The role of language and memory skills in sentence repetition in Arabic diglossia

Abstract

In diglossic Arabic there is a huge linguistic distance between StA and SpA (Saiegh-Haddad & Spolsky, 2014) and this creates a situation where words may be lexically and/or phonologically novel. Word can also encode novel morpho-syntactic units such as case/mood inflections. This linguistic distance impacts the establishment and access to phonological representations in long term memory (LTM) (Saiegh-Haddad & Haj, 2018) and phonological processing in working memory (WM) (Saiegh-Haddad & Ghawi-Dakwar, 2017) and is thus central to understanding language processing in Arabic (Saiegh-Haddad, 2018). The study aimed to test the contribution of language and memory skills to sentence repetition in Arabic among kindergarten children using a standard Arabic (StA) sentence repetition task. We focused on the accuracy of repetition of the individual words within sentences. Two questions were addressed: The first pertains to the errors that children make when they are asked to repeat simple sentences in StA. This question targeted linguistic distance and compared accuracy for identical, cognate and unique standard words; it also targeted novel and non-novel morpho-syntactic units including determiners, case/mood inflections and clitics. Error analysis was also conducted. The second question addressed the contribution of language and memory skills (vocabulary, language comprehension, memory) to repetition accuracy.

The results showed significant difference in the performance on sentence repetition according to stem type. Furthermore, a significant interaction was found between stem type and error type. In addition, correlations were found between language and working memory skills and sentence repetition task, where sentence repetition task was highly correlated with vocabulary skills.

First, the results support previous studies which shows that diglossia affect phonological representations of words in LTM (Saiegh-Haddad & Ghawi-Dakwar 2017). Secondly, the

results also support previous findings that the performance on sentence repetition task cannot rely only on working memory skills, however, sentence repetition task highly draws on linguistic abilities (Polišenská et al., 2015). To conclude, the present study highlights the sensitivity of sentence repetition task as a tool for assessing diglossic features in Arabic which may impact the acquisition of standard Arabic.

Background

Sentence repetition

Sentence repetition task (SRep) taps into a person's ability to recall and reproduce the exact wording from a previous hearing (Theodorou, Kambanarous, Grohmann, 2017). From one hand, this type of task is easy to administer but from the other hand it allows a controlled evaluation of specific target structures (Fraser, Bellugi, & Brown, 1963). Moreover, SRep is a complex task which involves syntactic knowledge, language processing, memory, and it requires auditory perception of stimuli and their reproduction (Leclercq, Quemart, Magis & Maillart, 2014; Marinis & Armon-Lotem, 2015). It is important to note that in SRep children repeat sentences rather than passive sequences of sounds (Polišenská et al., 2015). It is widely accepted that the way that a child repeats a sentence, particularly the changes he/she makes to the original model, can provide valuable information about the processing of the sentence (Karmiloff & Karmiloff-Smith 2001).

Sentence repetition task has been used to examine different developmental aspects of syntactic and morphosyntactic abilities (Kidd, Brandt, Lieven, & Tomasello, 2007; Komeili & Marshall, 2013; Devescovi & Caselli, 2007; Komeili & Marshall, 2013; Novogrodsky, Meir & Michael, 2018) in monolingual and bilingual children (Meir & Armon-Lotem, 2017; Tuller, Hamann, Chilla, Ferre, Morin, Prevost, Dos Santos, Abed – Ibrahim & Zebib, 2018) and it has been claimed to be a possible clinical marker for diagnosing developmental language disorder (Conti-Ramsden, Botting, & Faragher, 2001; Riches, 2012; Stokes, Wong, Fletcher, & Leonard, 2006; Leclercq et al., 2014). It was also used in different standardized batteries for language developmental assessments (Lee, 1971).

Syntactic complexity has been targeted in SRep as a measure of syntactic skills and research has indeed shown that more complex sentences in terms of syntactic structure are more difficult to process syntactically and are acquired later (Theodorou et al., 2017). Moreover, studies have shown that the sentence`s length (e.g. Devescovi & Caselli, 2007) and phonotactic frequency may also affect the performance of the children on SRep tasks (Coady, Evans, Kluender, 2010).

In the current study SRep is not used as a measure of syntactic ability, rather it is used as elicitation tool via which interference from the spoken dialect in children`s production of MSA sentences maybe tested. Therefore, even though the sentences targeted in the study were combined simple and complex syntactic structures, this syntactic property was not taken into account in our word level analysis of the repeated sentences.

