JSLHR

Research Article

Narrative Language Markers of Arabic Language Development and Impairment

Natalia V. Rakhlin,[a] Nan Li,[b] Abdullah Aljughaiman,[c] and Elena L. Grigorenko[b,d,e,f,g]

Purpose: We examined indices of narrative microstructure as metrics of language development and impairment in Arabic-speaking children. We examined their age sensitivity, correlations with standardized measures, and ability to differentiate children with average language and language impairment.
Method: We collected story narratives from 177 children (54.2% boys) between 3.08 and 10.92 years old (M = 6.25, SD = 1.67) divided into six age bands. Each child also received standardized measures of spoken language (Receptive and Expressive Vocabulary, Sentence Imitation, and Pseudoword Repetition). Several narrative indices of microstructure were examined in each age band. Children were divided into (suspected) developmental language disorder and typical language groups using the standardized test scores and compared on the narrative indicators. Sensitivity and specificity of the narrative indicators that showed group differences were calculated.
Results: The measures that showed age sensitivity included subject omission error rate, number of object clitics, correct use of subject–verb agreement, and mean length of utterance in words. The developmental language disorder group scored higher on subject omission errors (Cohen's d = 0.55) and lower on correct use of subject–verb agreement (Cohen's d = 0.48) than the typical language group. The threshold for impaired performance with the highest combination of specificity and sensitivity was 35th percentile.
Conclusions: Several indices of narrative microstructure appear to be valid metrics for documenting language development in children acquiring Gulf Arabic. Subject omission errors and correct use of subject–verb agreement differentiate children with typical and atypical levels of language development.

Developmental language disorder (DLD) is a high prevalence disorder, which, according to United States–based studies, affects 5%–12% of preschool children (Law et al., 2000). Despite its high long-term persistence rate (40%–60% of cases) and adverse lifelong consequences (Nelson et al., 2006), only a small percentage of children with delayed language get referred for speech-language therapy (Wittke & Spaulding, 2018). This number is even lower in countries where standardized language assessment methodologies are not developed and there is little awareness of DLD among the general public or even specialists.

Identification of DLD[1] is notoriously challenging. A key source of this challenge is a dearth of knowledge regarding language development and impairment in languages outside monolingual acquisition of mainstream dialects of English and a few other languages. Differences in language structure likely introduce certain differences in the patterns or timing of language acquisition across diverse languages, which makes studies of children from linguistically and culturally diverse backgrounds critically important for understanding both universal aspects of language acquisition and those subject to variation.

[a]Wayne State University, Detroit, MI
[b]University of Houston, TX
[c]King Faisal University, Al Ahsa, Kingdom of Saudi Arabia
[d]Baylor College of Medicine, Houston, TX
[e]Saint Petersburg University, Russia
[f]Moscow State University for Psychology and Education, Russia
[g]Yale University, New Haven, CT

Correspondence to Natalia V. Rakhlin: natalia.rakhlin@wayne.edu

Editor-in-Chief: Stephen M. Camarata
Editor: Krystal L. Werfel

Received February 24, 2020
Revision received April 19, 2020
Accepted July 7, 2020
https://doi.org/10.1044/2020_JSLHR-20-00082

[1]We will use the term *DLD* throughout to refer to a disorder of language acquisition in children observed in the absence of a concomitant genomic, neurological, sensory, or another developmental disorder that can account for language impairments. We will use this term even when referring to studies that used other labels common in research literature (e.g., specific language impairment, language impairment, language disorder).

Disclosure: The authors have declared that no competing interests existed at the time of publication.

This study focuses on Gulf Arabic (GA), a variety of vernacular Arabic spoken on the Arabian Peninsula along the Persian Gulf, specifically in parts of Kuwait, Bahrain, Qatar, United Arab Emirates, Oman, and Saudi Arabia (Simons & Fennig, 2018). The data reported here were collected in the Al-Ahsa-Dammam region, in the Eastern province of Saudi Arabia.

Arabic is a macrolanguage, diverse spoken varieties of which are native for about 313 million people (Simons & Fennig, 2018), including nearly 1 million individuals over 5 years of age in the United States (Ryan, 2013). Despite the high number of speakers, there is limited information on the typical patterns of language development and few resources available to clinicians working with Arabic-speaking children. The current study aims at beginning to address this knowledge gap.

Our goal was to investigate validity and potential diagnostic utility of a set of indices of narrative microstructure, known as ecologically valid measures of language development in English (Botting, 2002; Pavelko & Owens, 2017). To this end, we derived a set of indices of narrative microstructure hypothesized to have clinical promise in Arabic and investigated (a) their construct validity as measures of language acquisition in GA by examining age-related differences in the narrative skills in children of ages 3.08–10.92 years and convergent validity of these measures by examining their associations with standardized language measures known to have psychometric properties; and (b) their diagnostic utility by testing whether these measures differentiate children with typical language (TL) from those with suspected language disorder, as ascertained by standardized measures, and calculating sensitivity and specificity for those measures that show DLD–TL group differences.

Indices of Narrative Microstructure as a Measure of Language Development

Studies with English-speaking children demonstrated that narratives provide a rich source of linguistic data for the study of language growth in children. Various indicators derived from children's narratives reliably differentiate children with language impairment from their typically developing peers (Justice et al., 2006; Kaderavek & Sulzby, 2000; Reilly et al., 2004; Schneider et al., 2006; Scott & Windsor, 2000). Language sample analysis is recognized as a standard part of language assessment for children with suspected or identified language impairment and is of particular importance for children from diverse linguistic and cultural backgrounds for whom valid standardized norm-referenced tests may be lacking (Paul & Norbury, 2012). Language sampling offers some advantages over the use of standardized tests, as an ecologically valid measure, sampling connected speech and providing a rich source for the analysis of multiple aspects of linguistic and communicative competence, which can be tailored to suit assessment needs of an individual child. It can be collected repeatedly, tracking developmental trajectories and selecting short-term, as well as long-term, language therapy goals. A major disadvantage of narrative measures is that high levels of specificity (few false positives) generally come at the cost of low levels of sensitivity (a high rate of false negatives) reported in the literature and vice versa.

Narrative language is typically analyzed along two levels of structure: macro- and microstructure (Justice et al., 2006). Macrostructural analysis evaluates discourse-level organization of stories by rating them on the elements of story grammar and the quality/complexity of episodic structure. Microstructure analysis quantifies children's productive use of various types of lexical, morphosyntactic, and syntactic units. It typically includes general measures of linguistic productivity and complexity, such as mean length of utterance in morphemes or words (MLUm and MLUw, respectively), total number of words (NW), frequency of subordinate clauses, measures of lexical diversity (e.g., number of different words), and measures of accuracy (e.g., number of errors of specific types).

The vast majority of quantitative data on narrative microstructure analysis comes from English, with less research in other languages, including Spanish (Gutiérrez-Clellen et al., 2000; Muñoz et al., 2003), Swedish (Reuterskiöld et al., 2011), Québec French (Thordardottir et al., 2010), and Persian (Foroodi Nejad, 2011). One observation that emerged from the cross-linguistic research concerns the attempts to develop adaptations of the morphemic utterance length measure for languages that are more highly inflected and synthetic than English, which showed to be problematic (Thordardottir & Weismer, 1998). The MLUm measure tends to be inflated in languages in which bare stems are not possible words, errors of omission are not common, and instead errors of substitution are made. In addition, the decision regarding what constitutes a morpheme in child language is not trivial and becomes even more difficult in templatic languages like Hebrew and Arabic, where portmanteau morphemes encode more than one grammatical category and it is difficult to establish what constitutes a base versus a derived form (Dromi & Berman, 1982).

Second, research in English determined that morphosyntactic errors with tense marking (omissions and overregularizations) and omissions of other elements of verb finiteness (e.g., auxiliaries, subject–verb agreement marker) are suited for identifying language impairment (Botting, 2002; Norbury & Bishop, 2003; Reilly et al., 2004). However, because languages vary so widely with respect to inflectional morphology, omission of the elements marking verb finiteness was shown not to be a universal linguistic marker of DLD (Leonard, 2017).

DLD Across Languages

A number of proposals for what constitutes the core underlying deficit in DLD were universalist, such as agreement deficit (Clahsen et al., 1997), extended optional infinitive phase (Wexler, 1998), and representational deficit for dependent relations (Van Der Lely & Battell, 2003).

However, research found that presentation of DLD in different languages varies based on the characteristics of the language. Thus, a deficit using tense and/or subject–verb agreement morphology was documented as a linguistic marker of DLD in a number of languages, including English, German, Dutch, and Greek (Clahsen & Dalalakis, 1999; Conti-Ramsden et al., 2001; Hamann, 2015; Rice & Wexler, 1996; Verhoeven et al., 2011). However, this deficit was not found to be as pronounced in French, Spanish, and Italian (Bedore & Leonard, 2001; Bortolini et al., 1997; Thordardottir & Namazi, 2007). Such findings gave rise to relativist proposals, which suggested that typological properties of the child's language influence clinical presentation of DLD, that is, what features and to what degree tend to be impaired (Leonard, 2017).

