JSLHR

Research Article

Nonword Repetition Skills in Gulf Arabic–Speaking Children With Developmental Language Disorder

Saleh Shaalan[a]

Purpose: This study examined the performance of Gulf Arabic–speaking children with developmental language disorder (DLD) on a Gulf Arabic nonword repetition (GA-NWR) test and compared it to their age- and language-matched groups. We also investigated the role of syllable length, wordlikeness, and phonological complexity in light of NWR theories.
Method: A new GA-NWR test was conducted with three groups of Gulf Arabic–speaking children: school-age children with DLD, language-matched controls (LCs), and age-matched controls (ACs). The test consisted of two- and three-syllable words that either had no clusters, medial clusters, final clusters, or medial + final clusters.
Results: The GA-NWR distinguished between the performance of children with DLD and the LC and AC groups. Results showed significant syllable length, wordlikeness, and phonological complexity effects. Differences between the DLD and typically developing groups were seen in two- and three-syllable nonwords; however, when compared on nonwords with no clusters, children with DLD were not significantly different from the LC group.
Conclusions: The GA-NWR test differentiated between children with DLD and their ACs and LCs. Findings, therefore, support its clinical utility in this variety of Arabic. Results showed that phonological processing factors, such as phonological complexity, may have stronger effects when compared to syllable length effects.
Supplemental Material: https://doi.org/10.23641/asha.12996812

The ability to repeat nonwords is considered a potent predictor of language impairment, especially during the early stages of language development (Chiat & Roy, 2008; Gathercole, 2006). A vast number of studies have shown that children with developmental language disorders (DLDs; formerly known as specific language impairment) have significant problems in nonword repetition (NWR; Archibald & Gathercole, 2006; Bishop et al., 1996; Conti-Ramsden, 2003; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Gathercole & Baddeley, 1990; Gray, 2003; Montgomery, 1995, 2002; Munson et al., 2005; Oetting & Cleveland, 2006; Roy & Chiat, 2004; Snowling et al., 1991). These consistent impairments in NWR lead researchers to investigate processes that underlie these deficits in children with language impairments in general and those with DLD in particular. Some attribute difficulties in NWR to a "central" deficit in phonological short-term memory (PSTM; Gathercole, 2006; Gathercole & Baddeley, 1990), whereas others argue that, along with PSTM deficits, there are other contributing factors to NWR deficits, such as deficits in phonological processing skills (Chiat, 2001; Snowling et al., 1991).

The PSTM account of NWR (Gathercole, 2006; Gathercole & Baddeley, 1990) is based on the working memory model of Baddeley (2003), which argues that deficits in the phonological loop component and especially in the phonological store are the main cause of language deficits in children with DLD. Deficits in this part of the working memory can cause problems in forming appropriate phonological representations and learning new words (Archibald & Gathercole, 2006; Baddeley et al., 1998; Gathercole, 2006; Gathercole & Baddeley, 1990). Deficits in PSTM can be assessed using NWR tasks, such as the Children's Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996) or the Nonword Repetition Test (Dollaghan & Campbell, 1998). Studies of NWR in children with DLD have shown that, as the number of syllables increases, the performance of those children drops significantly when

[a]Department of Allied Health Services, Mohammed bin Rashid Center for Special Education—Operated by The New England Center for Children, Abu Dhabi, United Arab Emirates

Correspondence to Saleh Shaalan: sshaalan@neccuae.org

Editor-in-Chief: Stephen M. Camarata
Editor: Mary Alt

Received February 2, 2020
Revision received May 5, 2020
Accepted July 27, 2020
https://doi.org/10.1044/2020_JSLHR-20-00040

Disclosure: The author has declared that no competing interests existed at the time of publication.

compared to language-matched control (LC) and age-matched control (AC) groups (Archibald & Gathercole, 2006; Bishop et al., 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990). Both the CNRep and the Nonword Repetition Test are widely used in NWR studies. When Archibald and Gathercole (2006) compared these two tests, they found that they both distinguished between children with DLD and their ACs with a high level of accuracy. However, only on the CNRep, where there are many nonwords with clusters or sublexical units, did children with DLD perform significantly less well than their LCs (Archibald & Gathercole, 2006). Although in the latest version of the PSTM theory, its proponents acknowledge that NWR is multiply determined and deficits in phonological storage may not be the sole cause of language impairment, they stress that difficulties in NWR are largely due to deficits in the storage of phonological representations (Gathercole, 2006).

Proponents of the phonological processing account of NWR explain that there are various auditory, phonological, and motor processes involved in NWR (Bowey, 1996; Chiat, 2001; Snowling et al., 1991). Deficits in one or more of these processes might affect children's performance on NWR. Chiat (2001) explains that impaired phonological processing causes subsequent disruption of the mapping process, which is responsible for establishing word and sentence structures. Therefore, these basic phonological limitations affect lexical and syntactic development. Subsequent studies of some phonological factors found that NWR performance was affected by output processes (Krishnan et al., 2013; Pigdon et al., 2020), phonotactic probabilities and wordlikeness (Edwards et al., 2004; Roy & Chiat, 2004; Szewczyk et al., 2018), and phonological complexity represented by the presence of consonant clusters (Gallon et al., 2007; Marshall & van der Lely, 2009). In the next section, we examine some of the evidence for the two phonological processing factors that we investigate in this study, namely, wordlikeness and consonant clusters.

Wordlikeness. It refers to "the extent to which a sound sequence is typical of words in a language" (Bailey & Hahn, 2001, p. 568), and it could be determined by factors such as phonotactic probabilities or lexical neighborhood (Bailey & Hahn, 2001). While NWR tests maximize the use of unwordlike stimuli to assess the influence of phonological memory, including wordlike stimuli will draw on long-term lexical knowledge (Gathercole, 1995) and will provide some advantage to typically developing (TD) children over those with language impairment. Results from the Preschool Repetition Test (Roy & Chiat, 2004) with 66 TD children aged 2–4 years showed that these children were sensitive to lexical familiarity as they scored better on real words than they did on nonwords. These results were replicated in a larger sample of 315 children (Chiat & Roy, 2007). Furthermore, Gathercole et al. (1991) found that nonwords rated as being closer to real words were recalled more easily than those rated as less wordlike on the CNRep test. Armon-Lotem and Meir (2016) used an NWR test that was designed to assess the effects of syllable length, wordlikeness, and phonological complexity with 230 monolingual and bilingual children with and without DLD in Hebrew and Russian. Results showed that, along with sentence repetition, the NWR task helped in differentiating monolingual and bilingual children with DLD from their TD peers, although separate bilingual cutoff points were needed to reach the acceptable level of accuracy. The NWR test used was a short version of the test designed by Armon-Lotem and Chiat (2012), who defined wordlikeness in terms of similarity to the target language, so wordlike stimuli have the morphophonological structure of target language, while nonwordlike stimuli did not provide this information (Armon-Lotem & Chiat, 2012). Studies of word repetition and NWR in Arabic reported that children with language impairments performed significantly better on repeating real words than they did on nonwords in Palestinian Arabic (PA; Saiegh-Haddad & Ghawi-Dakwar, 2017) and Gulf Arabic (GA; Khater, 2016).

Consonant clusters. Earliest remarks to the role of consonant clusters in NWR came when Gathercole and Baddeley (1989) studied NWR skills in 104 TD children between the ages of 4 and 5 years and found that children at the age of 4 years were sensitive to the presence of consonant clusters; however, by the age of 5 years, they were less affected by consonant clusters. Marshall and van der Lely (2009) designed an NWR test that consisted of only nine phonemes and three-syllable nonwords to examine the performance of three groups of children (those with DLD only, those with dyslexia only, and those with DLD and dyslexia) and compared their NWR skills to three groups of TD younger children (aged 5, 7, and 9 years). They found that all three clinical groups had particular difficulties with medial clusters when compared to the TD groups. They used these findings to argue that NWR tests are not mere indices of PSTM and that phonological complexity has unique contributions to NWR. Moreover, Gallon et al. (2007) and Marshall et al. (2002) found that difficulties with repetition of nonwords with phonological complexities persisted and were observed even in older children and adolescents with DLD. These effects were also reported in bilingual populations. Dos Santos and Ferré (2018) developed an NWR test to help differentiate language disorders in bilingual children in multilingual settings. They conducted this test with 67 monolingual and bilingual children with and without DLD, aged between 5;6 and 8;6 (years;months). Many of the bilingual participants were French–Arabic bilinguals with varying degrees of proficiency and coming from different linguistic backgrounds (e.g., Lebanese Arabic, Moroccan Arabic, Tunisian Arabic, Algerian Arabic, and Libyan Arabic). This test, where nonwords were no longer than three syllables and with limited wordlikeness effects, was able to identify bilingual children with DLD. They noted that the scores of monolingual and bilingual children with DLD dropped significantly on nonwords with consonant clusters. Therefore, the authors concluded that there were stronger effects of consonant clusters in comparison to the weak effects of phonological storage.