The underlying mechanisms that tap into sentence repetition

There is a little consensus on the underlying mechanisms that are involved in SRep task. The central issue has been determining whether performance on sentence repetition mirrors linguistic knowledge (Klem, Melby-Lervag, Hagtvet, Lyster, Gustsfsson & Hulme, 2015; Pena & Bedore, 2020; Polišenská, Chiat & Roy, 2015) or memory capacity (Gathercole, 2005; Riches, 2012; Poll, Miller, Mainela-Arnold, Adams, Misra & Park, 2013).

Some researchers tested the association between memory skills, linguistic skills and SRep tasks and been argued that different mechanisms underlie the SRep performance depending on different factors (Klem at al., 2015). For example, Pena and Bedore (2020) examined the relationships between SRep, memory and lexical knowledge among TLD and DLD Spanish-English children between the ages of 6;10 and 9;11. They propose that SRep task differentially tap bilinguals' memory skills and lexical knowledge according to the language proficiency levels the child is exposed to. Pena and Bedore (2020) show different associations between memory and language skills and SRep performance for each group of TLD and DLD children depending on language proficiency. Their results indicate that TLD children rely more on linguistic skills rather than memory skills. Whereas in TLD children expressive vocabulary in English and Spanish explained 16% and 23% respectively of additional variance, non-word repetition (NWR) only explained 1% of additional variance. In contrary, among DLD children expressive vocabulary explained 15% of additional variance only in Spanish, whereas NWR in English accounted for 12% of additional variance but this effect was not evident in Spanish. Dosi and Koutsipetsidou (2019) highlight the impact of linguistic and cognitive abilities on SRep performance too for different measures; mainly accuracy and grammaticality. As a matter of fact, they found that cognitive and linguistic abilities impact differently these two measures of SRep. They tested 30 monolingual Greek-speaking DLD children and children with dyslexia with mean age of 8;3 years and they were tested on SRep task, verbal working memory (VWM) task and vocabulary. Their results give evidences that VWM predicts accuracy, while vocabulary knowledge predicts grammaticality.

Other stream of research also explored the involvement of different mechanisms that take part in SRep performance with emphasizing the role of linguistic skills rather than memory skills. The study of to Polišenskáet al., (2015) provides new insights into the interface between short term memory (STM) and long-term language knowledge. According to Polišenskáet al., (2015), long-term memory (LTM) storage is inextricably linked to repetition of sentences. In fact, they showed how linguistic knowledge helps determine the success rate of accurate repetitions (e.g., grammar, plausibility, prosody, and lexicality) in a SRep task among English-and Czech-speaking children between the ages 4- to 5-year-old emphasizing that there are relationships between STM and language that establish themselves in these years. The authors focused on immediate verbatim recall, by maintaining the just-heard phonological form and the linear sequence of the lexical items. According to Polišenskáet al., (2015) in a familiar language, verbatim recall is much more than a phonological recitation, as capacity varies with familiarity and knowledge of the material that should be recalled. Klem at al., (2015) also emphasizes the role of language abilities on the performance of SRep task. Thus, Klem et al. (2015) did not provide evidence for causality between working memory (WM) capacity and linguistic development and they claimed that SRep should be a measure of language skills and not to consider it as a separate element of memory that is related to language. Rather than, the authors found support for considering SRep as "a reflection of an underlying language ability factor rather than as a measure of a separate construct with a specific role in language processing" (Klem et al., 2015, p. 146). This conclusion come from the evidence that there were no continuing relationships between earlier SRep and later language abilities among 216 Norwegian children that were recruited from day-care centers. Due to SRep's satisfactory correlation with other measures of language ability, the authors believe that SRep should be regarded as a compound linguistic task that combine integration of different levels of language processing including grammatical processing but not only.

Okura and Lonsdale (2012) questioned whether SRep focuses on language proficiency or it is a mechanical repetition. The study was administered on 94 students studying English as a second language and they were tested on working memory task, and SRep task and English language proficiency test. The researchers assume that if the two constructs do not correlate, that means that they measure different skills. The results showed that there was no significance for the correlations between WM and SRep scores, or between WM scores and English language proficiency levels, but SRep scores and English language proficiency levels did reach significance. The absence of significant correlations between WM and English SRep scores and between WM and English proficiency levels, and the significant correlation between English SReps cores and English proficiency levels suggest that SRep task relies on language proficiency rather than on WM capacity. Similarly, Riches (2012) show that despite of various predictors that were linked to sentence repetition performance; yet, the best predictor for children with DLD was a structural priming task that addressed syntactic knowledge. Moreover, Riches proposes that phonological short-term memory plays a significant role in SRep for DLD rather than for TLD children.

Sentence repetition scoring

Different ways of scoring were used in SRep task according to the aim of the analysis. Some studies have focused on coding the sentence level (e.g. Taha, Stojanovik & Pagnamenta, 2021; Theodorou, Kambanaros & Grohmann, 2017; Marinis & Armon-Lotem, 2015), and others have focused on the word/morpheme level (e.g. Moll, Hulme, Nag & Snowling 2015; Marinis & Armon-Lotem, 2015).