It was proposed that, in languages with sparse or inconsistent inflectional paradigms, where overtly inflected forms coexist with bare stems (i.e., English), a frequent error type is omission of the inflectional elements, or in languages that overtly mark infinitival forms (e.g., German or Dutch), a substitution of a finite form with an infinitival form (and word order errors involving verb placement, which is related to verb finiteness). On the other hand, in languages with uniformly inflected verbs and regular and transparent inflectional paradigms (like Italian and Spanish), difficulties with verb morphology for children with DLD is alleviated (Leonard et al., 1992). Instead, DLD in these languages manifests itself in underuse of object clitics (i.e., unstressed pronominal forms that involve a noncanonical word order). These must be placed before the verb, rather than after the verb (e.g., Anna la (her) vede (sees), "Anna sees her" vs. Anna vede la ragazza, "Anna sees the girl"). Children with DLD avoid their use, replacing them with full noun phrases (Arosio et al., 2014; Bedore & Leonard, 2001; Bortolini et al., 2006, 1997; Paradis et al., 2003).

Morphologically isolating languages, such as Cantonese or Mandarin, have few elements of any kind that can be considered inflectional. One type of functional morphemes these languages have is aspectual markers, which are optional, as the verbs do not require them to be grammatical (e.g., tā chī yú ("he eats fish") versus tā zài chī yú ("he is eating fish"). DLD in these languages is characterized by underuse of aspectual markers (Fletcher et al., 2005).

In morphologically complex agglutinating languages, such as Hungarian and Finnish, where verbs (and nouns) are consistently inflected with a sequence of affixes, children with DLD do not make errors of omission or morpheme ordering. Instead, they make substitution errors replacing more complex forms with a simpler morphological default form (Kunnari et al., 2011; Leonard et al., 2009); for example, past-tense third-person plural with third-person singular.

In Hebrew, a language typologically similar to Arabic, a somewhat different phenomenon was observed. Hebrew verbs consist of consonantal roots, expressing the core meaning of the word, interleaved with vowel "patterns" or templates, referred to as "binyanim," which augment the roots' meaning. Consonantal roots are unpronounceable in isolation from the pattern in which they occur. There exist multiple verb patterns. In some patterns, the contrast between present and past tense involves a change in the vowel sequence; in others, it involves a syllable deletion or addition of a syllabic prefix, along with a change in the interconsonantal vowels. Research found that tense and agreement errors in Hebrew-speaking children with DLD have a rather circumscribed character (Dromi et al., 1999; Leonard et al., 2000). Children's error rate with verbs was higher than that of their typically developing peers only with verbs of one specific binyan, suggesting incomplete morphological learning of verbs of that class, rather than a more generalized weakness with agreement or tense.

Because language production markers of DLD vary across languages, language processing measures, namely, nonword repetition or sentence imitation, were proposed as universalist unbiased probes for language impairment. Such measures were shown to have good sensitivity and specificity in English (Bishop et al., 1996; Conti-Ramsden, 2003; Conti-Ramsden et al., 2001) and differentiate typically developing children from children with DLD in other languages, including Arabic (Saiegh-Haddad & Ghawi-Dakwar, 2017, Italian, Spanish, and French (Bortolini et al., 2006; Girbau & Schwartz, 2007; Thordardottir et al., 2011). Although nonword repetition failed to discriminate children with DLD from their typically developing children in some languages, such as in Cantonese and Czech, sentence imitation remained a sensitive measure in these languages (Smolík & Vávrů, 2014; Stokes et al., 2006). This suggests that a combination of language-specific production measures with universally applicable processing measures should be used for identification of DLD.

Challenges in Identification of DLD

Because of considerable heterogeneity of DLD and lack of specific neurobiological markers or known causes, its identification remains challenging even in languages where this phenomenon has received a lot of attention. The difficulty lies in determining what indicators are best at differentiating affected from unaffected children and what degree of language impairment is necessary for a diagnosis of mild-to-moderate forms of DLD. A relatively low cutoff score (i.e., 1.5 or 2.0 *SD*s below the population mean) is often used to designate the line between normal and impaired level of language. However, a review of standardized tests (Spaulding et al., 2006) showed that, instead of representing the low end of the normal distribution, children with DLD represent a downward-shifted distribution, with the majority scoring above the cutoff of −1.5 *SD* below the mean on most of the reviewed tests, with 27% of the tests yielding scores of within 1 *SD* of the mean. This revelation is at odds with the standard practice of expecting a child to score well below 1 *SD* of the mean for a diagnosis of DLD.

Another important observation concerns variability of diagnostic performance among existing tests and naturalistic measures. In a systematic review of diagnostic procedures for preschool children (Shahmahmood et al., 2016), tests and measures ranged in their sensitivity from 16% to 100% and specificity from 14% to 100%. Sensitivity and specificity of a test (i.e., its ability to correctly designate affected and unaffected individuals as such) are dependent on the cutoff score chosen as the threshold between unimpaired and impaired performance. Choosing a lower cutoff tends to improve specificity but decrease sensitivity, and vice versa. A cutoff score that allows an optimal combination of sensitivity and specificity for one test can differ significantly from that of another test, even when these tests were validated on the same sample of children. The same measure would perform differently when applied in linguistically diverse populations. Therefore, the right combinations of tools and cutoffs for clinical use must be derived for each language.

Language Acquisition and Language Disorder of Arabic

Speech-language pathology is an emerging field in the Arabic-speaking world. It has been reported that only 42 speech-language pathologists work in government and private sector in the entire country of Kuwait (Al-Khaledi et al., 2009) and there are fewer than five speech-language pathologists per million population in Egypt (Kotby et al., 2010). For Saudi Arabia, the country in which our research was conducted, only one speech-and-hearing clinical organization is listed in an online directory of speech-language pathology institutions in the Middle East and North Africa (speech-language-therapy.com). Addressing this acute shortage of speech-language services requires systematic studies of language development in Arabic-speaking children.

There are few published studies of child language acquisition of Arabic. In a study of the acquisition of Egyptian Arabic (Omar, 1973), children were reported to acquire most aspects of grammar at ages in line with other languages (between 2 and 5 years old). A longitudinal study of language growth in six Emirati Arabic–speaking children used indices of linguistic microstructure (e.g., MLUm and MLUw, number of utterances (NU) per turn, number of different words, and type–token ratio) showing age-related changes for most measures (Ntelitheos & Idrissi, 2017). Some studies focused on specific areas of syntax and morphosyntax, namely, comprehension of various word orders in Kuwaiti Arabic, noun plurals in Kuwaiti and Palestinian Arabic (Abdalla et al., 2013; Ravid & Farah, 1999; Saiegh-Haddad et al., 2012), and tense and subject–verb agreement (Abdalla & Crago, 2008; Aljenaie, 2010; Qasem & Sircar, 2017).

A study of 2.00- to 5.17-year-old children acquiring Urban Hijazi Arabic (Abdalla & Crago, 2008) found a high accuracy rate (over 94% correct) of tense and subject–verb agreement marking in present and past tense for even the youngest group of children with TL development. Children with DLD underperformed relative to both age-matched and MLU-matched TL peers, with only 68% correct on tense and 77.7% correct on subject–verb agreement (Abdalla & Crago, 2008). Errors primarily occurred with third person and feminine agreement, with the second-person masculine imperative used as a default form in substitutions by both children with TL and DLD. High accuracy with verb finiteness by children with TL was confirmed in a small study of Kuwaiti Arabic, in which children 3.0 years old and younger showed accuracy of 90% with verb tense and agreement marking (Aljenaie, 2010) and used bare imperfective masculine stems as default.

Finally, a small study of the acquisition of Egyptian Arabic (Fahim, 2017) reported substantial variability in children's use of verbal morphology, measured as either mean number of morphemes per utterance or as percentage of verb error patterns. Among the six typically developing participants, ages 2.03–4.06 years, morphemes per utterance ranged from 2.1 to 7.5, and percentage of verb error patterns ranged from 0% to 57%. The study reported both errors of omission and substitution. The key finding was that children tended to use the correct form of the verb stem (root and pattern), but either omitted or substituted the affixes or clitics.

Based on the available evidence, we expect the narrative measures used in DLD evaluation in English to behave somewhat differently in Arabic. First, we may expect analyses based on calculating morpheme density to be problematic (due to complexity in applying morpheme analyses in a root-and-pattern-based morphological system). Second, given low error rates with verb morphology reported in Arabic children, measures based on the accuracy of tense and agreement marking may require higher cutoffs for identifying DLD. Furthermore, additional measures reflecting Arabic grammar may serve as linguistic markers of DLD.

Arabic is characterized by a considerable variation among its spoken varieties. Ethnologue (Simons & Fennig, 2018) lists over 20 varieties of Arabic used across the Middle East and North Africa, in addition to Standard Arabic, a lingua franca used as the language of education, media, government, and religion throughout the Arab world. A systematic parallel use of two distinct language varieties (the vernacular and the standard), each for a different purpose, known as *diglossia* (Ferguson, 1959), likely impacts language acquisition. Establishing dialect-neutral language measures that can be applied to children learning a variety of Arabic dialects in the context of diglossia is thus needed.

Brief Sketch of Arabic Grammar

Arabic word structure, similar to Hebrew binyanim, is analyzed as an abstract template, which includes the slots for the root consonants and the vocalic patterns (Habash, 2007). The consonantal roots encode the base meaning, while the vowel pattern adds to the meaning (Holes, 2004). The information contributed by the pattern (which, in addition to vowels, may contain consonants, particularly in verb patterns, word initially) is specific and predictable, including syntactic properties (part of speech: noun, adjective, etc.) and semantic properties (in the case of verbs, valence, transitivity, etc.; for more, see Saiegh-Haddad & Henkin-Roitfarb, 2014.)