NWR has been found to be unaffected by dialectal, socioeconomic differences (Burt et al., 1999; Chiat &

Polišenská, 2016; Dollaghan & Campbell, 1998; Engel et al., 2008; Oetting & Cleveland, 2006) or differences in IQ (Bishop et al., 1996; Conti-Ramsden et al., 2001; Ellis Weismer et al., 2000). These qualities of NWR make it useful where there is a dire lack of formal and informal tools to assess language skills in languages such as GA (Shaalan, 2009), which is spoken in Bahrain, Kuwait, Qatar, the United Arab Emirates, and the eastern part of Saudi Arabia. Each of these regions has their own variety of GA, though they share many linguistic features and therefore are considered subdialects of GA (Holes, 2000). In this study, we are examining the GA variety spoken in Qatar.

There are a few studies that looked at NWR skills in children with language impairments in Arabic varieties in general and GA in particular. However, there were studies that examined the role of NWR and other phonological processing skills in reading in Arabic (e.g., Elbeheri et al., 2011; Mahfoudhi et al., 2020; Taibah & Haynes, 2011). Saiegh-Haddad and Ghawi-Dakwar (2017) conducted a word and nonword repetition test with PA-speaking children and found that children with DLD performed significantly less well than their age-matched peers on both tasks, adding evidence for the utility of NWR tests in this variety of Arabic. They also studied the impact of diglossia and phonological novelty on these children by manipulating the phonological distance between Modern Standard Arabic (MSA) and spoken PA. Diglossia in Arabic is characterized by the presence of two linguistic systems that differ significantly from each other in phonology, semantics, syntax, and morphology and exist in different contexts. While the spoken variety is the native language of the child and she is exposed to it most of the time, MSA is the variety used in education, literacy texts, news and media, and formal correspondence. Saiegh-Haddad and Ghawi-Dakwar (2017) showed that NWR in children with DLD was affected by the phonological distance between PA and MSA. They demonstrated that kindergarteners and first-grade PA-speaking children with DLD had more difficulties repeating nonwords that comprised novel phonemes (phonemes existing in MSA but not in PA) and across all syllable lengths. The stimuli used were one- to four-syllable long nonwords and had simple syllabic structures (i.e., none of the stimuli had consonant clusters). These findings revealed that the phonological skills of PA-speaking children with DLD were particularly impacted by this phonological novelty when compared to their TD peers. GA, however, has a different phonological system, and therefore, this effect of phonological distance may warrant further investigation in children with DLD.

In GA, Khater (2016) developed a GA word and nonword repetition test and used it with toddlers and preschoolers with and without language impairment. This word and nonword repetition test (WNRep) followed the design of the Preschool Repetition Test (Roy & Chiat, 2004). Forty-eight 1- to 3-syllable words and nonwords were created by transforming nonwords from existing words (e.g., the word /ki:s/ [bag] was transformed as the nonword /sa:k/). The participants consisted of 44 TD children and a clinical group of 15 children aged 2–4 years. The clinical group consisted of children who were referred to the speech therapy department at the main public hospital in Qatar and who did not have any other concomitant disorders. The results showed significant difference on the WNRep test between the two groups, and this supported the utility of this NWR test in GA. However, there is some uncertainty about the characteristics of the clinical group in this study. These very young children did not receive diagnoses, some could be late talkers, and some of them performed within the TD range. Moreover, the WNRep test did not control for wordlikeness and sublexical effects. For example, "la.kus" (which is transformed from the real word "se:.kel" [bike]) has the syllable "la," which means "no" in GA.

The results from Saiegh-Haddad and Ghawi-Dakwar (2017) and Khater (2016) show that NWR tasks can be a potential clinical marker of language impairment in Arabic. The design of these tests, however, did not try to control systematically for wordlikeness effects. Some of the nonwords followed some common patterns in Arabic, for example, /qa:.fu:s/ in Saiegh-Haddad and Ghawi-Dakwar (2017) and /yas.sa:.ri/ in Khater (2016). Neither of these tests did have nonwords with consonant clusters as they are not preferred in certain contexts in PA and MSA (Saiegh-Haddad & Ghawi-Dakwar, 2017) and due to the young age of participants in the study of Khater (2016). Both of these studies compared the performance of children with DLD to a group of their age-matched peers only. Many studies of DLD in older children prefer to add a language-matched group, as this will help identify processes and linguistic structures that reveal where children with DLD lag behind their age- and language-matched peers. The presence of such persistent deficits supports the deficit versus delay hypothesis of language impairment and help identify potential clinical markers of DLD across languages.

In the following study, we examine the performance of school-age children with DLD and their ACs and LCs on a new Gulf Arabic nonword repetition (GA-NWR) test to assess if it will be able to distinguish between the performance of children with DLD and their TD peers. Such finding will support the viability of this GA-NWR test in identifying children with DLD in this population. We will also examine the contributions of three important NWR factors, namely, syllable length (as an index of PSTM) and wordlikeness and phonological complexity (as phonological processing factors).

Method

Participants

Thirty-three GA-speaking children from Qatar participated in this experiment: 11 diagnosed with DLD (*M* = 7;9, *SD* = 10.5 months), 11 TD AC children (*M* = 7;9, *SD* = 10.7 months), and 11 TD LC children (*M* = 6;0, *SD* = 8.3 months). The LC group was matched with the DLD group based on their scores on the Gulf Arabic Sentence Comprehension Test (Shaalan, 2017). Children were recruited from two kindergartens (for the LC group) and four primary schools, and three children were recruited through personal acquaintance. All institutional review board and necessary administrative approvals were obtained before the start of the study. All the participating children lived in Doha, the capital of Qatar, and come from Qatari GA–speaking households. They all received general language tests, and children with DLD were selected based on the following criteria: All of them scored −1.5 *SD* or below on at least two out of four language tests from Shaalan (2017), or they had a score of −2.0 *SD* on one of these tests. The tests were the Sentence Comprehension Test, the Expressive Language Test (which examines the production of various morphosyntactic structures in GA), the Sentence Repetition Test, and the Arabic Picture Vocabulary Test (APVT; see Shaalan, 2017, for a description of these tests). These tests showed high levels of reliability and validity; however, no information was available on the specificity and sensitivity of the cutoff points for these tests. All children scored within normal range on either the Test of Nonverbal Intelligence, Third Edition (Brown et al., 1997) for children aged 6;0 and above or on two performance subtests from Wechsler Preschool and Primary Scale of Intelligence–Third Edition (Wechsler, 2002) for children who were younger than 6 years. The Wechsler Performance IQ subtests were the Block Design and the Picture Completion subtests, and they were recommended as an appropriate short form of nonverbal IQ (see LoBello, 1991; Tomblin et al., 1997). Both the DLD and AC groups consisted of eight boys and three girls, while the LC group had six boys and five girls. Table 1 summarizes participants' characteristics.

There was a good matching between the DLD group and the LC group based on their scores on the Sentence Comprehension Test. The DLD group was not significantly different from the LC group on the Digit Span subtest score (which included forward and backward digit span) of the Wechsler Preschool and Primary Scale of Intelligence–Third Edition (Wechsler, 2002) or the APVT (Shaalan, 2017). The LC group was not significantly different from the AC group on the APVT or the Digit Span subtest. The difference on the nonverbal IQ scores between the DLD and AC groups was not significant. All children passed articulation and childhood apraxia of speech screening tests.