Coding at the sentence level varied between from one study to the other. One common coding is binary scoring. In binary coding the child gets one score if he repeats the whole sentence accurately, and a score of 0 if he doesn't repeat the sentence as he heard it. Binary scoring could be limited since it doesn't reflect the number of errors were made by the child. This kind of coding refers to the number of errors the child makes when repeating the sentence. Score of 3 is given when the child repeats the sentence accurately, a score of 2 when the child makes one error, score of 1 for two to three errors and score of zero for a repetition with more than three errors (e.g Taha, Stojanovik & Pagnamenta, 2021; Theodorou, Kambanaros & Grohmann, 2017). Other types of scoring in the level of the sentence were directed to syntactic structure and to grammaticality. One score was assigned when the child preserved the syntactic structure of the sentence was given or when the sentence was grammatically correct (Marinis & Armon-Lotem, 2015; Taha, Stojanovik & Pagnamenta, 2021; Vinther, 2002).

However, the word-level analysis does not refer to the overall score of the sentence, but in this analysis the sentence is broken down into units of words and a score is given for each word within the sentence. In this way, it is possible to check how the child behaves at the level of each and every word and not just in its function in repeating the whole sentence (Moll, Hulme, Nag & Snowling 2015). Other word-level analyses are sometimes done intentionally for different types of words, for example content efficiency and function efficiency (Marinis & Armon-Lotem, 2015).

In the present study, morpheme-level analysis was conducted. Verbatim scoring for each morpheme unit in the sentence was coded; the stem unties, the inflectional morphological units and the morphosyntactic units. In the current study coding was related to diglossic features and children got a score of 1 if they had said the target unit or zero if they did not. Error analysis was conducted too for stem units, inflectional units and morpho-syntactic units.

Arabic language and diglossia

Arabic diglossia offers another natural setting in which within subject variations in extent of language experience on repetition ability may be tested. This is because native speakers in Arabic diglossia, and even the young ones among them, acquire two linguistic systems for two complementary sets of social functions: one for everyday speech and another for formal speech and writing. As a result, for most of the words they know, Arabic speakers store two phonological forms: one spoken/colloquial and another standard/written. Moreover, the two forms of many words in their lexicons may vary in one constituent phoneme, with the standard word embodying a standard novel phoneme that is not within the spoken variety of children or more (Saiegh-Haddad & Haj, 2018).

Relatedly, in Arabic diglossia, it is possible to tease apart phonological novelty from lexical novelty. Because words may have two different phonological forms, the lexical store of Arabic speaking children may be broken down into four types of words: (a) lexically and phonologically non-novel, (b) lexically non-novel but phonologically novel, (c) lexically novel but phonologically non-novel, and (d) lexically and phonologically novel (Saiegh-Haddad, 2004; Saiegh-Haddad and Spolsky, 2014). In turn, it is possible to test the independent contribution of lexical and phonological novelty to word repetition, and in our case to sentence repetition.

Diglossia: Impact on Language Processing

Arabic is a prototypical case of the concept diglossia as it was first outlined by Ferguson (1959): “a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety . . . . which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any section of the community for ordinary conversation” (p. 336). Local spoken vernaculars are collectively known as Spoken Arabic (or Colloquial Arabic). This variety is acquired naturally as a mother tongue. As opposed to modern standard Arabic (StA), which is the language of literacy tasks (reading and writing) and formal speech, with a strong focus on grammatically accurate reading and writing; It is a modern descendant of Classical Arabic and of Literary Arabic and is pretty uniform throughout the Arabic-speaking world. Therefore, once students enter school in the Arabic-speaking world, it is mandatory for them to learn Modern Standard Arabic as a language of reading and writing. The spoken interaction occurs, even within the classroom, in Spoken Arabic, or in the semi-standard Arabic known as Educated Spoken Arabic (Badawi, 1973), except perhaps for Arabic lessons, where Standard Arabic is dominant, at least in aspiration (Amara, 1995). Arabic is the native language of the vast majority of Palestinians in Israel, and most of them enroll in Arabic-medium schools from preschool through high school. In these schools, Arabic is the only language of instruction and textbooks, and all school subjects are taught exclusively in Arabic, including math and science. Beginning in grades three and four, Hebrew and English are both taught as second/foreign languages, respectively (Saiegh-Haddad and Everatt, 2017).

In spite of such a deceiving dichotomy, native Arabic speakers, including young children are actively engaged with both Spoken Arabic and Standard Arabic. They pray, do their homework, prepare for exams, and watch TV shows and dubbed series in Standard Arabic. As a result, linguistic development in Arabic requires, not just proficiency in spoken Arabic, but also in Standard Arabic2.

