammar, and pragmatics (see Figure 1 for the Vocabulary domain and supplementary material for the full APPL). For each domain, the APPL outlines the range of demonstrated skills that could meet criteria for each phase: Phase 2: First Words; Phase 3: Word Combinations, Phase 4: Sentences, or Phase 5: Complex Language, based on minimum criteria suggested by Tager-Flusberg et al. (2009). A child who...
does not meet the First Words phase criteria is considered to be at Phase 1: Pre-verbal. Assessors can use a range of possible assessment approaches (natural language samples, parent rating forms, and/or direct assessment measures; see Table 2). If multiple approaches are used to assess the same language domain (e.g. a natural language sample and a parent rating form to assess Vocabulary), the child need only meet minimum criteria based on one of these sources.

Consistent with recommendations by Tager-Flusberg et al. (2009, p. 647), to meet a specific level of overall language functioning (e.g. overall level of First Words), the child must meet at least one benchmark in every language domain that defines that phase. On the APPL, minimum criteria for language benchmarks can be met either using spoken language or an AAC system. In this study, children could move beyond Phase 1 while remaining non- or minimally verbal, if they met criteria using an AAC system.

**Current study.** In this study, the APPL was completed by 13 different S-LPs working in six regions of the NS EIBI program. It was administered at the beginning of the NS EIBI program for all children in this study and after approximately 10 months of intervention for a representative sub-sample. Prior to this study, S-LPs in the NS EIBI program were already assessing children’s language skills using natural language samples and standardized assessment measures as part of routine clinical care. The APPL was introduced in the context of the Preschool Autism Treatment Impact Study (see Smith et al., 2019). Clinicians from across the province attended an hour-long teleconference session in which the APPL was introduced and described. The APPL form was then distributed. The only guidelines regarding measure administration and selection were those built into the APPL measure itself, as described below (see also Table 2 and supplementary material).

Although not an explicit instruction, clinicians completing the APPL in this study consistently administered a natural language sample. Natural language samples were part of routine clinical practice in the NS EIBI program and can be used to complete the APPL for all language domains at all phases. APPL guidelines indicate that the natural language sample should involve at least 50 spontaneous utterances, if feasible, for a given child, be at least 20-min long, and involve a range of communicative contexts as appropriate to the child, including opportunities to request, comment, answer questions, have a conversation, and tell a narrative. In this study, mean duration of natural language samples was 31.39 (SD = 18.73) min (based on data available from 78% of assessments) and mean length of utterances was 64.44 (SD = 45.32). All natural language samples were at least 10-min long, and 83% were at least 20-min long. Ninety-four percent of natural language samples were either 20-min long or contained at least 50 spontaneous utterances. All natural language samples were administered in children’s homes or community childcare settings, in the context of play. The APPL does not specify that a specific communication partner (e.g. parent and clinician) interact with the child for the natural language sample. As play partner can vary, the APPL contains guidelines indicating that the clinician must ensure that there are frequent presses to communicate and a range of communicative contexts appropriate to the child’s developmental level (e.g. opportunities to request, answer questions, or have a conversation). In this study, natural language samples consistently involved the child interacting with one adult at a time in the

| Table 2. Assessment sources appropriate for demonstrating attainment of specific phases for each language domain on the APPL. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Phonology**   | **Vocabulary**  | **Grammar**     | **Pragmatics**  |
| Phase 2         | NLS             | NLS             | NLS            |
|                 | Parent report   | Direct assessment | Parent report |
| Phase 3         | NLS             | NLS             | NLS            |
|                 | Parent report   | Direct assessment | Parent report |
| Phase 4         | NLS             | NLS             | NLS            |
|                 | Direct assessment | Direct assessment | Parent report |
| Phase 5         | NLS             | NLS             | NLS            |
|                 | Direct assessment | Direct assessment | Parent report |

NLS: natural language sample.
context of play in a familiar setting (e.g. home or pre-
school). Twenty-three percent of S-LPs administered
the natural sample themselves, and 34% combined a
segment they administered with a segment administered
by a parent or familiar front-line staff member. For the
remaining videos, S-LPs ensured that a sample of the
child interacting with a parent (31%) or with a familiar
staff member (13%) met APPL criteria. S-LPs were
often present when the natural language sample was
obtained, which provided the opportunity to support
the communication partner in providing specific
presses, if needed. A quality review was carried out
by author FD for a subsample of 30 video-recorded
natural language samples, which confirmed that 100%
involved frequent presses to communicate, and 97%
involved a range of appropriate communicative
contexts.