Table 1. Descriptive summary data for the children with developmental language disorder (DLD; *n* = 11), language-matched control group (LC; *n* = 11), and age-matched control group (AC; *n* = 11).

Group	Age	SC	APVT	DS	TONI-3
DLD					
M (years;months)	7;9	25.0	48.5	8.1	93.0
SD (months)	10.5	4.5	15.6	1.2	7.4
Range (years;months)	6;3–9;1	18–31	28–76	6–11	85–109
LC					
M (years;months)	6;0	26.2	59.2	9.1	[a]
SD (months)	8.3	2.9	16.6	2.9	[a]
Range (years;months)	5;0–6;11	22–31	38–89	5–14	[a]
AC					
M (years;months)	7;9	33.1	71.7	11.3	99.5
SD (months)	10.7	2.3	19.9	1.3	7.8
Range (years;months)	6;3–9;0	30–38	47–110	9–14	88–111

Note. SC = Sentence Comprehension Test raw score (Shaalan, **2017**); APVT = Arabic Picture Vocabulary Test raw score (Shaalan, **2017**); DS = Digit Span task from the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 2002) raw score; TONI-3 = Test of Nonverbal Intelligence, Third Edition standard score.

[a]All the language-matched control children scored above the cutoff score of 16 on the shorter version of Wechsler Performance IQ (see LoBello, 1991; Tomblin et al., 1997).

Materials and Procedure

About the Phonology of Qatari GA

Arabic poses some challenges to those who want to create an NWR test. Arabic, such as other Semitic languages, is a nonconcatenative language where words consist of intertwined roots and templates or patterns (McCarthy, 1982). Roots consist of three or four consonants that carry the basic semantic meaning. For example, the root "K T B" (write) is used to derive words where vocalic templates are intertwined with the roots, so the template "ɪ-a:" and the same root produce /kɪta:b/ (book), while the template "a:- ɪ" with the same root produce /ka:tɪb/ (writer) and so forth. Therefore, the phonological structure of Arabic is constrained by its morphological structure, and it is difficult to tease them apart.

The Sound System of Qatari GA

Qatari GA has 30 consonants and the following eight vowels: /a/ and /a:/, /ɪ/, /i/, /ʊ/, /ʊ/, /e:/, and /o:/ (Mustafawi, 2006). When compared to MSA, Qatari GA shares all the MSA consonants except /dˤ/, and in addition, it has the following phonemes that are not part MSA: /tʃ/, /g/, and /lˤ/ (Bukshaisha, 1985). Bukshaisha (1985) listed 12 types of syllables in Qatari GA; 10 of them are common, while the other two are not. The 10 common types are /cv/, /cv:/, /ccv/, /ccv:/, /cvc/, /cv:c/, /cvcc/, /ccvc/, /cv:cc/, and /ccv:c/.

Stress in Qatari GA. Like many other varieties of Arabic, stress in Qatari is regular and depends on syllable weight. The final syllable is stressed if it has a long vowel /cv:/ or consonant clusters (cvcc), including geminate consonants; otherwise, stress falls on the penultimate syllable (Holes, 1989).

Variables considered in the design of the GA-NWR test. The design of NWR tests and the stimuli chosen can affect children's performance on the test (see Archibald & Gathercole, 2006). The GA-NWR test was designed to compare the effects of phonological storage (indexed by syllable length), wordlikeness effect, and the effects of phonological complexity (consonant clusters). Therefore, careful consideration was taken to control variables that have been found to influence NWR. These include articulatory complexity (output processes), lexicality effects, respecting phonotactic rules of Arabic, morphological information, syllable number, and wordlikeness.

Articulatory complexity. In order to control for effects of articulatory complexity, all consonants chosen were early developing consonants and are typically mastered by the age of 4 years in GA (Alqattan, 2015). Therefore, while Qatari GA has 30 consonants, only nine consonants were selected to form the nonwords. These sounds were /b/, /d/, /t/, /k/, /f/, /s/, /m/, /n/, and /l/. Moreover, following the recommendation of Dollaghan and Campbell (1998), only tense vowels were chosen. Therefore, short tense vowels (a, u, ɪ), which are common in MSA, GA, and most spoken varieties of Arabic, were employed to form the nonwords. No diphthongs or long vowels were included.

Lexicality effects. In order to reduce lexicality effects and neutralize the influence of previous vocabulary knowledge, an effort was made to minimize the number of syllables that are actual words inside the nonwords. Due to the design of the test and the limited number of consonants, it was difficult to eliminate all syllables that can be actual words. Therefore, out of the 140 syllables, 18 were possible words (i.e., 12.9% of the total number of nonwords). However, many of these syllables are MSA words that may not be in the lexicon of these children at this time (e.g., /ʁad/ "worked hard," /mas/ "touched"). As for sublexical effects, which are related to phonotactic probability of phoneme sequences (see Gathercole, 2006; Stokes et al., 2006), no database that lists consonant probabilities in GA is available, and therefore, it was difficult to determine the influence of phonotactic probability on the performance of children with DLD and TD children in this experiment. Although it is expected there is little phonotactic probability effects due to the use of nonexistent roots.

Language-specific phonotactic rules. Alongside attempts to control articulatory complexity and lexicality effects, nonwords in the GA-NWR test obeyed the phonotactic rules of Arabic. Therefore, no words with initial clusters were included, because most of these initial clusters are formed by shortening a vowel and then deleting it to form a consonant cluster (e.g., /ħɪsˤaːn/ "horse" to /ħsˤaːn/), and sometimes, a short vowel is introduced in front of initial consonant clusters /ɪħsˤaːn/ (Bukshaisha, 1985). To avoid this controversy of whether there is an initial cluster or not, none of the nonwords composed started with an initial cluster. Language-specific phonotactic rules were respected when forming the trilateral nonroots. For example, the Obligatory Contour Principle on place of articulation (OCP-Place) was respected. The OCP-Place states that roots with homorganic consonants are disfavored or rare in Arabic (Frisch et al., 2004; Frisch & Zawaydeh, 2001; McCarthy, 1986). Therefore, certain consonants belonging to some classes (e.g., labials, coronal obstruents, gutturals) are not commonly found in proximity to each other. For example, the cooccurrence of labial consonants (b, f, m) in the same root is infrequent compared to other consonants. The OCP-Place constraint is gradiently influenced by the similarity of the consonants within each consonant class (see Frisch et al., 2004). Based on these phonotactic constraints and the consonants chosen, the following seven roots were selected: /S T L/, /K D F/, /D L S/, /S B N/, /D N F/, /K M S/, /D F L/. All these roots are nonexistent in GA. These roots were checked in dictionaries (e.g., Holes, 2000; Qafisheh, 1997), and their nonexistence was confirmed. When consulting the biggest dictionary of Classical Arabic compiled in the 13th century (Ibn Manzur, 1290/1981), two of these roots were found, namely /D L S/ and /D N F/, but it was very unlikely that any of the children in the study had encountered any of these two roots from Classical Arabic. Ten college-educated teachers were given these seven roots in the common a–a vocalic pattern (e.g., /dalas/, /kadaf/). Two out of these 10 teachers identified the root /D L S/ and knew its meaning, while the rest did not identify the meaning of any of the roots, though they recognized that they could be possible Arabic words.

Morphological information. Another language-specific factor that was controlled in the design of the GA-NWR test was accessing morphological information. In Arabic, grammatical morphemes are affixed initially, medially, or finally to the root and therefore the nonwords were carefully selected to avoid including such morphemes. Therefore, none of the nonwords started with /b/ (a preposition in Arabic as in "bi" [in]), /f/ (a conjunction as in "fa" [and] or preposition as in /fi/ [in]), /l/ (a preposition as in "li" [for]), /n/ (a pronoun as in "naʔkʊl" [we eat]), /t/ (a feminine third person pronoun), and /m/ (which is commonly used to derive nouns, places, etc.). Possible suffixes, such as /m/ (used in plural third person pronouns), /k/ (second-person pronoun), and /t/ (used in feminine pronouns and to indicate past tense) were avoided. Based on this, all nonwords included in the test ended with the following consonants only: /b/, /f/, /l/, /n/, /s/, and /d/. Moreover, since some of the Arabic patterns could have some extra consonants in nonroot internal positions (e.g., /t/, /s/, and /n/), these consonants were not added to the consonantal roots used in this experiment.