Since StA is the language of formal speech and reading/writing, it permeates the speech of many speakers (phonology, syntax, morphology, lexicon). Consequently, it is often difficult to distinguish between spoken and written norms. Although Ferguson proposes a distinction between the spoken and written varieties, he himself acknowledges it to be an abstraction. There are a variety of ways in which one can understand Arabic diglossia as ranging between colloquial/vernacular and literary/standard forms (Blanc, 1960; Badawi, 1973; Meiseles, 1980; Boussofara-Omar, 2006).

Standard Arabic and Spoken Arabic are phonologically and lexically distant (for a comprehensive discussion, see Saiegh-Haddad and Henkin-Roitfarb, 2014). Different Arabic-speaking regions might have different form of distance. Interestingly, neither Spoken Arabic nor Standard Arabic share the same exact set of phonemes, or lexical items (Maamouri, 1998). In the domain of phonology, Standard Arabic contains 28 consonantal phonemes in addition to six vowel phonemes: three short vowels: low /a/, high front /i/, and high back /u/, and three corresponding long vowels: /a:/, /i:/, and /u:/. Further, in Standard Arabic all syllables are preceded by a consonant (C), which serves as the onset and is followed by a vowel (V), which serves as the nucleus. However, this phonological structure varies from that of many varieties of Spoken Arabic which usually have a smaller set of consonants and a larger set of vowels. For instance, interdental consonants do not feature in many dialects of Palestinian Arabic spoken in the north of Israel. As a result, Cognate words, which are also used in these varieties of Spoken Arabic, acquire a different phonological form from that in StA, with StA interdental phonemes being replaced by corresponding phonemes in these varieties of Spoken Arabic (StA /8aPlab/; SpA /taPlab/ "fox"). In these dialects, the glottal stop phoneme is not preferred at the end of words, especially when preceded by a long vowel. Hence, cognate words ending in glottal stops often remove this phoneme and reduce the vowel preceding it (StA /sama:P/; SpA /sama/ "sky"). Lastly, consonantal clusters (in pausal non-inflected form), which are common in monosyllabic StA words, are avoided in these dialects and are usually broken by an epenthetic vowel (StA /bahr /; SpA /bahir / or /bahar/ "sea").

There is a wide lexical distance between Standard and Spoken Arabic. To measure the extent of this gap, SaieghHaddad and Spolsky (2014) analyzed 4,500 wordtypes derived from a pool of 17,500 words collected from 5-year-old native speakers of a local dialect of Palestinian Arabic spoken in Israel. This study showed that in the child's spoken lexicon, only 21.2% of the words were identical (e.g., /na:m/, "slept"; /daftar/, "notebook"), whereas, the remaining words were divided approximately evenly between common words, which are shared over the two varieties, but have partly overlapping phonological forms in each (e.g., SpA /dahab / vs. StA /Dahab/ "gold"), and unique SpA words, which have their own lexico-phonological forms completely different from those in StA (e.g., SpA /juzda:n/ vs. StA /haqi:ba / “bag”).

Language and literacy development is rarely studied in terms of diglossia, namely the linguistic distance between SpA and StA. Nevertheless, it receives increasing attention particularly within the framework of comparative linguistics and its effect on language development in bilingual and bilingual children (Rowe and Grohmann, 2013, 2014; Grohmann and

Kambanaros, 2016; Grohmann et al., 2016). Researchers tested the impact of the linguistic distance between Spoken and Standard Arabic on the development of literacy-related skills, including phonological awareness, pseudo-word decoding, and word reading, (Saiegh-Haddad, 2003, 2004, 2005, 2007; Saiegh-Haddad et al., 2011; Saiegh-Haddad and Schiff, 2016; Schiff and Saiegh-Haddad, 2017). The phonological distance between SpA and StA affects the development of literacy-related phonological skills in StA Arabic. Study results suggest that phonological distance between SpA and StA has an impact on literacy-related phonological skills in StA Arabic. For instance, Saiegh-Haddad (2003) compared children's phonological awareness for Spoken Arabic as against Standard Arabic phonemes and found that, even after children's production of StA phonology had normalized, children had more difficulty isolating StA than SpA phonemes. In addition, first graders found it challenging to decode pseudo words encoding letters that map to StA phonemes. It has been found that these effects, known as the Linguistic Affiliation Constraint (Saiegh-Haddad, 2017) persist across early elementary grades, surface equally strongly on both production and recognition tasks and have cross-dialectal validity (Saiegh-Haddad, 2007). Likewise, research has supported phonological distance's role in letter naming (Asaad and Eviatar, 2013) and in reading speed and accuracy (Saiegh-Haddad and Schiff, 2016; Schiff and Saiegh-Haddad, 2017).