If a natural language sample is available and com-
prehensively coded, as was the case in this study, direct
assessments and parent forms are not needed to

**VOCABULARY**

| Phase | Description |
|-------|-------------|
| **PHASE 2** | In a 20-minute period involving frequent presses to communicate, uses at least 5 different words referentially, and says words at least 20 times (i.e., 5 types, 20 tokens; e.g., *more* 10 times, *tree* twice, *ball* 5 times, *up* twice, *down* once) |
| | 15-month level on a direct assessment or parent report measure of expressive language |
| **PHASE 3** | In a 20-minute period involving frequent presses to communicate, uses at least 30 different words referentially |
| | 24-month level on a direct assessment or parent report measure of expressive language |
| **PHASE 4** | In 50 utterances, uses at least 70 word roots* |
| | 36-month level on a direct assessment measure of expressive language |
| **PHASE 5** | In 50 utterances, uses at least 105 word roots* |
| | 48-month level on a direct assessment measure of expressive language |

* A word root is the basic lexical unit of a word, without pre-fixes and suffixes; e.g., if the child says “sit”, “sat”, “cat”, and “cats” they get credit for two root words, cat and sit.

**Vocabulary Phase Attained**

Locate the highest phase with at least one ✓ or AAC above. This is the child’s language phase, even if other items within or below that phase are rated X or N/A.

| Phase 1 “Pre-verbal communication” | Phase 2 “First words” | Phase 3 “Word combinations” | Phase 4 “Sentences” | Phase 5 “Complex language” | Comments: |
|-------|-------------|-----------------|-----------------|-----------------|-----------------|

**Figure 1.** The vocabulary domain of the Assessment of Phase of Preschool Language (APPL).
determine APPL phase. For most assessments in this study (59%), clinicians chose to supplement information from natural language samples with information from at least one standardized parent rating form or direct assessment measure. The APPL contains examples of specific direct assessment measures and parent rating forms for characterization of language phase within specific expressive language domains (see supplementary material), and a wide range of different measures was administered (see Table 3). The decision to administer standardized measures was influenced by a number of variables. In some cases, a clinician may have chosen to administer a direct assessment measure as they believed that it might add meaningfully to characterization of a child’s skills (e.g. a child who appeared anxious and reserved during unstructured play). Other variables influencing administration included duration since a child’s most recent comprehensive language assessment and variables related to intervention (e.g. administration of a direct assessment of phonology prior to targeting articulation). To determine the influence of adding information from standardized measures to information from natural language samples when characterizing language phase, we compared classifications including all available measures (characterizations used throughout this paper) with classifications based only on natural language samples, using weighted kappa for ratings from the beginning of intervention. Agreement was excellent between classification approaches for all domains: Phonology ($K_w = .83$, 95% CI (.67-.98), $n = 24$), Vocabulary ($K_w = .86$, 95% CI (.78-.94), $n = 43$), Grammar ($K_w = 1.00$, 95% CI (1.00–1.00), $n = 35$), and Pragmatics ($K_w = .90$, 95% CI (.76–1.00), $n = 26$). Inclusion of direct assessment measures never increased language level by more than one phase.

**Inter-rater reliability.** The inter-rater reliability of the APPL was assessed for 22 children using weighted kappa ($K_w$; 23% of sample, from five geographical

### Table 3. Measures administered by community-based speech-language pathologists completing the APPL.

| Measure Type                                   | Prior to intervention ($n = 95$) | Following intervention ($n = 46$) |
|------------------------------------------------|----------------------------------|----------------------------------|
| Natural language sample                        | 100%                             | 100%                             |
| Direct assessment measure of phonology         | 25%                              | 22%                              |
| Parent-report measure of expressive language   | 6%                               | 7%                               |
| MacArthur-Bates communication development inventory (Fenson et al., 2007) | 1%                               | 2%                               |
| Direct assessment measure of expressive language | 19%                             | 4%                               |
| Expressive domain, Clinical Evaluation of Language Fundamentals - 4 (Semel, Wiig, Secord, & Langdon, 2006) | 5%                              | 4%                               |
| Expressive domain, Clinical Evaluation of Language Fundamentals Preschool—2 (Wiig, Secord, & Semel, 2004) | 11%                             |                                   |
| Expressive domain, Rossetti Infant-Toddler Language Scale (Rossetti, 2006) | 8%                              | 2%                               |
| Expressive domain, Preschool Language Assessment Instrument—2 (Blank, Rose, & Berlin, 2003) | 4%                              |                                   |
| Test of Early Language Development—3 (Hresko, Reid, & Hammill, 1999) | 1%                              |                                   |
| Parent-report measure of pragmatics            | 27%                              | 22%                              |

Note. A range of measurement sources can be used to complete the APPL. Evidence from at least one appropriate source is needed to meet criteria within each domain.

*Natural language samples were administered for all children and are used to complete all APPL domains.

*Parent-report measures of expressive language were used to complete the Vocabulary domain of the APPL.

*Direct assessment measures of expressive language were used to complete the Vocabulary and the Grammar domains of the APPL.

APPL: Assessment of Phase of Preschool Language.
sites). To do so, the natural language sample administered as part of the APPL was video-recorded and shared with the research team; relevant scores from parent ratings and direct assessment measures for these children were also shared, all with parental consent. Two trained S-LP student research coders used all information combined from these sources to categorize language phase for each domain and for overall language as falling at either the Pre-verbal, First Words, Word Combinations, Sentences or Complex Language phase. Children included in reliability analyses did not differ significantly from others in the study (p ≥ .62 for age, autism symptom severity, and overall expressive language phase at intervention start). Weighted kappa for overall APPL score was 0.77, 95% CI (0.62–0.91), which is considered substantial (Landis & Koch, 1977). Reliability for specific domains were: phonology, Kw = 0.76, 95% CI (0.61–0.92); vocabulary, Kw = 0.75, 95% CI (0.57–0.93); grammar, Kw = 0.73, 95% CI (0.89–0.82); and pragmatics, Kw = 0.70, 95% CI (0.51–0.90).