Syllable number. GA can have up to seven syllabic words (Bukshaisha, 1985). However, most of the words longer than three syllables are formed by adding inflectional morphemes; therefore, all the nonwords included in the task were either two or three syllables. Due to the root-and-pattern nature of the language, it was not possible to create monosyllabic nonwords that are phonotactically possible. Creating one-syllable nonwords that respect Arabic phonotactics could add some lexical or morphological information that will provide advantage to TD children over those with DLD. For example, one of the one-syllable nonwords created by Khater (2016) was "saːk," which has the vocalic pattern indicating past tense in GA and many of the other single syllable nonwords could be parts of other words. Many studies found that differentiation in performance of children with DLD starts on three-syllable words and upwards (Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Montgomery, 2004). Therefore, syllable number effects can be examined by comparing the performance of children on two- versus three-syllable nonwords.

Wordlikeness. Since Arabic is a root-and-pattern language, triliteral roots cannot exist by themselves and need a pattern of vocalic sounds. However, patterns should respect phonotactic rules of Arabic, and therefore, using a nonexistent pattern will violate these rules. Therefore, we opted to use patterns that are infrequent in GA. The following vocalic patterns were employed to generate the experimental nonwords in combination with the consonantal roots: a-ʊ, a-ʊ-a, ʊ-ɪ, and ʊ-ɪ-a. These vocalic patterns are used in MSA to form passive voice structures (e.g., /kʊsɪr/ and /kʊsɪra/ "was broken"), but passive voice in GA has a different structure (e.g., /ɪnkɪsar/ "was broken") and does not employ the one from MSA. The test included eight control nonwords, which had one of the most frequent vocalic patterns in Arabic, namely, /a–a/ and /a–a–a/ (Holes, 2000). Therefore, the experimental nonwords were characterized by their low wordlikeness effect, while the control nonwords had higher wordlikeness effects.

About the GA-NWR Test

The NWR test consisted of 56 nonwords: 48 experimental nonwords and eight control nonwords. The experimental stimuli contained six nonexistent triconsonantal roots that do not appear in the Qatari GA lexicon, and they were used to construct two- and three-syllable nonwords with four types of cluster conditions (no clusters, medial clusters, final clusters, and medial + final [M + F] clusters), so each root was used to construct eight nonwords. The vocalic patterns used with these roots were existing but infrequent patterns in GA. See Table 2 for an example for one of the six triconsonantal roots.

The eight control nonwords were created by taking a nonexistent root and using the same types of clusters; however, the vocalic patterns used were ("a–a" and "a–a–a"), which are considered two of the most frequent vocalic patterns in Arabic.

These 56 nonwords were recorded by a female native speaker of Qatari GA. Recording of stimuli was conducted in a soundproof room using the Computerized Speech Lab (CSL 4300, Kay Elemetrics). The stimuli were then randomized and put into two lists, and children were assigned randomly to one of these nonword lists (see List A and List B in Supplemental Material S1). A *t* test was conducted at the end of data collection, and it showed no significant difference between children's performance on these two lists, $t(31) = -0.37$, $p = .71$.

Procedure

All sessions were conducted in a quiet room by the author who is a licensed speech-language pathologist and native speaker of GA. The average time it took to complete the experiment was 8–10 min; children expressed no fatigue, and none asked for breaks as this was the only task completed on that day. The instructions for each child were the equivalent of the following (in Arabic): "You will listen to funny and mixed up words and I want you to repeat them the way you hear them. Now let's try this…." This was followed by four trial items. Stimuli were presented from a laptop through a pair of external speakers. Children's productions were audiotaped through a Sony microphone attached to the laptop and using PRAAT software (Boersma & Weenink, 2004). Children's responses were transcribed online by the examiner, and when needed, the recorded stimuli were checked. Each repetition was scored either *correct* (1) or *incorrect* (0). Minor misarticulations (especially distortion of /s/ or substituting /θ/ for /s/) were counted as correct. No repetition of the stimuli was allowed. Another speech-language pathologist, who is a native speaker of GA, scored the full recordings of five children (two with DLD and three TD), and the interrater reliability was 95%.

Results and Analysis

NWR accuracy was scored at word level, so each word received a score of 1 (*correct*) or 0 (*incorrect*). Raw scores were then converted to percentages. Table 3 shows the percentage of correctly recalled words for all groups. It is evident from Table 3 that the DLD group found the NWR task more challenging than both control groups, especially as the number of marked structures increases.

The distribution of the scores of all children on the NWR test is displayed in the box plot in Figure 1. It clearly shows the significant difference between the group of children with DLD and the two TD groups, especially the clear lack of any overlap between AC children and those with DLD.

The data were analyzed using mainly analyses of variance (ANOVAs) and *t* tests for follow-ups due to the robustness of ANOVAs with small populations; however, due to concerns about normal distribution, we will also refer to the results of nonparametric tests (e.g., the Kruskal–Wallis test, the Mann–Whitney *U* test, and the Wilcoxon signed-ranks test). The skewness and kurtosis of the scores on the GA-NWR test for the three groups were as follows: DLD (0.75 and 2.0, respectively), LC (−1.7 and 3.7, respectively), and AC (−0.4 and −1.3, respectively). Shapiro–Wilk tests of normality results were significant for the LC group only ($p = .5$ for the DLD group, $p = .035$ for the LC group, and $p = .13$ for the AC group). The use of Kruskal–Wallis test was recommended by Khan and Rayner (2003),

Table 2. An example of a root and vocalic patterns used to create a list of two- and three-syllable nonwords.

Root	No. of syllables	Pattern	Syllable type: No cluster	Medial cluster	Final cluster	M + F cluster
/S T L/	2 syllables	a–ʊ	Sa.tʊl	Das.tʊl	Sa.tʊlb	Das.tʊlb
	3 syllables	a–ʊ–a	Da.sʊ.tal	Das.bʊ.tal	Da.sʊ.talb	Da.sʊm.talb

Note. Full stops indicate syllable boundary. M + F = medial + final.

Table 3. Group descriptive statistics (in percentages of correct repetitions) for the children with DLD (*n* = 11), age-matched control (AC) children (*n* = 11), and language-matched control (LC) children (*n* = 11) on the nonword repetition task.

Syllable length	Cluster type	DLD (*n* = 11)		LC (*n* = 11)		AC (*n* = 11)		All groups	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
2 Syllables	No cluster (*n* = 6)	86.3	16.4	92.4	11.5	100	0		
	Medial cluster (*n* = 6)	77.2	23.9	87.8	10.8	97	6.9		
	Final cluster (*n* = 6)	68.2	21.7	89.4	17.1	93.9	11.2		
	Medial + final cluster (*n* = 6)	42.5	25.1	71.2	31.7	80.4	22.1		
	All 2 syllables	68.6	26.9	85.2	20.7	92.8	4.5	82.3	16.7
3 Syllables	No cluster (*n* = 6)	63.6	20.9	78.8	15	92.4	11.5		
	Medial cluster (*n* = 6)	43.9	27.2	69.7	31.5	77.2	17.1		
	Final cluster (*n* = 6)	30.3	25.6	68.2	22.9	83.3	16.7		
	Medial + final cluster (*n* = 6)	21.3	22.5	45.6	22.5	53.1	26.7		
	All 3 syllables	39.8	28.4	65.5	25.9	76.6	23.3	60.7	23.6
Overall mean (*n* = 56)		55.8	17	77.4	16.4	84.1	9.6		