Moreover, the Arabic orthographic system provides letters that represent Arabic consonants and long vowels. In addition, Diacritics are an optional part of the Arabic orthography system. There are two categories of diacritics based on function, distribution, and form. Phonemic diacritics are in the first group, consisting of five diacritics. They represent three short vowels (high in front and low in back), doubled consonants, and null vocalization in Arabic. They may appear on any of the letters of the word and provide semantic contrast phoneme information. Alternatively, the second type of diacritics is morphosyntactic. These diacritics are found only at the end of the stem and refer to abstract syntactic roles and properties (properties): the case of nouns (and adjectives) and the mood of verbs; they are retained in StA, but disappeared from all dialects of SpA (Maamouri, 1998). Case endings of definite nouns and modal endings of verbs are composed of three short Arabic vowels and are spelled with the same phonetic diacritics. Therefore, the case ending of an indefinite noun in a non-paused state is تنوين "nunation". They differ from other diacritical marks in pronunciation and spelling: waladun (nominative), waladan "a boy" (accusative) and waladin "a boy" (possessive). Generally, morphosyntactic diacritics appear at the end of a word. It is important to note that the main difference between morphosyntactical diacritics and phonemic diacritics is that the first set of diacritics is needed for vocabulary access as they represent semantically contrasting phonetic information, while the second set merely map grammatical patterns that are often not used for reading comprehension or word recognition.

The objective of the current study is twofold, first we want to investigate the underlying mechanisms that impact SRep in StA: oral language vs. memory skills. And second, to investigate the role of linguistic the distance in SRep in StA : lexical-phonological distance vs. morpho-syntactic distance.

To achieve the former objectives, the following questions have been asked:

1. What is the contribution of language vs. memory skills in SRep for the verbatim repetition of morphemes and for the syntactic score?
2. Is the verbatim repetition of morphemes affected by diglossia: phonological distance vs. lexical – phonological distance for free and bound morphemes?
3. Is the contribution of language vs. memory skills to morpheme repetition affected by diglossia?

~~The following two questions will be addressed in the study:~~

1. ~~what type of errors the children make when they are asked to repeat simple sentences in StA? This question targeted linguistic distance and compared accuracy for identical, cognate and unique standard words; it also targeted novel and non-novel morpho-syntactic units including determiners, case/mood inflections and clitics. Error analysis was also conducted.~~
2. ~~What is the contribution of language and memory skills: vocabulary, language comprehension and memory to repetition accuracy?~~

~~To answer these questions, we hypothesize:~~

1. ~~Children will be more accurate on identical words more than cognate and unique words.~~
2. ~~Error types will differ according to stem type. Phonological errors will be more evident in cognate words more than identical and unique words, and lexical errors will be more evident in unique words more than the two others. What about identical words??~~
3. ~~Types of errors will differ for morpho-syntactic units: determiners, case/mood inflections and clitics than stem words.~~
4. ~~There will be a positive relationship between sentence repetition task and language and memory skills (vocabulary, language comprehension, WM, PWM) on the accuracy performance of the children in the sentence repetition task.~~
5. ~~What is the unique contribution of the children's performance on the linguistic and working memory tasks to the Explained Variance of the performance on the sentence repetition task beyond the children's gender and age?~~

Methodology

Participants

The sample size was determined a priori by using G*power software. For ANOVA with repeated measures (within factors) analyses and for the test parameters: effect size = 0.20, α error = 0.05, power = 0.90 and correlation among repeated measures = 0.3, the total required sample size was 76 participants. For linear multiple regression analyses and for the test parameters: effect size $f^2 = 0.20$, α error = 0.05, power = 0.90, and number of predictors = 6. the total required sample size was 94 participants. In order to increase power and sensitivity, the present study comprised of 116 children (49 boys and 67 girls). The children's ages ranged from 58 to 71 months (*M* = 64.67, *SD* = 3.41). All children spoke the northern Palestinian dialect and were from low-middle SES class. This sample is a part of a larger sample of an intervention project which was conducted in 113 different Arabic speaking kindergartens in Israel and received approval of funding from the Ministry of Education in Israel. Parental consent was demanded for the children participation. All children who participated in the study were typically developing with no reported problems related to hearing loss, developmental delay or language difficulties.

Descriptive table - participants

??

Materials

SR Task

The current SRep task contains 15 sentences which were designed specifically for this study. The sentences were in modern standard Arabic (MSA) and they differed in their syntactic structure and syntactic complexity (simple, compound and relative clauses) and number of words. In addition, the sentences varied in their diglossic features where the diglossic features reflected the lexical-phonological and morpho-syntactic distance between the SpA and the StA varieties of the Arabic language. To be specific the sentences differed in the following features:

1. Lexical – phonological status of the word/morpheme: identical, cognate and unique words/morphemes were used in the sentences.