Validity. To carry out a preliminary exploration of the validity of the APPL, we explored whether children at higher overall APPL language phases had greater mean age-equivalent scores on the expressive language sub-domain of the VABS-II (available for 83% of the sample). Similar to the APPL, the VABS-II measures a child’s functional communication skills in their everyday environment. A one-way ANOVA confirmed that VABS-II age-equivalent scores are higher at higher APPL phases, F (3, 75) = 54.68, p < .001. Adjusted post-hoc pairwise comparisons showed that VABS-II scores at Phases 1–4 differed significantly from one another (all p ≤ .021).

Intervention

The NS EIBI program is publicly funded and involves up to one year of behavioral intervention in natural settings (see Bryson et al., 2007; Smith, Flanagan, Garon, & Bryson, 2015). Children in this study participated for an average of 10 months (SD = 2 months). EIBI staff members provide direct intervention to children in homes and/or community daycares for 15 h/week for the first six months, 12 h/week for the next three months, and 6 h/week for the final three months. Individualized parent coaching is an integral part of the program, with the goal of extending intervention across daily routines. Parents receive four consecutive two-hour coaching sessions at the beginning of intervention, with monthly follow-up.

The main treatment targets of the NS EIBI program are expressive communication, play, and other functional skills (e.g. toileting, safety, and reducing challenging behavior). The primary treatment approach is Pivotal Response Treatment (see Koegel, Openden, Freeden, & Koegel, 2006), which is a NDBI (Schreibman et al., 2015). To build expressive communication with PRT, clinicians and parents follow a child’s lead to establish preferred objects and activities as part of daily routines in the natural environment. They then pause activities using a strategy called shared control and present clear language opportunities. For example, if a child wants a ball, the adult might hold the ball (i.e. shared control) and provide a verbal model such as “ball”, “throw ball”, or “throw me the red ball”, tailored to the child’s language level. Adults support children to provide reasonable attempts at responding to language opportunities before allowing play to continue (i.e. providing reinforcement). Reasonable attempts vary across children, and can include intentional vocalizations for Pre-verbal children.

In the NS EIBI program, AAC systems are introduced for Pre-verbal children if they do not progress using PRT. The usual AAC approach is PECS (Frost & Bondy, 1994), with a focus on ensuring that a child progresses through Phases 1 (How to Communicate) and 2 (Distance and Persistence) before beginning Phase 3 (Picture Discrimination). Some children are transitioned to an electronic tablet-based AAC system after attained Phase 3B of PECS (Picture Discrimination with Preferred Items; for this subgroup, a decision about whether to introduce an iPad®-based system is made collaboratively with parents). When supporting a child to use an iPad®-based AAC system, clinicians ensure that the child maintains skills learned through the PECS system, such as travelling to the communication partner and attaining his/her attention prior to communicating.

Results

Language phase and profiles in four- and five-year olds prior to intervention

The APPL was completed for 95 children at the beginning of the NS EIBI program. The most common overall language phase was Word Combinations (44%), followed by the Pre-verbal phase (26%), the Sentences phase (20%), and the First Words phase (10%). No child’s language fell at the Complex Language phase (0%) prior to intensive service (see Table 4).

For all study results, language phase was determined based on skills using either spoken language or AAC, whichever was highest (see Methods). When the NS EIBI program began, 11 of the 95 children were using an AAC system. Seven were using PECS (four at a very early stage), two were using Proloquo2Go (Sennott &
For all 11 children using an AAC system at the beginning of intervention, spoken language skills for vocabulary, grammar, and pragmatics were at the Pre-verbal phase. For 7 of the 11 children, inclusion of skills using AAC did not affect characterization of language level, as AAC skills were also at the Pre-verbal phase. For four children (4% of overall sample), inclusion of AAC skills did affect language phase for at least one domain, with overall language phase affected for one child (1% of overall sample; see Table 4). Notes in Table 4 outline proportions of the sample at each language phase considering only spoken language (i.e., excluding AAC skills).

Table 5 shows expressive language profiles at the beginning of intervention. Only three profiles were present in more than 5% of the sample. First, 24% of children had phonology, vocabulary, grammar, and pragmatics skills at the same level (even profiles). A second relatively common profile was phonology, grammar, and vocabulary skills at the same level and weakness in pragmatics (21% of the sample). The third profile was vocabulary, grammar, and pragmatics skills at the same level and strength in phonology (18% of the sample). The remaining children had different profiles of expressive language skills; 19 other profiles each occurred in ≤ 4% of the sample (see Table 5).