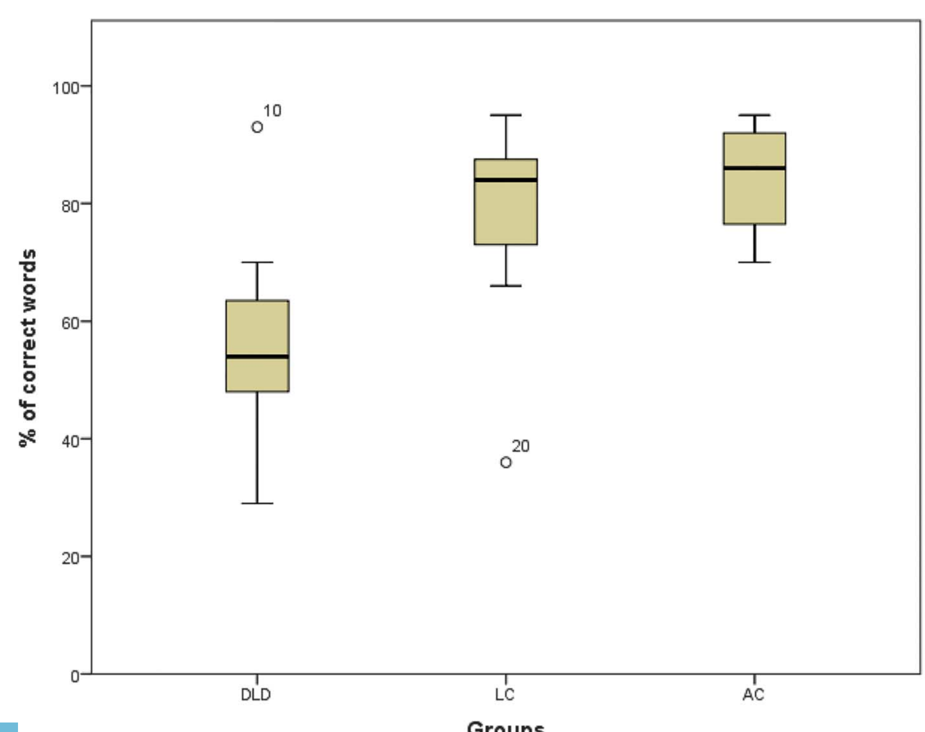
especially when kurtosis is high. For the sake of brevity, we will only report areas of agreements and/or disagreements between parametric and nonparametric tests.

A 3 ANOVA group (DLD, LC, AC) × 2 length (two syllables, three syllables) × 4 cluster type (no clusters, medial clusters, final clusters, and M + F clusters) was conducted. It revealed a significant main effect of group, $F(2, 30) = 12.4$, $p < .001$, $\eta_p^2 = .45$; syllable length, $F(1, 30) = 71.7$, $p < .001$, $\eta_p^2 = .70$; and cluster type, $F(3, 90) = 60.9$, $p < .001$, $\eta_p^2 = .67$. Cluster × Group interaction was also significant, $F(6, 90) = 2.5$, $p = .021$, $\eta_p^2 = .15$; however, Syllable Length × Group interaction was not significant, $F(3, 90) = 2.7$, $p = .08$. Length × Cluster interaction was not significant either, $F(3, 90) = 1.5$, $p = .22$, nor was the Group × Syllable × Cluster interaction significant, $F(6, 90) = 1.25$, $p = .29$. In the following, the main effects of independent variables are discussed. This is followed by examining the significant interaction effects.

Analysis of Main Effects

Main effects of group. Post hoc test with Bonferroni correction revealed that the DLD group performed significantly

Figure 1. A box plot showing the distribution of scores of all three groups. DLD = developmental language disorder group; LC = language-matched control group; AC = age-matched control group.



worse than both the LC group ($p < .01$) and the AC group ($p < .001$) on the overall accuracy of the NWR test. There was no significant difference between the AC and LC groups. Similar results were obtained from the nonparametric tests (the Kruskal–Wallis test and the Mann–Whitney *U* test).

Main effects of syllable length. The NWR task used bisyllabic and trisyllabic nonwords only. Two syllable nonwords were recalled more easily than three-syllable words, $t(32) = 9.1$, $p < .001$. This significant difference was also seen in the results of the Wilcoxon signed-ranks test.

The results of the Kruskal–Wallis test showed that there was a statistically significant difference between the three groups on both two- and three-syllable nonwords. A Mann–Whitney *U* test with Bonferroni correction showed that there was no significant difference between the DLD and LC groups on both the two- and three-syllable nonwords; however, the AC group performed significantly better than the DLD group, with the two TD groups not differing in their performance on either syllable length. These are consistent with the lack of Syllable × Group interaction reported in the main ANOVA.

Main effects of cluster types. The ANOVA showed a significant main effect of cluster type, $F(3, 90) = 60.9$, $p < .001$, $\eta_p^2 = .67$. There were four types of nonwords used in the NWR task: nonwords with no clusters, nonwords with a medial cluster only, nonwords with a final cluster only, and those with M + F clusters. Table 4 summarizes the overall performance of groups on the four types of clusters.

Further analysis using post hoc tests with Bonferroni correction revealed that there was a significant difference on the performance of all groups on nonwords with no clusters versus all other types of clusters. There was a significant difference between nonwords with M + F clusters on one hand and M + F clusters on the other hand, $t(30) = 23.3$, $p < .001$, and $t(30) = 20$, p > .001, respectively. However, there was no significant difference between medial-only and final-only clusters. The results from Wilcoxon signed-ranks tests were similar to the results of the *t* tests. Therefore, the following generalization about hierarchy of cluster difficulty holds:

0 cluster > 1 cluster (M or F) > 2 clusters (M + F)

Table 4. Means and standard deviations (in percentage of) correct nonword repetitions for each type of cluster.

Cluster type	DLD *M* (*SD*)	LC *M* (*SD*)	AC *M* (*SD*)	Total *M* (*SD*)
No cluster	75.0 (16.2)	85.6(12.6)	96.2 (5.7)	85.6 (14.8)
Medial cluster	60.6 (22.8)	78.8 (18.0)	87.1 (8.8)	75.5 (20.4)
Final cluster	49.3 (19.7)	78.8 (18.1)	88.6 (10.7)	72.2 (23.4)
M + F clusters	31.8 (21.0)	58. (22.5)	66.7 (19.7)	52.3 (25.4)
Overall mean	54.2 (23.0)	75.4 (20.4)	84.7 (14.0)	

Note. DLD = developmental language disorder group; LC = language-matched control group; AC = age-matched control group; M + F = medial + final.

Analysis of Interactions

Group × Cluster Type interaction. There was a significant Group × Cluster Type interaction. Figure 2 depicts the interaction between cluster types and groups. It clearly shows a pattern of increasing difficulty with nonwords as they increase in phonological complexity. This effect had a greater impact on the DLD group than it did on the other TD groups.

One-way ANOVAs were performed to investigate the effect of groups at each cluster level. Results showed that the three groups differed significantly at each level: no clusters, $F(2, 30) = 8.27$, $p = .001$, $\eta_p^2 = .36$; medial clusters, $F(2, 30) = 6.7$, $p = .004$, $\eta_p^2 = .31$; final clusters, $F(2, 30) = 16.8$, $p < .001$, $\eta_p^2 = .53$; and M + F clusters, $F(2, 30) = 8.2$ $p = .001$, $\eta_p^2 = .35$.

Post hoc tests with Bonferroni correction revealed that the AC group consistently performed better than the DLD group on all types of clusters and the AC and LC groups were not significantly different on any of the cluster types. Therefore, the following section will focus on the difference between the DLD group and the LC group on each cluster condition.

On nonwords with no clusters, the DLD group was not significantly different from the LC group, $t(30) = -10.6$, $p = .15$. The difference between these two groups on medial cluster nonwords using Bonferroni correction was close to significance, $t(30) = -18.3$, $p = .06$. However, on final cluster nonwords, the difference was significant, $t(30) = 30.0$, $p = .001$. In addition, on nonwords with M + F clusters, the difference between the DLD and LC groups was significant, $t(30) = -9.0$, $p = .019$.