For example, the triconsonantal root "ktb" with the base meaning "write" is the basis for the formation for such words as *kitaab* ("book"), *maktaba* ("library"), *maktuub* ("letter"), *kitaaba* ("writing"), and *kataba* ("to write"). GA verbs are morphologically marked for tense (present and past) and subject–verb agreement (for person, number, and gender). Table 1 illustrates a verb paradigm for the verb "write."

In summary, Arabic word structure is characterized by a high degree of abstractness, nonlinearity, and discontinuity, as roots and vocalisms do not constitute continuous separable phonetic entities and are instead interwoven with each other onto word templates. To be acquired, they must be inferred from underlying distributional patterns, relying on skills related to grammar induction (procedural rather than declarative memory involved in rote learning), to a much greater degree than word learning requires for languages such as English (Boudelaa et al., 2010; Ullman, 2016).

Like many other languages with rich agreement, Arabic permits pronominal subjects to be dropped (Holes, 2013; Kenstowicz, 1989). In fact, full forms of subject pronouns (listed in Table 1) are only used for emphasis and judged by native speakers as unnatural in neutral, nonemphatic contexts. In addition to subject drop, pronominal objects in GA are expressed as clitics, weak pronominal forms suffixed to the verb, rather than full pronouns. Thus, a sentence equivalent to English "I saw her" would consist of the verb "see" in the form "past first-person singular" with the third-person singular feminine object clitic affixed to it.

The use of null subjects and objects is pragmatically regulated. All pronouns (both overt or null) must have a referent established and unambiguously identifiable in the preceding discourse, that is to be referring to a discourse entity not only previously introduced but salient (i.e., at the center of attention of the hearer) at the time of the utterance containing the pronoun. A new entity introduced into discourse on its first mention must be realized with a full indefinite noun phrase (NP), and on subsequent mentions, it must be realized with a pronoun or a specific noun phrase. The latter is needed if other salient discourse entities make reference ambiguous. This is illustrated in the mini-discourse in (1) below, where the sentence in (c) is pragmatically infelicitous because the referents of the null subject and object cannot be identified unambiguously, as all three discourse entities previously introduced are masculine singular (i.e., the boy, the older brother, and the camel).

(1) a. ʔl-walad kan ʕəndah ħaflat ʕid.milad.
The-boy have.3.SG.M.Past party birthday
'The boy had a birthday.'
b. ʔxuh al-kəbir ʕəṭ-ah ʒamal kə-hadijah l-ʕid.milad-ah.
Brother.his the-old give.3.SG.M.Past camel as-present for-birthday-his
'His older brother gave him a camel as a present for his birthday.'
c. # ʔaxað-ah bə-kil mukan raħ-ah.
take.3.SG.M.Past-him in-all place go.3.SG.M.Past
'#He took him everywhere he went.'

In short, correct use of NS requires both the knowledge of the grammatical constraints governing their use (mastering the verb agreement system) and the pragmatic rules of structuring discourse to accommodate the hearer's information state (along with the capacity to maintain accurate representation of other minds). Grammatical knowledge is expected to develop early (during preschool age) and reach high levels of accuracy before the age of 5 years. On the other hand, pragmatic knowledge and social cognitive skills required for its use develop in a more protracted way, as children learn how to estimate, establish, and maintain common ground with their interlocutors past the age of 6–7 years.

The diglossic nature of the language learning environment in Arabic-speaking countries and the distinctive typological properties of Arabic require research that establishes benchmarks in the typical trajectory for acquiring this complex system in order to understand how this complexity affects language acquisition under conditions of impairment. Currently, such research on the acquisition of Arabic is insufficient (Mahfoudhi & Abdalla, 2017).

Table 1. Subject–verb agreement and pronouns in Gulf Arabic (based on Holes, 2013).

Person	Number, gender	Subject pronouns	Object clitics	Verb paradigm ("write," k-t-b)	
				Present	Past
1st	Singular	ána	-ni	á**ktib**	**kítab**-t
	Plural	iħna	-na	ni-**ktib**	**kítab**-na
2nd	Singular masculine	ínta	-ik/-ak	ti-**ktib**	**kítab**-t
	Singular feminine	ínti	-ich/-ach	ti-**kítb-**i	**kítab**-ti
	Plural masculine	íntu	-kum	ti-**kitb-**u	**kítab**-taw
	Plural feminine	íntu	-kum	ti-**kitb-**án	**kítab**-taw
3rd	Singular masculine	húwwa	-ah	yí**ktib**	**kítab**
	Singular feminine	híyya	-ha	ti-**ktib**	**kítb**-at
	Plural masculine	húmma	-hum	yi**kítb**u	**kítb**aw
	Plural feminine	húmma	-hum	yi**kitb**-an	**kítb**aw

The Current Study

The goal of the study was to investigate the properties of indices of narrative microstructure as metrics of language development in children acquiring GA and markers differentiating typical from impaired language in children between the ages of 3.08 and 10.92 years. Our principal aim was to derive a set of naturalistic indicators that can be used for tracking language development and as a clinical tool (i.e., for DLD screening, selection of measurable intervention goals, and assessing a child's progress throughout language intervention). The clinical and developmental subaims were closely related to each other, in accordance with the descriptive developmental approach to DLD we adopted as a best-evidence approach (Paul & Norbury, 2012). According to this approach, DLD is viewed in continuity with TL development, and detailed linguistic description is seen as the key to mapping out a successful intervention strategy for a child with DLD. Furthermore, the sequence of language skills in normal language development informs the selection of treatment goals for children with DLD. Thus, establishing a set of indices of narrative microstructure that can reliably track developmental change in typical development is directly relevant for clinical judgments involved in identifying and treating children with delayed (or disordered) language development.

We pursued this dual goal by investigating a set of narrative indices for their (a) correlations with standardized language measures of GA known to have psychometric properties, (b) age sensitivity, (c) differences between children with average language and children with suspected language disorder, and (d) diagnostic accuracy in predicting TL and DLD status.

Method

Participants

The data were collected from a community sample of 177 children (54.2% boys) between the ages 3.08 and 10.92 years (M = 6.25, SD = 1.67). Assessment took place in schools and preschools located in the Al-Ahsa-Dammam area of Saudi Arabia after informed consents were collected from the parents of the participants. Each narrative was recorded and transcribed by a native speaker of Arabic from Saudi Arabia with an MA degree in Linguistics. In addition, about 14% of the narratives were transcribed by a second Saudi MA-level linguist to demonstrate interrater agreement, which was 96.7%.

For the purposes of the analyses, the children were divided into six age bands of 12 months each: 3.00–3.99, 4.00–4.99, 5.00–5.99, 6.00–6.99, 7.00–7.99, and 8.00–10.92 years. Table 2 reports the demographic characteristics of the sample and the descriptive statistics.

Procedure and Measures

Each child was evaluated individually in a quiet room by a teacher, professionally trained in narrative elicitation by the members of the research team. Because of the cultural features of the Saudi Arabia society, teams of male and female assessors had to be recruited to collect data from male and female participants, respectively. All teachers received a 2-day training on the basics of psychological and language assessment procedures in general and the details of the Arabic Language: Evaluation of Function (ALEF; Rakhlin et al., 2019) administration in particular. The training was conducted remotely by the U.S. members of the research team who had developed the ALEF (which included the first author). Subsequently, all teachers were continuously monitored by the Saudi member of the research team (the third author).

Language Sampling

Language sampling task was story generation. For the purpose of the assessment, we developed a narrative instrument based on an original wordless picture-story (analogous to "Frog, Where Are You?" by M. Meyer). Each child was asked to look through the pictures and generate a story to go with the pictures. The narrative was recorded and transcribed using the International Phonetic Alphabet, followed by morpheme-by-morpheme glossing and English translation. Each utterance was coded as either verbal (containing a verb) or not. All verb forms were coded as correct or incorrect. Object clitics were coded as correct or incorrect in all transitive utterances in which the verb was marked with an object clitic. In all verbal utterances, sentence subjects were coded as overt (full pronoun or a full NP) or null. Null subjects were coded as felicitous or not based on the First Mentions Protocol from the Edmonton Narrative Norms Instrument (Schneider et al., 2006).

Following coding, each transcription was manually analyzed to derive a set of indices of narrative microstructure. Measures of linguistic productivity included Number of Words (NW), Number of Utterances (NU), and Number of Verbal Utterances (NVU; a measure that excludes utterances consisting of noun phrases or other labeling phrases). Measures of linguistic complexity included Mean Length of Utterances in Words (MLUw), Overt Subjects (OS; derived by calculating the number of verbal utterances with an overt pronoun or a full NP subject); Null Subjects (NS; derived by calculating the number of verbal utterances without an overt subject); and Number of Object Clitics (NOC; derived by calculating the total number of verbal utterances, in which the verb had an object clitic). Measures of grammatical accuracy included Subject-verb Agreement Errors (SAE, derived by calculating the number of verbs with incorrect forms of person-gender-number agreement); Object Clitic Errors (OCE; derived by calculating the number of transitive verbs with incorrect forms of the object clitic); Total Agreement Errors (TAE, a sum of SAE and OCE), as well as Correct Use of Subject-verb Agreement (CUSA; derived by calculating the number of correct verb forms). Finally, we included a measure of pragmatic felicity, Subject Omission Errors (SOE; the number of missing (null) subjects judged pragmatically infelicitous or ambiguous).

Table 2. Means (standard deviations) and range of scores on standardized tests (n = 177).