2. Morpho-syntactic diacritics: part of the words included morpho-syntactic diacritics and others did not. The distribution of the morpho-syntactic diacritics was different in each one of the sentences.

Internal consistency calculated using Cronbach's Alpha reliability was α = ??.

כאשר בודקים את העקיבות הפנימית ברמת ה units (160 יחידות שהיו במבחן) המהימנות עקיבות פנימית אלפא של קרונבאך גבוהה מאד והיא .955 אם בודקים את המהימנות ברמה של המשפט כולו שהילד התבקש לחזור עליו ולא ברמת היחידה של הניתוח אזי המהימנות ביצוע ברמה הגלובאלית הינה .76

The following table describes the total number of diglossic features that were applied in the sentence repetition task per morpheme:

Procedure for the SRep task

For this task, the children are requested to repeat the sentences exactly as said orally by the experimenter. The experimenter said each sentence in appropriate pace so that the children can hear each word of the sentence correctly. The experimenter recorded the repeated sentences on a form and each child was individually tested in a quiet room.

Transcription and coding

Coding rubric: Morpheme - based and diglossia-specific

Identical		**Cognate**		**Unique**		**Total units**
Words	Clitics	Words	Clitics	Words	Morpho-syntactic morphemes	160
25	13	55	28	12	27	

Sentences were coded and analysed for accuracy at the morpheme level to investigate the role of the lexical-phonological and morpho-syntactic distances between SpA and StA on the repetition of morphemes within sentences. All sentences for each child were transcribed by the experimenter who was Arabic native speaker. After the transcriptions, coding was conducted in two levels:

1. Syntactic scoring: we used a binary score of 1 or 0. Score one was assigned when the child had fully repeated the target syntactic structure, and a sore of zero was given when the child failed to repeat the exact syntactic structure.
2. Verbatim repetition: the verbatim repetition was accounted for the number of accurately repeated words per sentence. Thus, each correct repeated word in the sentence got one score. The maximum score for each sentence was its total words within the sentence.

Moreover, Error analysis was conducted for the free morpheme stems, and for the bound morphemes a distribution of errors was counted comparing identical, cognate and unique morphemes.

Vocabulary: receptive and productive

The vocabulary tests aimed to investigate the receptive and expressive vocabulary knowledge of the children. The two subtests included only nouns items which were divided into three categories: identical, cognate and unique words. The items differed in their frequency (high and low) based on a judge scale.

The receptive vocabulary test was designed after the PPVT test (ref.). The test consisted of 30 items; each item was paired with a picture together with 3 distractors on a PowerPoint Presentation slide. The experimenter would say the word and the child would point to the picture that matched the word pronounced by the experimenter.

The expressive vocabulary test consisted of 30 items, each item was presented in a picture on a PowerPoinr slide and the child was asked to name the picture. In case that the child gave the cognate or the unique word in spoken Arabic, he was encouraged to say it in MSA.

For both receptive and expressive vocabulary, one score was assigned for accurate performance and a zero score for inaccurate performance. No partial scores were assigned. Internal consistency calculated using Cronbach's Alpha reliability was α = 0.66 for receptive vocabulary and α = 0.79 for expressive vocabulary.

Comprehension: sentence comprehension and listening comprehension

These two tests tap into the comprehension abilities of children.

The sentence comprehension test was designed to examine specific forms of the Arabic grammatical structure (for example: formation of plural nouns, verb tenses, subject verb agreement, comparative form of adjectives, etc.). In this test we used 22/21? sentences in total, 17 sentences from the ALEF test (Arabic language: evaluation of function by … (), and 5 sentences were added in order to test complex structures such as relative clauses. Each item contained a sentence said by the experimenter and three pictures presented to the child: the correct picture matching the sentence and two distractors. The pictures were presented on a PowerPoint Presentation slide. Internal consistency calculated using Cronbach's Alpha reliability was α = 0.51.

The story listening comprehension test was based on the "Baby Birds" story from the MAIN project (Gagarina et al., 2014), the story script was translated into Standard Arabic and was approved by three judges. The child first heard a recorded story and after then the participants were asked questions about the story they heard. 10/13 ? questions were adapted from the MAIN project (Gagarina et al., 2014) and two questions were added by three judges. The questions from the MAIN included three questions targeting the goals, six questions elicit internal state terms and assess rational of the internal state term, and one question elicits Theory of mind/inferencing. The three questions that were added examine general comprehension of the total story. One score was assigned for accurate performance and a zero score for inaccurate performance. No partial scores were assigned. Internal consistency calculated using Cronbach's Alpha reliability was α = 0.66.