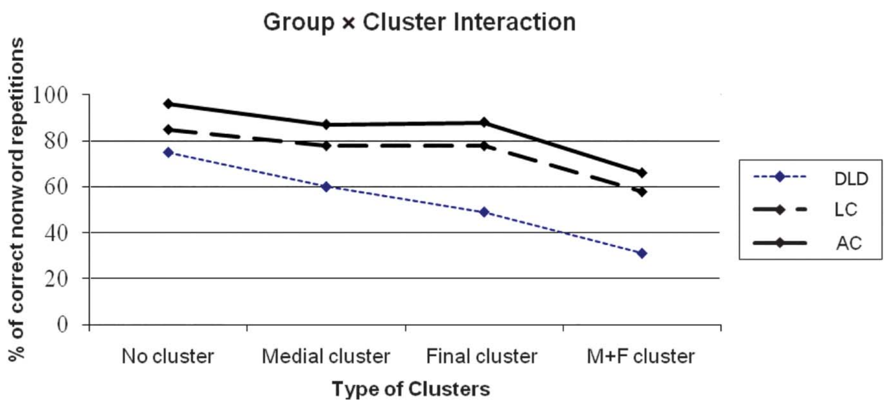
A Kruskal–Wallis test showed that there was a statistically significant difference between the three groups on all types of clusters. A Mann–Whitney *U* test with Bonferroni correction showed that the DLD group performed significantly worse than the AC group on all types of clusters, and it performed significantly worse than the LC group on nonwords with final clusters and M + F clusters only. The two TD groups were not significantly different from each other. Therefore, these results do not differ from those reported in the ANOVAs and follow-up *t* tests.

Wordlikeness Effects

To compare the effects of wordlikeness on the performance of the three groups, their scores on the experimental ($n = 48$) versus control nonwords ($n = 8$) were examined. While the experimental nonwords contained low-frequency patterns (i.e., patterns that are rarely used in GA), the control nonwords consisted of highly frequent vocalic patterns that are very common in GA and other varieties of Arabic. Table 5 summarizes the results of the three groups on both types of the nonwords patterns.

A repeated-measures ANOVA showed there was a main effect of wordlikeness, $F(1, 30) = 6.6$, $p = .015$, $\eta_p^2 = .18$. Overall, children found the high-frequency patterns, which were more wordlike, easier to recall than the other

Figure 2. Group × Cluster interaction. DLD = developmental language disorder group; LC = language-matched control group; AC = age-matched control group; M + F = medial + final.



less frequent patterns, $t(32) = 2.37$, $p = .02$. Results showed that there was a significant effect of group, $F(2, 30) = 7.0$, $p < .01$, $\eta_p^2 = .31$, with the DLD group being outperformed by both the LC and AC groups ($p = .009$ and $p = .009$). Moreover, there was a significant Wordlikeness × Group interaction, $F(2, 30) = 3.9$, $p = .03$ $\eta_p^2 = .20$. Subsequent analysis using Bonferroni correction showed that the DLD group did not benefit significantly from wordlikeness effects when compared to the LC group. A paired-sample *t* test with Bonferroni correction showed that the difference between the DLD group's performance on high- versus low-frequency patterns was not significant, $t(10) = 1.74$, $p = .11$. The LC group on the other hand benefitted significantly from wordlikeness effect with its performance increasing significantly on the high-frequency patterns, $t(10) = 5.0$, $p < .001$. As for the AC group, there was no significant difference between their performance on low- versus high-frequency patterns, $t(10) = -0.8$, $p = .4$. No significant difference was found between the two TD groups. A one-way ANOVA of nonwords with low-frequency (experimental) nonwords was significant, $F(2, 30) = 12.4$, $p \leq .001$, with the DLD group being outperformed by the LC group ($p = .006$) and the AC group ($p < .001$). A one-way ANOVA of nonwords with high-frequency (control) patterns was significant ($p = .04$); however, the only difference among the groups was that the DLD was outperformed by the LC only ($p = .04$).

Similar to the parametric test results, a Wilcoxon signed-ranks test found a significant effect of wordlikeness in that the three groups of children were more likely to correctly recall nonword with high-frequency patterns. The results of Wilcoxon signed-ranks tests on the performance of the three groups on low- versus high-frequency pattern nonwords were similar to the parametric tests. Moreover, a Kruskal–Wallis test showed that there was a statistically significant difference in the performance of the three groups on the nonwords with low-frequency patterns with the Mann–Whitney *U* test showing that the DLD group performed significantly worse than the AC group on low-frequency patterns. However, the difference between the DLD and LC groups did not reach significance ($p = .069$) due to the adjusted Bonferroni. This difference was significant in the parametric test. A Kruskal–Wallis test showed that the performance of the three groups on nonwords with high-frequency patterns did not reach significance ($p = .06$), thus differing from the reported parametric test.

Overall, although the results of the parametric tests might indicate the presence of a strong wordlikeness effect, with qualitative and quantitative differences in the performance of the three groups on nonwords with low- and high-frequency patterns, the fact that the effects of high-frequency patterns were seen in one group only (the LC in the parametric test) and the difference on worldlike nonwords did not reach significance in the nonparametric tests might imply a possibly weakened wordlikeness effect.

Table 5. Means and standard deviations (in percentages) of the scores of all groups on experimental nonwords that have nonfrequent patterns ($n = 48$) versus control nonwords that have very frequent patterns ($n = 8$).

Type of patterns		DLD (n = 11)	LC (n = 11)	AC (n = 11)
Nonfrequent (Experimental nonwords)	*M*	54.2	75.4	84.7
	SD	17.2	17.2	9.4
Frequent (Control nonwords)	*M*	66.0	89.9	80.9
	SD	26.2	19.0	18.0

Note. DLD = developmental language disorder group; LC = language-matched control group; AC = age-matched control group.

Discussion

This study set out to investigate two main issues about NWR. First is whether the NWR test devised in this study can distinguish the performance of GA-speaking children with DLD from their ACs and LCs. Second, this study endeavored to shed light on some of the competing theories of the nature of NWR deficits in children with DLD, especially those that attribute these impairments to a central impairment in phonological capacity (Gathercole, 2006; Gathercole & Baddeley, 1990) or phonological processing theories that argue that, in addition to phonological storage, there are other important factors that can impact NWR, such as wordlikeness and phonological complexity (Chiat, 2001; Snowling et al., 1991).

Clinical Implications for the Study of NWR in Arabic

The results of this investigation of NWR skills in GA-speaking children with DLD show that these children performed significantly worse than their TD peers matched on age or language abilities. Therefore, these results extend the viability of NWR task as a possible clinical marker of DLD to Qatari GA. These results support the findings of Saiegh-Haddad and Ghawi-Dakwar (2017), where they found that NWR test differentiated the performance of PA-speaking children with DLD from their age-matched peers. They also concur with the results from Khater (2016), whose clinical group had significant challenges with the NWR task when compared to their age-matched peers. Moreover, this study included both age- and language-matched groups.

The usefulness of this task may not be constrained to GA, as the design of the test and the stimuli used may render it useful and clinically viable in other varieties of Arabic. The stimuli used in this task consist of eight early developing sounds that exist in all Arabic dialects, and the syllable structures (cv), (cvc), and (cvcc) used are common in many dialects of Arabic (Watson, 2002). Therefore, this task might be a useful tool in the identification of children at risk of language impairments; especially with the paucity of assessment tools in Arabic. Moreover, many studies have found that NWR is less influenced by socioeconomic factors and is therefore less prone to bias than other conventional language measures (Campbell et al., 1997; Ebert et al., 2008; Ellis Weismer et al., 2000). The NWR test could avail itself to be used with a wider population than the current sample of children, who mostly come from middle-class households.

However, it is possible that dialect-specific factors may influence the performance of school-age children who are exposed to MSA and to their spoken variety. The phonological distance between the two sound systems may add another level of complexity to the NWR test (Saiegh-Haddad & Ghawi-Dakwar, 2017). Therefore, larger scale empirical studies are needed to confirm these findings in various varieties of Arabic and further examinations of the diagnostic accuracy of this test are warranted.