Age	n	Age range	% Boys	EV	RV	SI	PR
3 years	24	3.08–3.99	71%	49.08 (14.23)	24.17 (4.18)	24.5 (19.69)	17.88 (7.94)
Range				21–76	19–32	0–65	5–31
4 years	24	4.00–4.99	67%	54.21 (13.19)	26.46 (5.00)	36.96 (18.51)	21.83 (6.21)
Range				19–72	16–35	4–83	6–31
5 years	27	5.00–5.99	48%	66.33 (11.33)	28.56 (4.80)	51.04 (16.88)	24.00 (6.48)
Range				36–83	16–40	19–79	7–32
6 years	35	6.00–6.99	37%	68.69 (8.48)	30.34 (4.66)	55.66 (15.86)	23.26 (6.27)
Range				41–83	14–39	11–83	8–33
7 years	35	7.00–7.99	57%	67.37 (7.62)	31.6 (3.95)	57.11 (16.43)	25.86 (4.74)
Range				44–77	24–40	2–83	13–33
8 years	32	8.00–10.92	53%	67.47 (10.16)	30.88 (4.74)	54.47 (22.14)	26.25 (4.77)
Range				38–83	20–39	1–83	15–33

Note. The values are raw scores. EV = Expressive Vocabulary; RV = Receptive Vocabulary; SI = Sentence Imitation; PR = Pseudoword Repetition.

Because sentences with null subjects may result in ambiguous reference, making it impossible to determine the target agreement and its accuracy, the measures SAE, TAE, and CUSA were based only on the utterances with overt subjects, including demonstratives, and utterances with null subjects with unambiguous reference.

Standardized Language Measures

We used the following four subtests from the ALEF battery (Rakhlin et al., 2019), a standardized test of Arabic language development:

Receptive Vocabulary. Receptive Vocabulary was a picture-pointing task of 53 items (α = .90) with three alternatives to choose from. This subtest assessed the receptive knowledge of concrete and conceptual vocabulary using developmentally appropriate lexical items sampled from different grammatical categories (nouns, adjectives, verbs, adverbs, and prepositions) and semantic classes (animals, transportation, home/household items, body parts, nature, food, clothing, materials, emotions, actions, colors, and shapes). The verbs were sampled from verb classes I–X. Items were selected according to a three-tier taxonomy: (a) basic, high-frequency vocabulary learned in everyday social interactions; (b) high- and moderate-frequency words learned through educational experiences used across domains by more mature speakers; and (c) more advanced and context-specific words learned from exposure to schooling used in specific domains.

Expressive Vocabulary. Expressive Vocabulary was a picture-naming task with 53 items (α = .89), where a child was shown a series of pictures. Similar principles of item selection were used for this subtest as the Receptive Vocabulary.

Sentence Imitation. Research in typical language acquisition has established the important role of sentence imitation as a probe into the child's internalized grammatical system and an effective tool for identifying children with DLD (Conti-Ramsden et al., 2001). Elicited sentence imitation assesses children's knowledge of various grammatical elements based on the accuracy of the imitation. The assumption is that for the child to successfully imitate a sentence, its structure must be part of the child's grammar (Lust et al., 1987). Our test contained 35 sentences of increasing complexity (α = .97). The items targeted such structures as *wh*-questions of varying complexity, sentences with negation, sentences containing complex predicates and conjoined clauses, and sentences with subordinate clauses of various types and complexity. A child is asked to repeat them exactly as they are spoken by the adult, and the score for each item is assigned based on the number and type of errors (i.e., omission, substitution, permutation).

Pseudoword Repetition. Pseudoword repetition tasks, assumed to measure phonological working memory, have been previously identified as a robust clinical marker for DLD. In the task used for the study, each child repeated pseudowords of increasing length and complexity after the evaluator. The score reflected the number of correctly repeated items. There were 34 items (α = .90).

DLD identification procedure. Because there are no country-wide speech-language pathology services in Saudi Arabian schools, none of the children in our sample had an established diagnosis of DLD. To investigate whether indices of narrative microstructure differentiate children with TL from children with DLD, we used children's performance on the four subtests from the ALEF battery described above. We classified children into two groups, (suspected) DLD and TL, using the following procedure:

1. DLD: any child who had at least two standardized scores below the 16th percentile (n = 27);

2. TL: the remaining children, namely, the children who had at least three standardized scores above the 16th percentile (n = 150).

Results

Are Narrative Indicators Valid Measures of Language Development in GA?

Tables 2 and 3 report the descriptive statistics for the standardized and narrative measures (respectively) for each age band. We investigated whether the narrative indicators show convergence with existing standardized measures of language development and exhibit age-related sensitivity in GA. To this end, we first conducted a correlational analysis. Table 4 shows the Pearson intercorrelations between the narrative indicators, standardized measures, and age.

With respect to the relationship between the narrative and standardized measures, we found no correlations for the general measures of productivity (number of words, utterances, and verbal utterances). However, a number of correlations for the measures of complexity, wellformedeness, and pragmatic felicity were significant. Thus, we found significant inverse correlations between several standardized measures (including Expressive Vocabulary, Sentence Imitation, and Pseudoword Repetition) with the measures of subject omission errors and subject–verb agreement errors. We also found significant positive correlations between standardized measures and the narrative measures of overt subjects, number of object clitics, object clitic errors, correct use of subject–verb agreement, and MLUw (see Table 4).

With respect to correlations with age, we again found no significant correlations between age and measures of productivity (number of words, utterances, and verbal utterances). The measures that were significantly correlated with age (in addition to all four standardized measures) were the following: MLUw (r = .25), number of object clitics (r = .26), and correct use of subject–verb agreement (r = .28). There were no other significant correlations.

We followed up with a set of multiple linear regression analyses, in which each narrative indicator was regressed on gender, age, and then the interaction between gender and age was added. Age was centered at the sample mean (M_{age} = 6.25 years). We found a number of dependent variables with significant amount of variance explained by the predictors. The results of these analyses are given in Table 5.

First, we will discuss the results pertaining to the indicators of null subject use, which did not show bivariate correlation with age in our previous analyses and for which the three-factor regression model was not significant. We found the two null subject measures, null subjects and subject omission errors, were strongly correlated with each other (r = .79), suggesting a possibility that a major factor driving the number of subject omission errors was the number of null subjects, which, in turn, strongly correlated with the number of utterances (r = .72). Thus, the contribution of narrative length via the number of null subjects likely obscured the contribution of age to the frequency of null subject errors and needed to be controlled for. To control for the number of null subjects, we created a new dependent variable: a ratio of subject omission errors to the total number of null subjects, SOE/NS_Ratio, and included it in our multiple linear regression analysis.

The model for the SOE/NS_Ratio was significant, $F(3, 172) = 6.37$, $p < .01$, $R^2 = .08$. The interaction between age and gender was a significant predictor, indicating that the association between age and the SOE/NS_Ratio was moderated by gender. We proceeded with investigating the simple effect of age on the SOE/NS_Ratio for boys and girls separately. For boys, age was negatively related to the SOE/NS_Ratio ($b = -0.07$, $p < .01$), suggesting that a 1-year increase in age for a boy is expected to result in a 0.07-unit decrease in the SOE/NS_Ratio. However, for girls, age was not related to the SOE/NS_Ratio ($b = -0.01$, $p = .57$).

Table 3. Means and standard deviations for narrative measures (n = 177).

Age (years)	NU	NVU	NW	SOE	OS	NS	NOC	OCE	SAE	TAE	CUSA	MLUw
3	34.92	28.08	133.88	16.04	11.38	20.25	5.42	0.38	1.21	1.58	11.08	3.70
	(10.30)	(13.55)	(67.11)	(9.84)	(8.66)	(11.15)	(4.27)	(1.06)	(1.47)	(2.06)	(9.26)	(1.17)
4	36.33	32.75	131.96	18.88	11.21	23.38	6.79	0.54	0.58	1.13	12.96	3.43
	(15.39)	(17.43)	(76.44)	(13.79)	(9.98)	(13.51)	(5.91)	(0.88)	(0.78)	(1.30)	(10.23)	(1.16)
5	35.89	32.63	147.78	14.30	14.26	21.56	9.30	0.67	0.63	1.30	18.85	4.04
	(14.23)	(14.31)	(79.76)	(9.67)	(10.10)	(10.38)	(6.52)	(1.00)	(0.93)	(1.56)	(14.45)	(1.05)
6	33.20	30.00	133.63	11.83	14.43	18.23	8.14	0.54	0.74	1.29	18.00	3.99
	(9.96)	(9.16)	(55.56)	(8.56)	(10.27)	(7.59)	(4.40)	(1.09)	(1.07)	(1.79)	(10.88)	(1.03)
7	34.89	32.71	149.03	12.51	14.43	19.48	9.11	0.89	1.00	1.89	19.43	4.16
	(9.04)	(9.14)	(61.90)	(8.21)	(9.01)	(8.94)	(6.07)	(1.89)	(1.73)	(2.77)	(9.65)	(0.89)
8+	36.84	35.50	163.94	14.16	12.03	25.34	10.78	0.56	0.28	0.84	21.72	4.38
	(7.48)	(7.15)	(56.30)	(10.85)	(10.24)	(8.39)	(3.87)	(1.11)	(0.68)	(1.32)	(10.54)	(0.83)

Note. NU = number of utterances; NVU = number of verbal utterances; NW = number of words; SOE = subject omission error; OS = overt subjects; NS = null subjects; NOC = number of object clitics; OCE = object clitic errors; SAE = subject–verb agreement errors; TAE = total agreement errors; CUSA = correct use of subject–verb agreement; MLUw = mean length of utterance in words.

Table 4. Pearson correlations between narratives and standardized tests and age (n = 177).