Working memory tasks

To measure working memory, the 'Digit Span' subtest followed by *Wechsler (1974)* was used. This test consisted of two subtests: forward and backward digit span in which children were asked to repeat increasingly longer strings (from 2 up to 5 digits) of random digits both forward and backward. Each digit span test consisted of one list, two trials for each string. The experimenter presented the lists verbally (one digit per second) in an increasing order and wrote down the children's answers in each trial.

In the backward digit span task, the children were asked to repeat the digits they heard in reversed order than they heard from the experimenter. The experimenter presented the lists verbally in an increasing longer strings manner (from 2 up to 5 digits) and wrote down the children's answers in each trial.

The score of the forward and backward digit span task was calculated from the total number of the correct answers from the overall trials. If a child failed in two consecutive trials at any one span size, the test was stopped (Hester, Kinsella, & Ong, 2004). Internal consistency calculated using Cronbach's Alpha reliability was α = ?? for both forward and backward digit span task.

Phonological working memory task

In order to test phonological working memory, an adapted version of pseudo word repetition was used (Saiegh-Haddad and Ghawi-Dakwar 2017). This task measures the capacity of the phonological loop. However, three variables were manipulated in this task: phoneme novelty (*/θ/, /ðˤ/, /ð/, /q/*), phonological structure novelty (StA structure vs. SpA strucure) and word length (1-4 syllables). The test included 24 items; 12 items were in StA and the other 12 item were in SpA form. Internal consistency calculated using Cronbach's Alpha reliability was α = ??.

General procedure

The tasks were administered individually in a quiet room at the children's schools in two sessions, each phase lasted for almost 45 minutes. The order of the tasks was intermixed; each answer given by each child was documented by the experimenter on a special form for each one of the tasks.

Results

Prior to examining the research hypotheses, in order to examine whether the study variables were normally distributed, Shapiro-Wilk tests were conducted. The study variables were the children's performances on the linguistic and working memory tasks and their performance on the sentence repetition task. The results indicated that the study deviated significantly from normal distribution ($p < .05$). Therefore, we conducted both non-parametric analyses and parametric analyses. We used the Friedman and Wilcoxon tests as non-parametric analyses. The Friedman test examined the differences between the three stem types (identical, cognate, unique) in the performance on the sentence repetition task. The Wilcoxon test examined the differences between two stem types (paired comparisons between the three stem types) and between two error types (phonological errors, lexical errors). The findings of the non-parametric analyses indicated the same significance level of the differences of the parametric analyses. Therefore, only the findings of the parametric analyses were presented and instead of reporting the mean or the sum ranks, the means and Standard Deviations as well as the range of the study variables were reported. Furthermore, we conducted Mauchly's test in order to examine the sphericity assumption in the Repeated measures ANOVA analyses. The result of the Mauchly's test indicated that the assumption of sphericity was rejected. Therefore, we reported the adjusted degree of freedom (df) in decimal numbers as well as the adjusted F-value.

In order to examine the first research question with regard to the differences in the children's performance on the sentence repetition task by stem type, we conducted one-way repeated measures ANOVA analysis. The independent variable was stem type (identical, cognate, unique) as the within subject factor. The dependent variable was the performance on the sentence repetition task. Significant difference in the performance on the sentence repetition task was found according to the stem type, $F(1.311, 150.730) = 583.87$, $p < .001$, $\eta_p^2 = .84$. Bonferroni analysis indicated that the performance on the sentence repetition task was significantly higher in the identical ($M = 88.57$, $SD = 9.87$) compared to cognate ($M = 73.48$, $SD = 12.44$) and unique ($M = 52.28$, $SD = 19.78$) stems ($p < .001$). Furthermore, the performance on the sentence repetition task was significantly higher in the cognate stem compared to the unique stem ($p < .001$).

In addition, we examined the effect of the stem type on the percentage of phonological and lexical errors in the sentence repetition task. In order to examine this question, two-way repeated measures ANOVA (3x2) was conducted. The independent variables were the stem type (identical, cognate, unique) and error type (phonological errors, lexical errors) as the within subject factors. The dependent variable was the percentage of errors in the sentence repetition task.

The main effects of stems type and error type were significant [$F(1.247, 146.214) = 664.55$, $p < .001$, $\eta_p^2 = .85$ and $F(1,230) = 255.50$, $p < .001$, $\eta_p^2 = .69$. Furthermore, the interaction of stem type and error type was significant, $F(1.271, 146.214) = 540.51$, $p < .001$, $\eta_p^2 = .83$. Table 1 presents the percentage of phonological and lexical errors in the sentence repetition task in each diglossia.