To further understand profiles of expressive language skills in children with ASD, we explored which language domains were most likely to be areas of relative strength (higher than at least two of the other three language domains) or of relative weakness (lower than at least two of the three other language domains). As shown in Table 6, phonology was the most common area of relative strength (41% of children), and

---

Table 4. Language phase on the APPL at the beginning of intervention in four- to five-year olds with ASD; N = 95.

| Phase                  | Phonology n (%) | Vocabulary n (%) | Grammar n (%) | Pragmatics n (%) | Overall language n (%) |
|------------------------|-----------------|------------------|--------------|------------------|------------------------|
| Phase 1:               |                 |                  |              |                  |                        |
| Pre-verbal             | 11 (11.6%)      | 22 (23.2%)       | 31 (32.6%)   | 23 (24.2%)       | 25 (26.3%)             |
| Phase 2:               |                 |                  |              |                  |                        |
| First words            | 9 (9.5%)        | 9 (9.5%)         | 6 (6.3%)     | 9 (9.5%)         |                        |
|                        | n = 3 AAC<sup>a</sup> |                  | n = 1 AAC<sup>c</sup> |                  |                        |
| Phase 3:               |                 |                  |              |                  |                        |
| Word combinations      | 30 (31.6%)      | 24 (25.3%)       | 24 (25.3%)   | 45 (47.4%)       | 42 (44.2%)             |
|                        |                 |                  | n = 1 AAC<sup>b</sup> |                  |                        |
|                        |                 |                  | n = 1 AAC<sup>c</sup> |                  |                        |
| Phase 4:               |                 |                  |              |                  |                        |
| Sentences              | 25 (26.3%)      | 26 (27.4%)       | 14 (14.7%)   | 20 (21.1%)       | 19 (20.05%)            |
| Phase 5:               |                 |                  |              |                  |                        |
| Complex language       | 20 (21.1%)      | 14 (14.7%)       | 26 (27.4%)   | 1 (1.1%)         | 0 (0%)                 |

<sup>a</sup>For vocabulary, three children who attained the First Words phase did so using an Augmentative and Alternative Communication (AAC) system. Considering only spoken language/excluding AAC: 25 (26%) were Pre-verbal and 6 (6%) were First Words.

<sup>b</sup>For grammar, one child who attained the Word Combinations phase did so using AAC. Considering only spoken language/excluding AAC: 32 (34%) were at Phase 1 and 23 (24%) were at Phase 2.

<sup>c</sup>For pragmatics, one child who attained First Words and one child who attained Word Combinations did so using AAC. Considering only spoken language/excluding AAC: 25 (26%) were at Phase 1, 5 (5%) were at Phase 2, and 44 (46%) were at Phase 3.

<sup>d</sup>For overall language, one child who attained the First Words phase did so using AAC. Considering only spoken language/excluding AAC: 26 (27%) were at Phase 1 and (8%) were at Phase 2.
pragmatics was the most common area of relative weakness (43% of children). Relative weaknesses in vocabulary and strengths in grammar were also relatively common, each present in approximately 20% of the sample.

Changes in APPL scores after intervention

The APPL was completed at the end of intervention for 46 of the 95 children, at a mean age of five years, six months ($SD = 4$ months). Mean duration of intervention for these children was 10 months ($SD = 2$ months). A Wilcoxon signed-rank test showed that Overall Language Phase improved significantly ($Z = -3.84$, $p < .001$). Significant gains were observed in all language domains: Phonology ($Z = -3.05$, $p = .002$), Vocabulary ($Z = -3.90$, $p < .001$), Grammar ($Z = -4.25$, $p < .001$), and Pragmatics ($Z = -2.71$, $p = .007$).

We explored the proportions of children who experienced change in specific language domains during treatment (see Table 7). Children beginning at Phase 5 for a specific language domain typically remained at Phase 5 (10/11 for phonology, 14/14 for grammar, 4/6 for vocabulary, and 0/1 for pragmatics). Therefore, for each expressive language domain, children were included in analyses if they began treatment at Phase 4 or below. Approximately, half of the sample gained at least one phase in each of phonology, vocabulary, and grammar. Pragmatic language phase improved for approximately one in three children. Overall phase, which requires a child’s most impaired area of expressive language to change, improved for 37% of the sample. Decreases in language phase were rare, especially for vocabulary, grammar, and overall language. Overall language phase decreased for one child who was using an AAC system to communicate at both time points (this decrease was due to a change in pragmatics: the child did not comment at follow-up during the natural language sample).

We also explored changes in overall language phase for subgroups that began intervention at the Pre-verbal, First Words, Word Combinations, or Sentences levels (see Figure 2). Improvements in overall language phase were most common for children who began intervention at the Word Combinations phase (56% improved), followed by the First Words phase (33%), the Pre-verbal Phase (25%), and then the Sentences Phase (20%).