Measure	NU	NVU	NW	SOE	OS	NS	NOC	OCE	SAE	TAE	CUSA	MLUw	Age
EV	−.10	.03	.09	−.33**	.13	−.10	.16*	.02	−.13	−.07	.36**	.26**	.48**
RV	−.04	.07	.07	−.13	.07	−.03	.14	.17*	−.05	.08	.19*	.16*	.45**
SI	−.08	.03	.03	−.33**	.16*	−.13	.19*	.04	−.17*	−.09	.35**	.12	.45**
PR	.06	.15	.14	−.17*	.11	.04	.26**	.04	−.14	−.06	.32**	.14	.39**
Age	.01	.13	.14	−.13	.05	.06	.26**	.06	−.12	−.04	.28**	.25**	

Note. NU = number of utterances; NVU = number of verbal utterances; NW = number of words; SOE = subject omission error; OS = overt subjects; NS = null subjects; NOC = number of object clitics; OCE = object clitic errors; SAE = subject–verb agreement error; TAE = total agreement errors; CUSA = correct use of subject–verb agreement; MLUw = mean length of utterance in words; EV = Expressive Vocabulary; RV = Receptive Vocabulary; SI = Sentence Imitation; PR = Pseudoword Repetition.

*p < .05. **p < .01.

Second, the model statistically significantly accounted for 10% of the variance of the number of object clitics, $F(2, 174) = 11.11$, $p < .01$. Age was positively related to the number of object clitics controlling for gender, suggesting that, for boys, a 1-year increase in age was related to a 0.78-unit increase in the number of object clitics. Gender was positively related to the number of object clitics. controlling for age, suggesting that girls who were 6.25 years old scored 2.32 units higher on the number of object clitics than boys. However, the interaction between age and gender was not statistically significant for number of object clitics.

Next, the model also statistically significantly predicted correct use of subject–verb agreement, $F(3, 173) = 8.74$, $p < .001$, $R^2 = .12$. The interaction between age and gender was a significant predictor of correct use of subject–verb agreement, suggesting that the influence of age on correct use of subject–verb agreement depended on child

Table 5. Estimated regression model parameters for narrative indices (n = 177).

		Age		Gender		Age × Gender		Overall model		
Measure	Intercept	b	β	b	β	b	β	Adjusted R^2	F	df
NU	34.64	0.00	.00	1.37	.06			.00	0.34	2
	34.73	0.57	.09	1.43	.07	−1.46	−.14	.00	0.90	3
NVU	31.37	0.88	.12	1.48	.06			.01	1.86	2
	31.51	1.74*	.24	1.58	.07	−2.17*	−.19	.03	2.59	3
NW	143.15	5.27	.13	2.18	.02			.01	1.66	2
	143.93	10.09**	.26	2.75	.02	−12.24*	−.20	.03	2.47	3
SOE	13.85	−0.84	−.14	0.94	.05			.01	1.76	2
	13.79	−1.25*	−.20	0.90	.04	1.03	.11	.01	1.57	3
NS	20.71	0.32	.05	1.16	.06			.00	0.60	2
	20.76	0.62	.10	1.19	.06	−0.75	−.08	.00	0.61	3
SOE/NS_Ratio	0.64	−0.05***	−.27	0.03	.04			.06	6.78***	2
	0.63	−0.07***	−.41	0.02	.04	0.06*	.21	.08	6.37***	3
OS	13.12	0.28	.05	0.00	.00			.00	0.20	2
	13.16	0.55	.10	0.04	.00	−0.70	−.80	.00	0.33	3
NOC	7.38	0.78***	.24	2.32**	.21			.10	11.11***	2
	7.42	1.01***	.31	2.35**	.22	−0.61	−.12	.11	7.97***	3
OCE	0.44	0.03	.04	0.37	.15			.01	2.19	2
	0.43	−0.02	−.03	0.36	.14	0.13	.11	.02	1.91	3
SAE	0.78	−0.08	−.11	−0.10	−.04			.00	1.32	2
	0.79	−0.03	−.04	−0.09	−.04	−0.14	−.12	.01	1.38	3
TAE	1.22	−0.05	−.04	0.27	.07			.00	0.54	2
	1.22	−0.05	−.04	0.27	.07	−0.01	.00	.00	0.36	3
CUSA	17.21	1.90***	.28	0.56	.03			.07	7.58***	2
	17.42	3.15***	.46	0.71	.03	−3.17**	−.29	.12	8.74***	3
MLUw	4.00	0.15***	.25	−0.04	−.02			.05	5.59**	2
	4.02	0.22***	.36	v0.03	−.01	−0.18	−.18	.06	5.01**	3

Note. NU = number of utterances; NVU = number of verbal utterances; NW = number of words; SOE = subject omission error; NS = null subjects; SOE/NS_Ratio = subject omission errors divided by null subjects; OS = overt subjects; NOC = number of object clitics; OCE = object clitic errors; SAE = subject–verb agreement error; TAE = total agreement errors; CUSA = correct use of subject–verb agreement; MLUw = mean length of utterance in words.

*p < .05. **p < .01. ***p < .001.

gender. For boys, age was positively related to correct use of subject–verb agreement ($b = 3.15$, $p < .05$), suggesting that a 1-year increase in age for a boy is expected to result in 3.15 units of increase in correct use of subject–verb agreement. However, age was not related to correct use of subject–verb agreement for girls ($b = -0.02$, $p = .98$).

Finally, the model was significant for MLUw, $F(2, 174) = 5.59$, $p < .01$, $R^2 = .05$. Age was positively related to MLUw controlling for gender, whereas gender and the interaction between age and gender were not related to MLUw.

The Narrative Performance in Children With (Suspected) DLD

As described in the Method section, we used 16th percentile as the threshold between DLD and TL. We found that this was a clean cutoff, with few children ranking in the 15th and 17th percentiles. The comparison between the two groups showed that they did not significantly vary on gender composition ($p = .51$, Fisher's exact test). A two-group between-subjects multivariate analysis of variance was carried out using language status (TL or DLD) and age (i.e., covariate) as independent variables and 12 narrative measures as dependent variables (see Table 6 for these results). The Pillai's trace, $F(11, 164) = 2.56$, $p < .01$, partial $\eta^2 = .15$, suggests that the composite dependent variable was significantly influenced by whether a child belonged to the TL or DLD group. To follow up on the significant multivariate effect, we proceeded with univariate analyses of variance on each dependent variable separately. The two groups significantly differed on two measures: subject omission errors and correct use of subject–verb agreement (Cohen's $d = 0.55$ and 0.48, respectively).

Table 6. Comparing group performance of children with (suspected) developmental language disorder (DLD) and those with typical language (TL).

Variables	DLD (n = 27)	TL (n = 150)	ANOVA	Cohen's d
NU	38.33 (13.20)	34.71 (10.52)	$F(1, 174) = 2.49$ $p = .12$	0.30
NVU	33.41 (12.97)	31.8 (11.68)	$F(1, 174) = 0.27$ $p = .60$	0.13
NW	143.22 (68.25)	144.28 (65.51)	$F(1, 174) = 0.05$ $p = .83$	0.02
SOE	19.48 (12.51)	13.35 (9.49)	$F(1, 174) = 9.63$ $p = .00$	0.55
OS	11.85 (8.76)	13.35 (9.89)	$F(1, 174) = 0.61$ $p = .43$	0.16
NS	23.67 (12.15)	20.80 (9.62)	$F(1, 174) = 1.71$ $p = .19$	0.26
SAE	1.04 (1.61)	0.68 (1.12)	$F(1, 174) = 2.39$ $p = .12$	0.26
OCE	0.59 (1.12)	0.61 (1.27)	$F(1, 174) = 0.02$ $p = .89$	0.02
NOC	8.26 (6.29)	8.47 (5.29)	$F(1, 174) = 0.21$ $p = .65$	0.04
TAE	1.63 (2.11)	1.29 (1.88)	$F(1, 174) = 0.77$ $p = .38$	0.17
CUSA	13.26 (8.90)	18.22 (11.60)	$F(1, 174) = 6.30$ $p = .01$	0.48
MLUw	3.71 (1.02)	4.04 (1.04)	$F(1, 174) = 3.32$ $p = .07$	0.32

Note. ANOVA = analysis of variance; NU = number of utterances; NVU = number of verbal utterances; NW = number of words; SOE = subject omission error; OS = overt subjects; NS = null subjects; NOC = number of object clitics; OCE = object clitic errors; SAE = subject–verb agreement error; TAE = total agreement errors; CUSA = correct use of subject–verb agreement; MLUw = mean length of utterance in words.

Finally, we conducted analyses of sensitivity and specificity for two indicators that showed clinical promise: subject omission errors and subject–verb agreement errors. For the purposes of these analyses, we used the ratio of subject omission errors to null subjects (SOE/NS_Ratio), rather than the subject omission errors, to control for the confounding factor of narrative length. We also created the variable subject–verb agreement error rate (SAER), also allowing us to control for the narrative length by dividing the number of subject–verb agreement errors by the number of verbal utterances. We classified children as DLD if they had at least two standardized scores below the 16th percentile ("impaired") and as TL if they had at least two standardized scores at or above 50th percentile ("unimpaired"). According to the classification criterion, 31 participants did not belong to either group. There were two cases, for which two scores (Expressive Vocabulary and Sentence Imitation) above the 50th percentile overlapped with two scores (Receptive Vocabulary and Pseudoword Repetition) below the 16th percentile, which were excluded from these analyses. An additional case was excluded due to missingness on SOE/NS_Ratio.