Theoretical Implication

The findings of significant differences between children with DLD and the other two control groups on this NWR task that comprises only two- and three-syllable words organized according to their syllabic structures cannot be fully explained by the PSTM account of DLD as proposed by Gathercole and Baddeley (1990), as they are better accounted for by the phonological processing account of NWR. According to the PSTM account, limitations in the phonological loop, the part of working memory responsible for storing phonological information, are the "main" cause of deficits in NWR, vocabulary learning, and syntactic development in children with DLD (Baddeley et al., 1998; Gathercole & Baddeley, 1990). This account predicts a significant interaction between syllable length and groups, where children with DLD perform significantly poorer than their TD peers, indicating a stronger effect of phonological storage over other phonological processing factors. This prediction was not borne out in this study. Although we cannot rule out that the lack of significant interaction between syllable length and groups could be due to the small size of the three groups, the fact that there were significant Group × Cluster Type and Group × Wordlikeness interactions reveals that wordlikeness and, especially, phonological complexity might show better differentiation when compared to length effects.

A detailed examination of the relationship between length and cluster effects reveals that the presence or absence of clusters had a stronger effect on the performance of children with DLD than syllable length. Seventy-five percent of all nonwords used in this experiment had consonant clusters in them, and when children with DLD were compared with the LC children on bisyllabic and trisyllabic nonwords with no clusters, the difference was not significant. This is in contrast to the significant difference between the two groups (DLD vs. LC) on nonwords with clusters regardless of syllable length. This weak effect of clusterless nonwords versus the strong effect of nonwords with clusters is not supported by the strong PSTM account (Baddeley et al., 1998; Gathercole & Baddeley, 1990), which posits that PSTM is the main factor in determining NWR performance or even its latest version (Gathercole, 2006), which acknowledges the contribution of other phonological factors (such as prosody, wordlikeness, and lexicality effects), yet it continues to argue that PSTM has a central role in NWR and its contribution exceeds those of the aforementioned factors (Gathercole, 2006).

It is important, however, to reiterate that children with DLD performed significantly worse than their AC group on both two- and three-syllable nonwords. Therefore, the difference between the group of children with DLD and their ACs in terms of capacity limitations in PSTM is evident and cannot be ignored. However, this experiment shows that their performance is better differentiated from their LCs by phonological complexity and to a lesser extent by wordlikeness as predicted by phonological processing accounts.

The results of the GA-NWR show that children with DLD are sensitive to two important phonological processing factors, namely, wordlikeness and phonological complexity (consonant clusters).

Wordlikeness effects. Results of this study show that wordlikeness effects could have significantly influenced the performance of children with DLD on the GA-NWR task.

The experimental nonwords used in the Arabic NWR test were based on nonfrequent and therefore less wordlike patterns, while the control stimuli contained more frequent (more wordlike) nonwords. The results of both parametric and nonparametric tests show that the difference in accuracy on low- versus high-frequency pattern was not significant in the group of children with DLD, while the LC group found the high-frequency patterns significantly easier to recall. Therefore, unlike TD children, children with DLD did not benefit from previous linguistic knowledge stored in long-term memory to form new phonological representations. The AC group did not benefit from pattern effects due to possible maturational factors. Studies by Berman (2003), Karwar and Sakran (1998), and Malenky (1997) reported that pattern awareness develops at a later age (7–10 years old) in TD Hebrew and PA-speaking children. So, while the LC are developing their awareness of high- and low-frequency patterns, those with DLD seem to be lagging behind them, while the AC group could be in the process of mastering this awareness or have already mastered this skill.

Wordlikeness effects were reported in PA-speaking children with DLD (Saiegh-Haddad & Ghawi-Dakwar, 2017) and with GA-speaking children at risk of language impairment in Khater (2016), where all groups performed better on real words versus nonwords. One possible explanation for the lack of wordlikeness effects in the DLD in this group is the type of stimuli used here (nonwords with high-frequency patterns vs. nonwords with low-frequency patterns), while in the other two experiments, the stimuli were real words versus nonwords. Finally, it seems that despite the presence of a general strong wordlikeness effects in parametric and nonparametric tests and the presence of group and wordlikeness interaction effects in the parametric tests, the effects of high frequency patterns were seen in one of the three groups only (the LC) and in the ANOVA only, while the Kruskal–Wallis results did not reach significance (p = .06). Therefore, wordlikeness effects were not as strong as the effects of phonological complexity.

Phonological complexity (consonant clusters) effects. The results of the GA-NWR test show that children with DLD had increasing difficulties with nonwords as the number of consonant clusters increased, as predicted by the phonological processing account of NWR (Bowey, 1996; Chiat, 2001). The presence of these marked structures increases the phonological complexity of nonwords, and children with DLD seemed more prone to such complexity than their age- or language-matched peers. These results replicate the finding of Gallon et al. (2007) and Marshall et al. (2002) where children with DLD had difficulties repeating even monosyllabic and bisyllabic nonwords when they contained marked syllabic and metrical structures. Both Marshall et al. (2002) and Gallon et al. (2007) found that the performance of children with DLD deteriorated as the number of complex structures increased, similarly to our findings of increasing difficulties as the number of consonant clusters increases from 0 (*no cluster*) to 1 (*medial or final cluster*) to 2 (*M + F cluster*).

This strong phonological complexity effect was also found in bilingual French-Arabic–speaking children with and without DLD who came from different Arabic linguistic backgrounds (Algerian Arabic, Lebanese Arabic, Libyan Arabic, Moroccan Arabic, and Tunisian Arabic), despite the small number of syllables used in that study (no longer than three syllables). They also limited wordlikeness effects by not building their nonwords on any specific phonological system. Similar to our results, they reported a significant drop in the performance of monolingual and bilingual children with DLD on nonwords with clusters, while the decline from bisyllabic to trisyllabic items seemed smaller, thus supporting their conclusion of the presence of stronger phonological complexity effects.

Conclusions and Summary

This experiment examines the viability of NWR as a clinical marker of DLD in an understudied language, namely GA. The NWR test was designed to examine the role of syllable length, wordlikeness, and phonological complexity (consonant clusters) in light of two major theories of NWR: PSTM (Gathercole & Baddeley, 1990) and phonological processing theory (Chiat, 2001; Snowling et al., 1991).

Analysis of the data shows the GA-NWR test differentiated between the performance of children with DLD (M_{age} = 7;9) and LCs and ACs. Findings showed a stronger effect for phonological processing factors (especially consonant clusters) when compared to the role of syllable length. Results revealed significant difference in performance of children with DLD even on two-syllable nonwords due to the presence of more phonologically complex structures. Wordlikeness effects were also seen although they were stronger in parametric tests and were not as strong as the consonant cluster effects. Overall, these results show that deficits in PSTM alone cannot explain all the results of this NWR test. However, it is important to stress that phonological processing accounts do not neglect the role of phonological storage, as it is considered one of the important factors in NWR (along with auditory, perceptual, phonological, and output processes). One important factor to consider is that we examined the effects of syllable length based on two- and three-syllable nonwords only, and we were not able to compare the three groups on a wider range of syllable lengths. It is also possible that the lack of Syllable × Group interaction seen here could be due to the small size of the groups; however, the results support a more prominent role for phonological complexity and, to a certain extent, wordlikeness effects. It is expected that the combination of increasing length and complexity will pose a higher level of difficulty for children with DLD. Therefore, it seems NWR accounts such as the phonological processing skills (e.g., Chiat, 2001; Snowling et al., 1991), which account for the various processes influencing the performance of children with DLD, might provide better explanations for the nature of NWR deficits.

It is important to examine these findings taking into consideration that replication with a larger sample is needed and in different varieties of Arabic. Moreover, the diagnostic accuracy of this NWR test should be examined with a representative sample of GA-speaking children in order to assess its clinical utility and diagnostic accuracy, including sensitivity and specificity. Moreover, no information was available on the specificity and sensitivity of the cutoff points for the tests used to diagnose children with DLD. It is important to note that not all varieties of Arabic tolerate the presence of clusters in M + F positions. Moreover, as older Arabic-speaking children become more exposed to MSA patterns, the experimental patterns (or infrequent patterns) used in this study might be affected, and more attention should be paid to the influence of diglossia on language impairment with older school-age children.

Acknowledgment

The author received an Overseas Research Students Award from University College London.

References

Alqattan, S. (2015). *Early phonological acquisition by Kuwaiti Arabic children* [Unpublished PhD thesis]. Newcastle University, United Kingdom.

Archibald, L., & Gathercole, S. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49*(5), 970–983. https://doi.org/10.1044/1092-4388(2006/070)

Armon-Lotem, S., & Chiat, S. (2012). How do sequential bilingual children perform on non-word repetition tasks. In *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 53–62). Cascadilla Press.