INSERT TABLE 1 ABOUT HERE

Comparisons between the two error types for each stem type, using paired samples t-test revealed that while the percentage of lexical errors in the sentence repetition task was significantly higher compared to the phonological errors in the identical and unique stems [$t(115) = 5.06$, $p < .001$, *Cohen's d* = 0.47 and $t(115) = 9.08$, $p < .001$, *Cohen's d* = 0.84, respectively], the percentage of phonological errors in the sentence repetition task was significantly higher compared to the lexical errors in the cognate stem [$t(115) = 23.33$, $p < .001$, *Cohen's d = 2.17* (see Figure 1).

INSERT FIGURE 1 ABOUT HERE

In the current study, we also counted the number of other errors in the sentence repetition task: determiner errors, morpho-syntactic diacritic error and clitics errors. We calculated the percentage of these other errors in the sentence repetition task by dividing the number of errors by the total number of units (160 units). Table 2 presents the percentage of other errors in the sentence repetition task.

INSERT TABLE 2 ABOUT HERE

Before examining our second research question with regard to the correlation between the performance on the linguistic and working memory tasks and the performance on the sentence repetition task, we calculated the mean, SD and the range of the performance on the linguistic and working memory tasks. Table 3 presents these descriptive statistics measures.

INSERT TABLE 3 ABOUT HERE

In order to examine the correlation between the performance on the linguistic and working memory tasks and the performance on the sentence repetition task, Pearson correlation analyses were conducted. Table 4 presents the Pearson correlation coefficients between the performance on the linguistic and working memory tasks and the performance on the sentence repetition task.

INSERT TABLE 4 ABOUT HERE

As can be seen in Table 4, as hypothesized, significant positive correlations were found between the performance on the linguistic and working memory tasks and the performance on the sentence repetition task. Only the correlations between the performance on the non-word repetition and working memory and the performance on the identical units in the sentence repetition task did not reach statistical significance [p = .129 and p = .090, respectively]. These results indicated that as the performance on the linguistic and working memory tasks increases, the performance on the sentence repetition task increases, respectively.

Finally, in order to examine our third research question with regard to the unique contribution of the children's performance on the linguistic and working memory tasks to the Explained Variance (EPV) of the performance on the sentence repetition task beyond the children's gender and age, four hierarchical regression analyses were conducted. One analysis was conducted for the performance on the whole task of the sentence repetition task. The other three analyses were conducted for the performance on each stem type in the sentence repetition task. The children's background characteristics were entered into the regression model in the first step. The children's performance on the linguistic and working memory tasks were entered in the second step in order to examine their unique contribution to the EPV of the performance on the sentence repetition task beyond the children's background characteristics. The variables in the second step were entered into the regression model in a stepwise manner so that only variables that have a significant contribution to the EPV of the performance on the sentence repetition task were entered into the regression model in the second step. The variables were entered by order of significance. In this manner, the probability of multicollinearity is likely to be decreased. Table 5 presents the results of hierarchical regression for the performance on the sentence repetition task by the children's gender, age and their performance on the linguistic and working memory tasks.

INSERT TABLE 5 ABOUT HERE

As can be seen in Table 5, the children's background characteristics did not contribute significantly to the EPV of the performance on the whole task and of the performance on each stem type in the sentence repetition task. The children's vocabulary score was found to be the most explanatory variable to the EPV of the performance on the whole task (43.3%) and of the performance on each stem type in the sentence repetition task (26.8%, 43.7% and 39.6% for the identical, cognate and unique units, respectively). The positive β coefficients indicated that as the children's vocabulary score increases, the performances on the whole task and on each stem type in the sentence repetition task increase, respectively.

The results also indicated a significant unique contribution of the children's performance on the working memory tasks. The children's performance on the working memory tasks contributed 4% to the EPV of the performance on the whole task and 3.6% and 5.4% to the EPV of the performance on the cognate and unique units in the sentence repetition task beyond the children's background characteristics and their scores on the vocabulary tasks. The positive β coefficients indicated that as the children’s performance on the working memory tasks increases, the performances on the whole task and on the cognate and unique units in the sentence repetition task increase, respectively.

Finally, the results indicated a significant unique contribution of the children’s performance on the language comprehension tasks. The children’s performance on the language comprehension tasks contributed 2.5% to the EPV of the performance on the whole task and 3.4% to the EPV of the performance on the unique units in the sentence repetition task beyond the children’s background characteristics, their scores on the vocabulary tasks and their scores on the working memory tasks. The positive β coefficients indicated that as the children’s performance on the language comprehension tasks increases, the performances on the whole task and on the unique units in the sentence repetition task increase, respectively.