We used receiver operating characteristic (ROC) curves to evaluate classification accuracy. ROC graphs are bidimensional representations of the sensitivity (on the x-axis) against 1 − specificity (on the y-axis) for each cutoff point. The area under the curve is called ROC AUC, ranging from 0 to 1, with .5 as chance-level accuracy (50–50 guess) and 1 as perfect accuracy. A rough rule of thumb is that AUC values between .7 and .8 are considered acceptable, AUC values between .8 and .9 are considered excellent, and AUC values above .9 are outstanding (Streiner & Cairney, 2007). Table 7 shows the results of AUC related to SOE/NS_Ratio and SAER for all children and children for each age band. For all children, the AUC of SOE/NS_Ratio was acceptable (.746), whereas the AUC of SAER was not acceptable (.610). The AUC of SOE/NS_Ratio was excellent for children aged 5–7 years but unacceptable for children in the remaining age groups. The AUC of SAER was outstanding for 6-year-old children and excellent for 4-year-old children but unacceptable for children in the remaining age groups.

This analysis requires using a specific threshold score as a cutoff for predicting group membership. We examined three different thresholds for each of the two markers (a total of six analyses) for each age group and for the total sample: (a) the 16th percentile (approximately 1 *SD* from the population mean), (b) the 25th percentile (i.e., the first quartile

Table 7. Areas under the curve (AUC) and 95% confidence intervals (CIs) for predicting risk of developmental language disorder (DLD).

	SOE/NS_Ratio				SAER	
Group	DLD	TL	AUC	95% CI	AUC	95% CI
All participants	25	118	.746	[0.645, 0.848]	.610	[0.484, 0.736]
3-year-olds	2	13	.673	[0.275, 1.000]	.519	[0.000, 1.000]
4-year-olds	4	14	.500	[0.222, 0.778]	.830	[0.637, 1.000]
5-year-olds	3	18	.833	[0.655, 1.000]	.630	[0.300, 0.959]
6-year-olds	4	26	.837	[0.647, 1.000]	1.000	[1.000, 1.000]
7-year-olds	6	27	.802	[0.635, 0.970]	.414	[0.146, 0.681]
8-year-olds	6	20	.683	[0.406, 0.916]	.592	[0.315, 0.869]

Note. SOE/NS_Ratio = subject omission errors divided by null subjects; SAER = subject–verb agreement error rate; TL = typical language.

or −0.675 *SD* from the mean), and (c) the 35th percentile. Given that children in the TL group are expected to make fewer errors than children in the DLD group, children in each age band scoring at or above each cutoff on each measure were defined as "impaired" and children scoring below the cutoff as "unimpaired." Finally, the sensitivity and specificity of each test were calculated. Sensitivity was determined by dividing the number of children from the DLD group scoring at or above a cutoff point on each of the two narrative measures by the total number of children in the DLD group, yielding the probability that an affected child, as ascertained by the standardized tests, will be correctly identified as DLD by the narrative measure. Specificity was determined by dividing the number of children in the TL group scoring at or below a cutoff point on each narrative measure by the total number of children in the TL group, yielding the probability that an unaffected child, as ascertained by the standardized tests, would be correctly identified as unaffected by the narrative measure. Table 8 shows the results for these analyses.

With regard to SAER, none of the cutoffs resulted in acceptable sensitivity, with the exception of 6-year-olds, for whom it was 100% at all thresholds, and specificity ranged from 72% to 100%. With regard to SOE/NS_Ratio, neither the 16th nor the 25th percentile thresholds resulted in acceptable DLD group identification at any age, while the specificity at all thresholds ranged from 77% to 100%. However, the 35th percentile threshold produced high accuracy in group identification for 5-, 6-, and 7-year-olds, with sensitivity of 100%, 75%, and 83% and specificity of 72%, 69%, and 78%, respectively.

Discussion

In this study, we investigated the validity and potential diagnostic utility of indices of narrative microstructure in a representative sample of Arabic-speaking children. Most studies of narrative microstructure were done on data from English. However, the same units of linguistic structure or the same grammatical processes may not be suitable for measuring language development in other languages because of potential differences in the patterns of language acquisition and in ways DLD is manifested in typologically diverse languages. This necessitates research aimed at discovering relevant linguistic parameters that can be reliably used to measure language development and identify language impairment in languages other than English.

We found that general measures of language productivity (i.e., total number of words, total number of utterances, or total number of verbal utterances), commonly used for language sample analyses in English, did not show age sensitivity, convergence with standardized measures, or significant effects when we comparing children with TL and children with suspected DLD. However, we found four measures of language complexity and accuracy to be promising markers for tracking language development in Arabic: MLUw, the number of object clitics, the number of verbs with correct subject agreement, and the number or rate of subject omission errors. These measures showed convergence with multiple standardized language measures, indicating that they are valid measures of language development in GA. These measures also displayed age sensitivity in our regression analyses.

As we discussed above, the traditional measure of MLUm is difficult to implement in templatic languages such

Table 8. Sensitivity and specificity for narrative measures.

Age group	SOE/NS-Ratio: 16th percentile	SOE/NS Ratio: 25th percentile	SOE/NS Ratio: 35th percentile	SAER: 16th percentile	SAER: 25th percentile	SAER: 35th percentile
3-year-olds	0%/100%	50.0%/76.9%	50.0%/69.2%	50.0%/92.3%	50.0%/84.6%	50.0%/76.9%
4-year-olds	0%/100%	25.0%/78.6%	25.0%/64.3%	25.0%/92.9%	50.0%/85.7%	50.0%/78.6%
5-year-olds	33.3%/88.9%	33.3%/77.8%	100%/72.2%	33.3%/88.9%	33.3%/77.8%	33.3%/72.2%
6-year-olds	50.0%/92.3%	50.0%/80.8%	75.0%/69.2%	100.0%/100.0%	100.0%/88.5%	100.0%/73.1%
7-year-olds	33.3%/88.9%	33.3%/77.8%	83.3%/77.8%	16.7%/85.2%	16.7%/74.1%	16.7%/55.6%
8-year-olds	50.0%/95.0%	50.0%/85.0%	50.0%/70.0%	33.3%/90.0%	33.3%/85.0%	33.3%/85.0%
Overall	32.0% (8/25) /93.0% (110/118)	40.0% (10/25) /79.7% (94/118)	64.0% (16/25) /71.2% (84/118)	40.0% (10/25) /91.5% (108/118)	44.0% (11/25) /82.2% (97/118)	44.0% (11/25) /72.0% (85/118)

Note. The first number is sensitivity; the second is specificity. Children with at least two standardized scores below 16th percentile were classified to developmental language disorder for each age group. Children with at least two standardized scores above 50th percentile were classified as typical language for each age group. SOE/NS_Ratio = subject omission errors divided by null subjects; SAER = subject–verb agreement error rate.

as Arabic. MLUw appears a usable alternative, as it is straightforward to calculate and shows age sensitivity. However, we did not find robust increases in utterance length between each of our 12-month-age bands, suggesting that utterance length is not a key indicator of language development in GA-speaking children of ages 3 years and older. This measure also yielded no significant effects when we compared children with and without DLD, indicating its limited diagnostic utility. This measure may, however, be suitable as a criterion-referenced index for tracking a child's progress.

One of the most clinically promising measure we found was related to children's use of null subjects. As discussed above, Arabic is a null subject language, where subjects with a salient, unambiguous, previously mentioned referent are omitted (in neutral, nonemphatic contexts). Subject omission errors are instances when a subject is omitted without a clear referent provided by the discourse. The mastery of this aspect of language requires grammatical and pragmatic skills: knowing that overt pronominal subjects should be omitted, but only if certain discourse conditions for the omission are met.

In our sample, even the youngest group of children showed the knowledge of the grammatical principle that, in Arabic, null subjects are used in place of overt subject pronouns. The overall mean rate of SOE/NS_Ratio was 66% (SD = 0.33) compared to 3.84% (SD = 0.01) of overt pronominal subjects. We also found that subject omission does not significantly decrease with age but becomes progressively more pragmatically appropriate. This improvement likely reflects the growth of skills involved in keeping track of introducing discourse referents and/or gauging their salience for the listeners. While core grammatical knowledge (in all languages) develops early (by the age of 4 and 5 years), pragmatic competence continues to develop into school age, and a number of skills are formed around 7 years of age, such as skilled use of anaphoric reference (Karmiloff-Smith, 1985), metapragmatic skills (Andersen-Wood & Smith, 1997), and mastery of discourse markers (Kyratzis & Ervin-Tripp, 1999). The number of subject omission errors in our study decreases with age but reaches a plateau by the age of 8 years. This measure yielded significant effects when comparing children with TL and children with suspected DLD, with a moderate effect size, indicating a difference between the means of the two groups of approximately 0.5 SD. This measure also showed the best combination of sensitivity and specificity in predicting DLD and TL. While pragmatics is not considered a core deficit in DLD in English, these data suggest that null subject measures may be clinically useful in Arabic for children between the ages of 5 and 7 years.