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders, 51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44,* 568–591. https://doi.org/10.1006/jmla.2000.2756

Berman, R. A. (2003). Children's lexical innovations. *Language Acquisition and Language Disorders, 28,* 243–292. https://doi.org/10.1075/lald.28.13ber

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioral 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

Boersma, P., & Weenink, D. (2004). *Praat: Doing phonetics by computer* (Version 4.2. 28) [Computer program]. Institute of Phonetic Sciences, University of Amsterdam.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63*(1), 44–78. https://doi.org/10.1006/jecp.1996.0042

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *TONI-3: Test of Nonverbal Intelligence, Third Edition*. Stoelting.

Bukshaisha, F. (1985). *An experimental phonetic study of some aspects of Qatari Arabic* [Unpublished doctoral dissertation]. University of Edinburgh, Edinburgh, Scotland.

Burt, L., Holm, A., & Dodd, B. (1999). Phonological awareness skills of 4-year-old British children: An assessment and developmental data. *International Journal of Language & Communication Disorders, 34*(3), 311–335. https://doi.org/10.1080/136828299247432

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40*(3), 519–525. https://doi.org/10.1044/jslhr.4003.519

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises predictions and evidence. *Language and Cognitive Processes, 16*(2–3), 113–142. https://doi.org/10.1080/01690960042000012

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research, 59*(5), 1179–1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research, 50*(2), 429–443. https://doi.org/10.1044/1092-4388(2007/030)

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *The Journal of Child Psychology and Psychiatry, 49*(6), 635–645. https://doi.org/10.1111/j.1469-7610.2008.01881.x

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. *The Journal of Child Psychology and Psychiatry, 42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition, 25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish nonword repetition: Stimuli development and preliminary results. *Communication Disorders Quarterly, 29*(2), 67–74. https://doi.org/10.1177/1525740108314861

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Elbeheri, G., Everatt, J., Mahfoudhi, A., Abu Al-Diyar, M., & Taibah, N. (2011). Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia, 17*(2), 123–142. https://doi.org/10.1002/dys.430

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition

performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*(4), 865–878. https://doi.org/10.1044/jslhr.4304.865
Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research, 51*(6), 1580–1587. https://doi.org/10.1044/1092-4388(2008/07-0210)
Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory, 22,* 179–228. https://doi.org/10.1023/B:NALA.0000005557.78535.3c
Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language, 77*(1), 91–106. https://doi.org/10.1353/lan.2001.0014
Gallon, N., Harris, J., & van der Lely, H. K. J. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics and Phonetics, 21*(6), 435–455. https://doi.org/10.1080/02699200701299982
Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23,* 83–94. https://doi.org/10.3758/BF03210559
Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513–543. https://doi.org/10.1017/S0142716406060383
Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200–213. https://doi.org/10.1016/0749-596X(89)90044-2
Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J
Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. The Psychological Corporation.
Gathercole, S. E., Willis, C., & Baddeley, A. D. (1991). Nonword repetition, phonological memory, and vocabulary—A reply to Snowling, Chiat, and Hulme. *Applied Psycholinguistics, 12*(3), 375–379. https://doi.org/10.1017/S0142716400009280
Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36*(2), 129–151. https://doi.org/10.1016/S0021-9924(03)00003-0
Holes, C. (1989). *Gulf Arabic*. Routledge.
Holes, C. (2000). *Dialect, culture, and society in Eastern Arabia: Glossary*. Brill.
Ibn Manzur. (1290/1981). *Lisaan al 'arab*. Daar Al Ma'aarif.
Karwar, C., & Sakran, M. (1998). *Developmental aspects of phonological and morphological awareness in Palestinian Arabic* [Seminar paper]. Department of Communication Disorders, Tel Aviv University, Israel.
Khan, A., & Rayner, G. D. (2003). Robustness to non-normality of common tests for the many-sample location problem. *Advances in Decision Sciences, 7,* 187–206. https://doi.org/10.1155/S1173912603000178
Khater, M. (2016). *The relationship between nonword repetition, root and pattern effects, and vocabulary in Gulf Arabic–speaking children* [Unpublished Doctoral thesis]. City, University of London, London, United Kingdom.
Krishnan, S., Alcock, K. J., Mercure, E., Leech, R., Barker, E., Karmiloff-Smith, A., & Dick, F. (2013). Articulating novel words: Children's oromotor skills predict nonword repetition abilities. *Journal of Speech, Language, and Hearing Research, 56*(6), 1800–1812. https://doi.org/10.1044/1092-4388(2013/12-0206)
LoBello, S. G. (1991). A short form of the Wechsler Preschool and Primary Scale of Intelligence–Revised. *Journal of School Psychology, 29*(3), 229–236. https://doi.org/10.1016/0022-4405(91)90004-B
Mahfoudhi, A., Everatt, J., Elbeheri, G., & Roshdy, M. (2020). Development and standardization of a phonological processing test in Arabic. *Arab Journal of Applied Linguistics, 5,* 1–24.
Malenky, A. (1997). *The development of morphological awareness in Hebrew speakers* [Master's thesis]. School of Education, Tel Aviv University, Israel.
Marshall, C. R., Ebbels, S., Harris, J., & van der Lely, H. K. J. (2002). Investigating the impact of prosodic complexity on the speech of children with specific language impairment. *UCL Working Papers in Linguistics, 14,* 43–68.
Marshall, C. R., & van der Lely, H. K. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia. *Language, 85*(1), 39–57. https://doi.org/10.1353/lan.0.0081
McCarthy, J. J. (1982). Nonlinear phonology: An overview. *GLOW Newsletter, 8,* 63–77.
McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry, 17,* 207–263.
Montgomery, J. W. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics, 16*(4), 355–378. https://doi.org/10.1017/S0142716400065991
Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11*(1), 77–91. https://doi.org/10.1044/1058-0360(2002/009)
Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders, 39*(1), 115–133. https://doi.org/10.1080/13682820310001616985
Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*(1), 61–78. https://doi.org/10.1044/1092-4388(2005/006)
Mustafawi, E. M. (2006). *An optimality theoretic approach to variable consonant alternations in Qatari Arabic* [Unpublished doctoral dissertation]. University of Ottawa, Ottawa, Canada.
Oetting, J. B., & Cleveland, L. H. (2006). The clinical utility of nonword repetition for children living in the rural south of the U.S. *Clinical Linguistics & Phonetics, 20,* 553–561. https://doi.org/10.1080/02699200500266455
Pigdon, L., Willmott, C., Reilly, S., Conti-Ramsden, G., & Morgan, A. T. (2020). What predicts nonword repetition performance? *Child Neuropsychology, 26,* 528–533.
Qafisheh, H. (1997). *NTC's Gulf Arabic–English dictionary*. NTC Publishing Group.
Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research, 47*(1), 223–234. https://doi.org/10.1044/1092-4388(2004/019)
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
Shaalan, S. (2009). Considerations for developing and adapting language and literacy assessments in Arabic-speaking countries.

In E. Grigorenko (Ed.), *Multicultural Psychoeducational Assessment* (pp. 287–314). Springer.

Shaalan, S. (2017). Reliability and validity of four Arabic language tests: A comparison of performance of Qatari school-aged children with and without language impairment. *Arab Journal of Applied Linguistics, 2,* 20–48.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords and phonological processes: Some comments on Gathercole, Willis, Emslie and Baddeley. *Applied Psycholinguistics, 12*(3), 369–373. https://doi.org/10.1017/S0142716400009279

Stokes, S. F., Wong, A. M., 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)

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multifactorial analysis. *Cognition, 179,* 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Taibah, N. J., & Haynes, C. W. (2011). Contributions of phonological processing skills to reading skills in Arabic-speaking children. *Reading and Writing, 24,* 1019–1042. https://doi.org/10.1007/s11145-010-9273-8

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford University Press.

Wechsler, D. (2002). *WPPSI-III: Wechsler Preschool and Primary Scale of Intelligence–Third Edition*. Psychological Corporation. https://doi.org/10.1037/t15177-000