An AAC system was taught for 11 of the 46 children included in change-over-time analyses. Eight used PECS only, one used PECS and then Proloquo2go, one used PECS and then TouchChat (Silver Kite, 2016), and one used Speak for Yourself. Including AAC skills in APPL ratings affected characterization of change during intervention. Despite remaining at the Pre-verbal phase for vocabulary, grammar, and pragmatics based on spoken language after participating in the NS EIBI program, 45% attained Phase 2 for vocabulary using AAC (vs. 18% at baseline), 27% attained Phase 3 for grammar using AAC (vs. 9% at baseline), 18% attained Phase 3 for pragmatics using AAC (vs. 9% at baseline; none met Phase 2), and 18% attained Phase 2 for overall language using AAC (vs. 9% at Baseline). Notes in Figure 2 show the proportion of the sample that changed overall language phase considering only spoken language (i.e. excluding AAC skills). While 8% of Pre-verbal four- and five-year olds moved to the First Words phase during intervention based only on spoken language, 25% moved to the First Words based on either spoken language or AAC.

### Table 6. Areas of relative strength and weakness in specific language domains at the beginning of intervention; $N = 95$.

|          | Phonology | Vocabulary | Grammar | Pragmatics |
|----------|-----------|------------|---------|------------|
| Area of strength | 41%       | 11%        | 22%     | 2%         |
| Area of weakness | 9%        | 17%        | 9%      | 43%        |

Note: A domain was considered an area of relative strength if language phase was at least one phase higher in that domain than in two or three of the other three domains. A domain was considered an area of relative weakness if language phase was at least one phase lower in that domain than in two or three of the other three domains.

### Table 7. Change in specific language domains during community-based natural developmental behavioral intervention, $N = 46$.

|                          | Phonology ($N = 35$) | Vocabulary ($N = 40$) | Grammar ($N = 32$) | Pragmatics ($N = 45$) | Overall language phase ($N = 46$) |
|--------------------------|----------------------|-----------------------|--------------------|-----------------------|-----------------------------------|
| Improved two phases      | 9%                   | 3%                    | 16%                | 11%                   | 0%                                |
| Improved one phase       | 37%                  | 48%                   | 41%                | 22%                   | 37%                               |
| Stayed the same          | 46%                  | 48%                   | 44%                | 62%                   | 61%                               |
| Decreased one phase      | 9%                   | 3%                    | 2%                 | 2%                    | 2%                                |
| Decreased two phases     |                       |                       |                    |                       | 2%                                |

Note: For each domain, analyses include children who began intervention at Phase 4 or lower.
Discussion

There is a pressing need to establish consistent frameworks to characterize skill levels in children with ASD (Kasari & Smith, 2019). We developed a new rating form, the APPL, to facilitate consistent implementation of the Tager-Flusberg et al. (2009) language benchmarks framework by clinicians or researchers assessing individual children with ASD. The language benchmarks framework allows for the categorization of language phase based on a range of possible assessment sources, including natural language samples and parent-report measures, as well as standardized clinician-administered measures. Natural language samples and parent-report measures capture important information about communication, especially in early communicators and in children who may be less verbal in unfamiliar contexts such as testing. Importantly, the language benchmarks framework is designed to incorporate skills across all domains of language, including pragmatics and phonology, areas that are not consistently tapped by commonly used direct assessment measures.

On the APPL, raters can indicate whether a child met language benchmarks using either spoken language or an AAC system. This acknowledges that some children with ASD develop meaningful communication using AAC (large vocabularies, use of pictures across multiple functions, and including commenting). This is an important consideration when conceptualizing intervention outcomes and predicting later development, given that a functional communication system is generally a key goal of intervention programs. The same APPL standards are applied across communication

Figure 2. Changes in Overall language during intervention for children starting at different language phases (N = 46).
Note: Two children who started at Phase 1 advanced to Phase 2 using AAC, and the one child who decreased from Phase 2 to Phase 1 used AAC. Considering only spoken language/excluding AAC: 13 children starting at Pre-verbal phase, with 8% advancing to First Words and 92% remaining Preverbal; five children starting at First Words, with 40% advancing to Word Combinations and 60% remaining at First Words.

![Fig 2](image.png)
modalities. This is a high standard for AAC users and could be adapted to reflect emerging research specific to AAC users. In this study, language phase was determined based on skills using either spoken language or AAC, whichever was highest.

In this study, the APPL was administered by 13 S-LPs in a province-wide community-based NDBI program primarily providing PRT. The treatment program is publicly provided, and no children are excluded based on skill level or symptom severity. Ninety-five children were assessed at the beginning of intervention and of them, 46 were re-assessed after approximately 10 months of treatment. Inter-rater reliability of APPL characterization was substantial, and validity was supported by relationships between APPL scores and age-equivalent scores on the expressive language domain of the VABS-II.