It is of note that other types of error measures, namely, subject–verb agreement errors, object clitic errors, and total errors, did not show age effects nor differentiated children with DLD from children with TL. This was likely due to a floor effect, as errors were very infrequent even in the youngest age group (e.g., M = 1.21, SD = 1.47 for the 3-year-olds and M = 1.04, SD = 1.61 for the DLD group on SAEs, and even lower numbers for object clitic errors.) These results are in contrast with numerous studies of DLD in English and other Germanic languages, showing a pronounced deficit with tense and agreement markers (Clahsen et al., 1997; Hansson & Nettelbladt, 1995; Rice & Wexler, 1996). This is also at odds with the literature on object clitic errors as a marker of DLD in Romance languages (Arosio et al., 2014; Bortolini et al., 2006). Our findings, however, are in accord with the previous observations that, in Arabic, even young children show high accuracy with the agreement system (Abdalla & Crago, 2008; Aljenaie, 2010). The mechanisms underlying the observed cross-linguistic differences in the acquisition of the tense agreement marking are an important area for further investigations.

In place of morphosyntactic error measures, we found that measures focused on the correct production, number of object clitics and use of correct subject–verb agreement forms, showed age sensitivity, and the latter differentiated children with DLD from those with TL. This observation may reflect grammatical conservatism in language acquisition (Snyder, 2011), which leads children to limit the use of structures until they are established as part of their grammar, rather than using them inaccurately (i.e., with errors of substitution). Measuring the progressive increase in the use of correct structures appears to be a more effective strategy than focusing on errors in languages where errors are infrequent. Another potential strategy may be to construct elicitation probes targeting the production of various object clitic forms and subject–verb agreement and with a variety of verbs rather than relying on spontaneous production measures.

Gender moderated the relationship between age and several indices of narrative microstructure. Girls produced more object clitics than boys of the same age. Also, subject omission error rate and correct use of subject–verb agreement were associated with age for boys (negatively and positively, respectively), but not for girls. For example, the correlation between age and the subject omission error rate for boys was −.41, while for girls, it was −.06, likely reflecting gender differences in language development, with even the youngest girls making few errors. The presence of gender effects is in accord with findings from other languages showing that girls typically outpace boys with respect to language acquisition (Eriksson et al. 2012), underscoring the importance of taking gender into consideration when developing age expectations for language.

In our sample, we identified 15.25% of children as "suspected DLD." This is in line with the results of a United Kingdom–based population study of DLD prevalence (Norbury et al., 2016), in which 91 children out of 529 (17.2%) met the criteria for DLD. Estimates of DLD prevalence vary substantially based on the age and the criteria used for identification, ranging from under 2% to well over 10% (Law et al., 2000). No prevalence estimates exist for Arabic-speaking countries due to a lack of standardized instruments and low level of awareness of the disorder among educators and other providers of child services. Creating standardized measures and metrics for the use of informal diagnostic procedures therefore is highly important.

Finally, our investigation into diagnostic accuracy of narrative indicators showed that using measures of narrative microstructure requires adjusting one's expectations regarding the threshold separating children with DLD and TL. We found that to reach acceptable levels of sensitivity, the threshold must be kept quite high: 35th percentile in our study. We also found that the measures that worked well for children aged 5–7 years lacked diagnostic accuracy in children aged 4 years and younger and older than 7 years. This suggests that clinical information that results from narrative elicitation is rather subtle. It is meant to be used as a supplement to (not a substitute of) standardized measures, as the two offer complementary data with each affording important advantages and limitations. This also suggests that identification of language impairment in children younger than 5.0 years and older than 7.9 years requires more sensitive indicators than those used in our study.

Our study had a number of limitations, which are important to address in future research. First, we lacked a clinical group with preidentified DLD, a natural limitation since DLD is not a commonly diagnosed disorder in the Arab world. Second, our measures of subject–verb agreement in our linguistic analyses did not differentiate affixal processes from templatic nor examined errors in different verb classes. It is possible that a more fine-grained approach would yield interesting results. However, such analysis would require special sentence elicitation measures. Story narrative elicitation used in our study allowed children their own choice of verbs, which, given grammatical conservatism, may result in minimizing errors. Another limitation was the absence of measures of nonverbal intelligence as a means to test divergent validity and for more precision in classification of language disorder (e.g., ruling out cases of intellectual disability, specific and nonspecific language disorder). Finally, to make reliable generalizations, further studies with larger samples are needed, particularly in a language such as Arabic, with its considerable dialectal diversity.

In summary, our study offers quantitative data on the use of indices of narrative microstructure in Arabic-speaking children between the ages of 3 and 8 years. It shows that narrative elicitation yields a set of indices that can be used as valid metrics for documenting language development in children acquiring GA, including MLUw, subject omission error rate, number of object clitics, and correct use of subject–verb agreement. These measures correlated with standardized measures of language development for GA previously shown to have psychometric properties and showed age sensitivity. Measures of subject omission errors and correct use of subject–verb agreement differentiated children with typical and atypical levels of language development. The former yielded acceptable levels of sensitivity and specificity at the threshold of 35th percentile.

Acknowledgments

This work was supported with a grant from the Ministry of Higher Education, Saudi Arabia. We are grateful to Sergey Kornilov for his contributions to the Arabic Language: Evaluation of Function construction. We also thank Nabilah Halal for her help with the translation of the story narrative and item construction for the Arabic Language: Evaluation of Function battery, Natalie Banker for creating the artwork for the assessment, Mardheya Alsamadani for the assistance with transcribing and coding the narratives, and Faisal Alharbi for providing transcription accuracy check. Our special gratitude goes to the educators from the Al-Hasa area who assisted us with data collection and data entry and the children and their families who participated in this research.

References

Abdalla, F., Aljenaie, K., & Mahfoudhi, A. (2013). Plural noun inflection in Kuwaiti Arabic-speaking children with and without specific language impairment. *Journal of Child Language, 40*(1), 139–168. https://doi.org/10.1017/S0305000912000499

Abdalla, F., & Crago, M. (2008). Verb morphology deficits in Arabic-speaking children with specific language impairment. *Applied Psycholinguistics, 29*(2), 315–340. https://doi.org/10.1017/S0142716408080156

Aljenaie, K. (2010). Verbal inflection in the acquisition of Kuwaiti Arabic. *Journal of Child Language, 37*(4), 841–863. https://doi.org/10.1017/S0305000909990031

Al-Khaledi, M., Lincoln, M., McCabe, P., Packman, A., & Alshatti, T. (2009). The attitudes, knowledge and beliefs of Arab parents in Kuwait about stuttering. *Journal of Fluency Disorders, 34*(1), 44–59. https://doi.org/10.1016/j.jfludis.2009.02.003

Andersen-Wood, L., & Smith, B. R. (1997). *Working with pragmatics: A practical guide to promoting communicative confidence*. Speechmark.

Arosio, F., Branchini, C., Barbieri, L., & Guasti, M. T. (2014). Failure to produce direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children. *Clinical Linguistics & Phonetics, 28*(9), 639–663. https://doi.org/10.3109/02699206.2013.877081

Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(4), 905–924. https://doi.org/10.1044/1092-4388(2001/072)

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The Journal of Child Psychology and Psychiatry, 37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bortolini, U., Arfé, B., Caselli, C. M., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and nonword repetition. *International Journal of Language & Communication Disorders, 41*(6), 695–712. https://doi.org/10.1080/13682820600570831

Bortolini, U., Caselli, M. C., & Leonard, L. B. (1997). Grammatical deficits in Italian-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*(4), 809–820. https://doi.org/10.1044/jslhr.4004.809

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy, 18*(1), 1–21. https://doi.org/10.1191/0265659002ct224oa

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. (2010). Arabic morphology in the neural language system. *Journal of Cognitive Neuroscience, 22*(5), 998–1010. https://doi.org/10.1162/jocn.2009.21273

Clahsen, H., Bartke, S., & Göllner, S. (1997). Formal features in impaired grammars: A comparison of English and German SLI children. *Journal of Neurolinguistics, 10*(2–3), 151–171. https://doi.org/10.1016/S0911-6044(97)00006-7

Clahsen, H., & Dalalakis, J. (1999). *Tense and agreement in Greek SLI: A case study*. Department of Language and Linguistics, University of Essex Colchester.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46*(5), 1029–1037. https://doi.org/10.1044/1092-4388(2003/082)

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry, 42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Dromi, E., & Berman, R. A. (1982). A morphemic measure of early language development: Data from modern Hebrew. *Journal of Child Language, 9*(2), 403–424. https://doi.org/10.1017/S0305000900004785

Dromi, E., Leonard, L. B., Adam, G., & Zadunaisky-Ehrlich, S. (1999). Verb agreement morphology in Hebrew-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*(6), 1414–1431. https://doi.org/10.1044/jslhr.4206.1414

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology, 30*(2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Fahim, D. (2017). Verb morphology in Egyptian Arabic developmental language impairment. *Arab Journal of Applied Linguistics, 2*(1), 49–73.

Ferguson, C. A. (1959). Diglossia. *Word, 15*(2), 325–340. https://doi.org/10.1080/00437956.1959.11659702

Fletcher, P., Leonard, L. B., Stokes, S. F., & Wong, A. M.-Y. (2005). The expression of aspect in Cantonese-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(3), 621–634. https://doi.org/10.1044/1092-4388(2005/043)

Foroodi Nejad, F. (2011). *Towards the identification of linguistic characteristics of specific language impairment in Persian* (Doctoral dissertation). University of Alberta. https://era.library.ualberta.ca/items/71305aaa-8ef2-426b-8fc3-40480932308c.

Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 42*(1), 59–75. https://doi.org/10.1080/13682820600783210

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*(1), 88–98. https://doi.org/10.1044/0161-1461.3101.88

Habash, N. (2007). Arabic morphological representations for machine translation. In A. Soudi, A. van den Bosch, & G. Neumann (Eds.), *Arabic computational morphology: Knowledge-based and empirical methods* (Vol. 38, pp. 263–285). Springer. https://doi.org/10.1007/978-1-4020-6046-5_14.