First, we explored APPL profiles in four- and five-year olds with ASD at the beginning of intervention. APPL phase varied, with the largest number of children scoring at the Word Combinations phase. Including skills using AAC affected overall language phase for one child (1% of sample) and affected scores on at least one specific language domain for four children (4% of sample). Similar to findings with younger samples, a relatively small proportion of children had skills that were evenly developed across language domains: 24% in this study; 10% in Ellawadi & Weismer (2015); and 6–14% in Kover et al. (2014). Relative strengths in phonology and relative weaknesses in pragmatics were common (41% and 43% of our sample, respectively) but not universal, replicating findings by Ellawadi and Weismer (2015). These findings highlight the importance of evaluating distinct domains of language when assessing children with ASD. Assessments that focus only on one or two areas may fail to capture important strengths and challenges that affect communication. The observed variability in language profiles also lends support for the approach to characterizing overall language phase suggested by Tager-Flusberg et al. (2009). Defining overall language phase as the highest phase met across all domains may reduce the possibility that clinicians and researchers overestimate language level by failing to consider communication challenges influenced by specific language domains.

Using the APPL, we explored changes in children’s expressive language phase during participation in a community-based NDBI program. Pivotal Response Treatment was the main form of intervention for most children. PECS and other AAC systems were also taught. Thirty-seven percent of children moved to a new phase of overall language during intervention. The NS EIBI program was associated with more gains in phonology, vocabulary, and grammar (approximately half of the sample gained a phase) than pragmatics (approximately one-third of the sample gained a phase). Including AAC skills in ratings had an impact on characterization of change. Eleven of the 46 children (24%) used an AAC system during intervention, and many made gains in specific language domains despite remaining non-vocal. Twenty-five percent of Pre-verbal children moved to the First Words phase using either spoken language or AAC, versus 8% based on spoken language alone. Inclusion of skills using AAC allows for measurement of important treatment gains in key communication skills that can improve children’s and families’ quality of life. These findings add to results from our larger study demonstrating that children participating in the NS EIBI program experienced significant gains in adaptive functioning and significant reductions in challenging behavior during participation in the intervention (see Smith et al., 2019).

To our knowledge, this is the first study exploring changes during intervention using the language benchmarks framework in a sample of children with ASD that spans initial language levels. This study suggests that the NS EIBI model is especially likely to improve the overall language level of children in this four- to five-year age range who begin intervention at the Word Combinations phase (56% improved), although gains were observed for children starting at all levels.

Paul et al. (2013) compared the impact of a discrete trial training intervention versus a milieu communication training intervention on spoken language in pre-verbal preschoolers. A greater proportion of children in that study moved to the First Words phase (50%) than in this one (25%, with some meeting criteria using an AAC system). This may have been influenced by sampling differences (e.g. Paul et al. included pre-verbal children with a nonverbal mental age of at least 12 months and children in whom a generalized motor imitation repertoire could be established). It may also have been influenced by differences in application of the benchmarks framework. In Paul et al. (2013), commenting skills were not required to meet the First Words phase (any two pragmatic functions were accepted). In our study, whereas approximately half of the children at the Preverbal phase for Phonology and Vocabulary moved to the First Words phase in these areas (60% for phonology; 50% for vocabulary (some met vocabulary criteria using AAC)), only 25% met criteria for pragmatics, with 75% failing to meet the commenting requirement.

It is important to note that this study included only four- and five-year olds. Language profiles and changes during intervention are likely to differ for younger preschoolers. A further limitation of this study is that it did not include a control or comparison group. A number of children who were already four or five years old...
when treatment began made gains during 10 months of intervention, moving to a new phase of overall language. However, without a control group, it is not possible to know how many would have made gains without treatment. It is also important to acknowledge that the same clinicians often completed pre- and post-treatment assessments, and that knowledge of pre-treatment scores could have influenced follow-up ratings. Future studies could explore the impact of interventions on language phase using control or comparison groups, as well as blinded outcome ratings.

It is important to acknowledge that differences in measurement approach across APPL administrators may affect characterization of language phase. In previous studies applying the language benchmarks framework, research teams have characterized language phase using a consistent set of measures across all children. In contrast, in this study, S-LPs working in multiple regions within a community-based intervention program determined language phase using a flexible tool that allowed evidence from a range of possible sources to be sufficient in demonstrating that a child had attained a specific language phase for a certain domain (all assessments involved a natural language sample; and 59% incorporated information from at least one standardized direct assessment measure or parent rating form). This increased flexibility will reduce reliability compared to a more consistent testing approach (e.g. an approach in which all children complete a natural language sample during a specific task; and an approach in which grammar and vocabulary is always assessed using a specific standardized tool). For example, Kover et al. (2014) found that the context of the natural language sample affected language benchmark characterization. In the current study, children had varying levels of familiarity with the communication partner during the natural language sample, which could also influence outcomes. Similarly, although Condouris et al. (2003) found that scores from language samples and direct assessment measures were significantly correlated, and noted that measures from natural language samples may be a good proxy for standardized assessment of related constructs, different assessment sources (e.g. direct assessment measures versus natural language samples) may categorize children within different language domains. In the subset of children in this study for whom both a natural language sample and standardized measures were administered, characterization based on all available sources of information was similar to characterization based only on natural language samples, with excellent agreement across domains. There were instances in which administration of a direct assessment measure led to an increase by one phase in the specific language domain assessed. As the APPL defines a child’s phase as the highest level attained using at least one assessment source, the addition of direct assessment measures could not decrease phase below that attained with the natural language sample. Consistency of categorization may be lower if a study using the APPL included some children assessed using only natural language samples and other children assessed using only standardized measures. The APPL includes criteria aimed at minimizing variability in characterization across clinicians, such as specific criteria regarding language sample administration and criteria outlining appropriate rating forms and standardized measures to assess specific language domains. Importantly, characterization of overall language phase requires that a specific level of language be met across all four language domains, which minimizes the impact of differences in categorization in any specific domain.