Hamann, C. (2015). Specific language impairment in German speaking children. In S. Stavrakaki (Ed.), *Specific language impairment. Current trends in research* (pp. 215–251). John Benjamins. https://doi.org/10.1075/lald.58.10ham

Hansson, K., & Nettelbladt, U. (1995). Grammatical characteristics of Swedish children with SLI. *Journal of Speech and Hearing Research, 38*(3), 589–598. https://doi.org/10.1044/jshr.3803.589

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.

Holes, C. (2013). *Colloquial Arabic of the Gulf*. Routledge.

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology, 15*(2), 177–191. https://doi.org/10.1044/1058-0360(2006/017)

Kaderavek, J. N., & Sulzby, E. (2000). Narrative production by children with and without specific language impairment: Oral narratives and emergent readings. *Journal of Speech, Language, and Hearing Research, 43*(1), 34–49. https://doi.org/10.1044/jslhr.4301.34

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes, 1*(1), 61–85. https://doi.org/10.1080/01690968508402071

Kenstowicz, M. (1989). The null subject parameter in modern Arabic dialects. In O. A. Jaeggli & K. J. Safir (Eds.), *The null subject parameter* (pp. 263–275). Springer. https://doi.org/10.1007/978-94-009-2540-3_9.

Kotby, M. N., El-Sady, S., & Hegazi, M. (2010). Thirty-five years of care of child language in Egypt. *Topics in Language Disorders, 30*(1), 84–91. https://doi.org/10.1097/TLD.0b013e3181d08273

Kunnari, S., Savinainen-Makkonen, T., Leonard, L. B., Mäkinen, L., Tolonen, A.-K., Luotonen, M., & Leinonen, E. (2011). Children with specific language impairment in Finnish: The use of tense and agreement inflections. *Journal of Child Language, 38*(5), 999–1027. https://doi.org/10.1017/S0305000910000528

Kyratzis, A., & Ervin-Tripp, A. K. (1999). The development of discourse markers in peer interaction. *Journal of Pragmatics, 31*(10), 1321–1338. https://doi.org/10.1016/S0378-2166(98)00107-6

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language & Communication Disorders, 35*, 165–188. https://doi.org/10.1080/136828200247133

Leonard, L. B. (2017). Cross-linguistic studies of child language disorders. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 328–344). Psychology Press.https://doi.org/10.4324/9781315283531-13

Leonard, L. B., Caselli, M. C., Bortolini, U., McGregor, K. K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition, 2*(2), 151–179. https://doi.org/10.1207/s15327817la0202_2

Leonard, L. B., Dromi, E., Adam, G., & Zadunaisky-Ehrlich, S. (2000). Tense and finiteness in the speech of children with specific language impairment acquiring Hebrew. *International Journal of Language & Communication Disorders, 35*(3), 319–335. https://doi.org/10.1080/136828200410609

Leonard, L. B., Kas, B., & Pléh, C. (2009). The use of tense and agreement by Hungarian-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 52*(1), 98–117. https://doi.org/10.1044/1092-4388(2008/07-0183)

Lust, B., Chien, Y.-C., & Fynn, S. (1987). What children know: Methods for the study of first language acquisition. In B. Lust (Ed.), *Studies in the acquisition of anaphora. Vol. 2: Applying the constraints* (pp. 271–356). Reidel. https://doi.org/10.1007/978-94-009-3387-3_12

Mahfoudhi, A., & Abdalla, F. (2017). Language development and impairment in Arabic-speaking children: Introduction to the special issue. *Arab Journal of Applied Linguistics, 2*(1), 1–11.

Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools, 34*(4), 332–342. https://doi.org/10.1044/0161-1461(2003/027)

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the U.S. Preventive Services Task Force. *Pediatrics, 117*(2), e298–e319. https://doi.org/10.1542/peds.2005-1467

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders, 38*(3), 287–313. https://doi.org/10.1080/136820310000108133

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Ntelitheos, D., & Idrissi, A. (2017). *Language growth in child Emirati Arabic*. Paper presented at the Perspectives on Arabic Linguistics XXIX: Papers from the Annual Symposium on Arabic Linguistics, Milwaukee, Wisconsin, 2015. https://doi.org/10.1075/sal.5.10nte

Omar, M. (1973). The acquisition of Egyptian Arabic as a native language. *De Gruyter Mouton.*. https://doi.org/10.1515/9783110819335

Paradis, J., Crago, M., & Genesee, F. (2003). *Object clitics as a clinical marker of SLI in French: Evidence from French–English bilingual children*. Paper presented at the Proceedings of the 27th Annual Boston University Conference on Language Development.

Paul, R., & Norbury, C. (2012). *Language disorders from infancy through adolescence—E-book: Listening, speaking, reading, writing, and communicating*. Elsevier Health Sciences.

Pavelko, S. L., & Owens, R. E. (2017). Sampling Utterances and Grammatical Analysis Revised (SUGAR): New normative values for language sample analysis measures. *Language, Speech, and Hearing Services in Schools, 48*(3), 197–215. https://doi.org/10.1044/2017_LSHSS-17-0022

Qasem, F., & Sircar, S. (2017). Imperative as root infinitive analogue in Yemeni Ibbi Arabic: Two case studies. *Arab Journal of Applied Linguistics, 2*(1), 1–19.

Rakhlin, N. V., Aljughaiman, A., & Grigorenko, E. L. (2019). *Assessing language development in Arabic: The Arabic Language: Evaluation of Function (ALEF)* (pp. 1–16). Applied Neuropsychology: Child. https://doi.org/10.1080/21622965.2019.1596113.

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language, 19*(56), 187–206. https://doi.org/10.1177/014272379901905603

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language, 88*(2), 229–247. https://doi.org/10.1016/S0093-934X(03)00101-9

Reuterskiöld, C., Hansson, K., & Sahlén, B. (2011). Narrative skills in Swedish children with language impairment. *Journal of Communication Disorders, 44*(6), 733–744. https://doi.org/10.1016/j.jcomdis.2011.04.010

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239

Ryan, C. (2013). Language use in the United States: 2011. *American Community Survey Reports, 22,* 1–6.

Saiegh-Haddad, E., & Ghawi-Dakwar, O. (2017). Impact of diglossia on word and non-word repetition among language impaired and typically developing Arabic native speaking children. *Frontiers in Psychology, 8,* 2010. https://doi.org/10.3389/fpsyg.2017.02010

Saiegh-Haddad, E., Hadieh, A., & Ravid, D. (2012). Acquiring noun plurals in Palestinian Arabic: Morphology, familiarity, and pattern frequency. *Language Learning, 62*(4), 1079–1109. https://doi.org/10.1111/j.1467-9922.2012.00727.x

Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy* (pp. 3–28). Springer. https://doi.org/10.1007/978-94-017-8545-7_1

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument. *Journal of Speech Language Pathology and Audiology, 30*(4), 224.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324–339. https://doi.org/10.1044/jslhr.4302.324

Shahmahmood, T. M., Jalaie, S., Soleymani, Z., Haresabadi, F., & Nemati, P. (2016). A systematic review on diagnostic procedures for specific language impairment: The sensitivity and specificity issues. *Journal of Research in Medical Sciences, 21,* 67–67. https://doi.org/10.4103/1735-1995.189648

Simons, G. F., & Fennig, C. D. (2018). Ethnologue: Languages of the world (*21st ed.*). SIL International. http://www.ethnologue.com

Smolík, F., & Vávrů, P. (2014). Sentence imitation as a marker of SLI in Czech: Disproportionate impairment of verbs and clitics. *Journal of Speech, Language, and Hearing Research, 57*(3), 837–849. https://doi.org/10.1044/2014_JSLHR-L-12-0384

Snyder, W. (2011, November). Children's grammatical conservatism: Implications for syntactic theory. In A. Lukyanchenko, K. Gor, N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (Vol. 1, pp. 1–20). Cascadilla Press.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment. *Language, Speech, and Hearing Services in Schools, 37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry, 52*(2), 121–128. https://doi.org/10.1177/070674370705200210

Thordardottir, E. T., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research, 54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Thordardottir, E. T., Keheyia, E., Lessard, N., Sutton, A., & Trudeau, N. (2010). Typical performance on tests of language knowledge and language processing of French-speaking 5-year-olds. *Revue canadienne d'orthophonie et d'audiologie, 34*(1), 5–16.

Thordardottir, E. T., & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research, 50*(3), 698–715. https://doi.org/10.1044/1092-4388(2007/049)

Thordardottir, E. T., & Weismer, S. E. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language, 18*(52), 001–032. https://doi.org/10.1177/014272379801805201

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 953–968). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00076-6.

Van Der Lely, H. K., & Battell, J. (2003). Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language, 79*(1), 153–181. https://doi.org/10.1353/lan.2003.0089

Verhoeven, L., Steenge, J., & van Balkom, H. (2011). Verb morphology as clinical marker of specific language impairment: Evidence from first and second language learners. *Research in Developmental Disabilities, 32*(3), 1186–1193. https://doi.org/10.1016/j.ridd.2011.01.001

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*(1), 23–79. https://doi.org/10.1016/S0024-3841(98)00029-1

Wittke, K., & Spaulding, T. J. (2018). Which preschool children with specific language impairment receive language intervention? *Language, Speech, and Hearing Services in Schools, 49*(1), 59–71. https://doi.org/10.1044/2017_LSHSS-17-0024