We believe that important advantages are associated with the APPL’s flexible assessment approach. First, assessors can use clinical judgment to select measures that are the best fit for individual children. As clinicians in this study could determine an appropriate natural language sample context for an individual child, and decide whether to supplement this with information from standardized measures, they had the opportunity to maximize the quality of language phase characterization, increasing the validity of individual APPL scores. Specific assessment approaches may be better suited to certain children, especially when samples contain children at diverse ages and language levels. For example, a direct assessment measure may not sufficiently capture language skills in a child who has attentional or behavioral challenges that limit engagement in standardized testing. Conversely, a natural language sample may not sufficiently capture language skills in a child who is anxious or reserved during unstructured play. A second advantage to the APPL’s flexible measurement approach is that it may maximize the representativeness of study samples. For example, a child who could not be assessed using a specific assessment approach (e.g. could not obtain a basal level on a standardized measure), but could be assessed using another (e.g. a language sample) would not be excluded from analyses. Third, the APPL allows for research and program evaluation studies such as this that amalgamate information from diverse clinical or research sites that used varied measures to explore similar constructs. It provides a framework to ensure that scores from these measures are used to assess relevant language domains in a consistent way (e.g. the use of language to comment and for at least one other function to meet Phase 2 for pragmatics; the attainment of a consistent age equivalent score on standardized measures),
and that all domains of language are considered at all phases when defining overall language level.

This study presents the initial use of the APPL in a community-based context. Thirteen different S-LPs in a province-wide intervention program completed the APPL following a brief orientation to the measure. A limitation of this study is that the APPL guidelines were often, but not consistently, applied. Seventeen percent of natural language samples were shorter than 20 min, and 3% of those included in a quality scan did not involve the recommended range of appropriate communicative contexts. A more detailed orientation to the measure would likely have improved administration quality. In this study, all administrators were S-LPs who had received clinical training in the administration and scoring of natural language samples. Additional training in the use of the APPL may be beneficial in future studies, especially if coders have less experience administering and scoring natural language samples. In this study, natural language samples were consistently administered, and any of a wide range of standardized measures was administered. Future APPL users could increase reliability of categorization beyond the acceptable level reported in this study by adding context-specific assessment requirements beyond those specified by the APPL. For example, users could require that the natural language sample is administered with a specific communication partner, or that specified standardized measures are always administered to measure vocabulary and grammar skills.

To conclude, study authors developed a rating form that promotes consistent application of the language benchmark framework. Consistent characterization of language level is central to understanding language profiles and the impact of specific interventions. We found that four- and five-year-old children with ASD typically have language skills that vary across phonology, vocabulary, grammar, and phonology domains. We also found that participation in a community-based NDBI program was associated with gains in language phase, with gains varying based on initial language level, and varying across language domains.

Acknowledgements

This project would not have been possible without the support of families, staff and leadership from Nova Scotia Hearing and Speech Centres, the IWK Health Centre and the Nova Scotia Health Authority. We are also grateful to staff and volunteers from the IWK’s Autism Research Centre. Thank you to Dr. Joanne Volden for early feedback, and to reviewers for helpful suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Canadian Institutes of Health Research and the Nova Scotia Health Research Foundation.

Supplemental Material

Supplemental material is available for this article online.

ORCID iD

Helen E Flanagan https://orcid.org/0000-0003-2360-749X
Isabel M Smith https://orcid.org/0000-0001-5525-2123

References

Blank, M., Rose, S. A., & Berlin, L. J. (2003). Preschool Language Assessment Instrument–Second Edition (PLAI-2). Austin, TX: PRO-ED.

Bryson, S. E., Koegel, L. K., Koegel, R. L., Openden, D., Smith, I. M., & Nefdt, N. (2007). Large scale dissemination and community implementation of pivotal response treatment: Program description and preliminary data. Research & Practice for Persons with Severe Disabilities, 32(2), 142–153.

Bzoch, K. R., League, R., & Brown, V. L. (2003). Receptive-expressive emergent language test: Examiner’s manual. Austin, TX: Pro-ed.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. American Journal of Speech-Language Pathology, 12(3), 349–358.

Constantino, J. N., & Gruber, C. P. (2012). Social Responsiveness Scale-Second Edition (SRS-2). Torrance, CA: Western Psychological Services.

Costanza-Smith, A. (2010). The clinical utility of language samples. Perspectives on Language Learning and Education, 18, 9–15.

Dawson, J. I., & Tattersall, P. J. (2001). Structured photographic articulation test II. DeKalb, IL: Janelle Publications Inc.

Ellawadi, A. B., & Weism, S. E. (2015). Using spoken language benchmarks to characterize the expressive language skills of young children with Autism Spectrum Disorders. American Journal of Speech-language Pathology, 24(4), 696–707.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). MacArthur-Bates communicative development inventories. Baltimore, MD: Paul H. Brookes Publishing Company.
Frost, L. A., & Bondy, A. S. (1994). PECS: The picture exchange communication system training manual. Cherry Hill, NJ: Pyramid Educational Consultants Newark.

Gillberg, C., & Steffenburg, S. (1987). Outcome and prognostic factors in infantile autism and similar conditions: A population-based study of 46 cases followed through puberty. Journal of Autism and Developmental Disorders, 17(2), 273–287.

Guo, Y. L., & Eisenberg, S. (2015). Sample length affects the reliability of language sample measures in 3-year-olds: Evidence from parent-elicited conversational samples. Language, Speech, and Hearing Services in Schools, 46(2), 141–153.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. Journal of Child Psychology and Psychiatry, 45(2), 212–229.

Hresko, W., Reid, K., & Hamill, D. (1999). Test of Early Language Development 3. Austin, TX: PRO-ED.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. Journal of Child Psychology and Psychiatry, 54(2), 216–224.

Kasari, C., & Smith, T. (2016). Forest for the Trees: Evidence-Based Practices in ASD. Clinical Psychology: Science and Practice, 23(3), 260–264.

Koegel, R. L., Openden, D., Fredeen, R. M., & Koegel, L. K. (2006). The basics of pivotal response treatment. In R. L. Koegel & L. K. Koegel (Eds), Pivotal response treatments for autism: Communication, social, and academic development (pp. 3–30). Baltimore, MD: Brookes Publishing.

Kover, S. T., Davidson, M. M., Sindberg, H. A., & Weismer, S. E. (2014). Use of the ADOS for assessing spontaneous expressive language in young children with ASD: A comparison of sampling contexts. Journal of Speech, Language, and Hearing Research, 57(6), 2221–2233.

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. Biometrics, 33, 159–174.

Lord, C., Rutter, M., Sampson, P. C., & Le Couteur, A. (2003). Autism diagnostic interview – Revised. Los Angeles, CA: Western Psychological Services.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 38(8), 1426–1438.

Mayo, J., Chlebowski, C., Fein, D. A., & Eigsti, I. M. (2013). Age of first words predicts cognitive ability and adaptive skills in children with ASD. Journal of Autism and Developmental Disorders, 43(2), 253–264.

Mirenda, P. (2003). Toward a functional augmentative and alternative communication for students with autism: Manual signs, graphic symbols, and voice output communication aids. Language, Speech, and Hearing Services in Schools, 34(3), 203–216.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). Early social communication scales (ESCS). Coral Gables, FL: University of Miami.

O'Neill, D. (2009). Language Use Inventory: An assessment of young children's pragmatic language development for 18-to 47-month-old children [Manual]. Ontario, Canada: Knowledge in Development.

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 43(2), 418–431.

Rogers, S. (2006). Evidence-based interventions for language development in young children with autism. In T. Charman & W. Stone (Eds), Social & communication development in autism spectrum disorders: Early intervention, diagnosis, & intervention (pp. 143–179). New York, NY: Guilford Press.

Rossetti, L. M. (2006). The Rossetti infant-toddler language scale. East Moline, IL: LinguiSystems.

Rutter, M., LeCouteur, A., & Lord, C. (2003). Autism diagnostic interview – Revised. Los Angeles, CA: Western Psychological Services.

Silver Kite (2016). Touch Chat [Mobile application device]. Retrieved from https://itunes.apple.com/ca/app/speak-for-yourself/id482508198?

Smith, I. M., Flanagan, H. E., Garon, N., & Bryson, S. E. (2015). Effectiveness of community-based early intervention based on pivotal response treatment. Journal of Autism and Developmental Disorders, 45(6), 1858–1872.

Smith, I. M., Flanagan, H. E., Ungar, W. J., D'Entremont, B., Garon, N., den Otter, J., Murray, P. (2019). Comparing the one-year impact of preschool autism intervention programs in two Canadian provinces. Autism Research. doi: 10.1002/AUR.2072.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2005). Vineland-II adaptive behavior scales. Circle Pines, MN: AGS Publishing.

Speak for Yourself (2016). Speak for Yourself [Mobile application software]. Retrieved from https://itunes.apple.com/ca/app/speak-for-yourself/id482508198?

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

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., … Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. Journal of Speech, Language, and Hearing Research, 52(3), 643–652.
van der Meer, L. A. J., & Rispoli, M. (2010). Communication interventions involving speech-generating devices for children with autism: A review of the literature. *Developmental Neurorehabilitation, 13*(4), 294–306.

Wetherby, A. M., & Prizant, B. M. (2003). *CSBS manual: Communication and symbolic behavior scales*. Baltimore, MD: Brookes Publishing.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). *CELF preschool 2: Clinical evaluation of language fundamentals preschool*. San Antonio, TX: Pearson/PsychCorp.

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

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *The Preschool Language Scale – Fifth Edition*. San Antonio, TX: The Psychological Corporation.