

Arabic Diacritics and Their Role in Facilitating Reading Speed, Accuracy,

and Comprehension by English L2 Learners of Arabic

MOHAMMAD T. ALHAWARY¹ and ALI AL MIDHWAH²

¹*University of Michigan, Department of Middle East Studies, 202 S. Thayer Street, Suite 4111, Ann Arbor, MI 48104-1608 Email: abuamr@umich.edu*

²*King Saud University, Arabic Language Institute (ALI), Building 10 Riyadh, 11451, Saudi Arabia Email: aalmdwah@ksu.edu.sa*

<ABSTRACT>

The present study investigated the role of Arabic diacritics in word recognition and their impact on Arabic L2 learners' reading speed, accuracy, and comprehension at different stages of Arabic L2 acquisition. Fifty-four English L2 learners of Arabic from 3 proficiency levels (beginner, intermediate, and advanced) participated in the study. They belonged to 2 sets of groups: half with exposure to instructional materials containing diacritics, vowelized textbook (VT), and the other half with exposure to instructional materials not containing diacritics, un-vowelized textbook (UVT). Participants performed word-list reading, text reading, and target-word comprehension tasks under vowelized (V) and un-vowelized (UV) conditions. The results revealed participants in all VT groups did consistently better than their UVT counterparts. This positive role of diacritics in terms of Arabic word recognition and reading performance implies that inclusion of diacritics in words and texts within instructional materials does not only benefit Arabic L2 learners by removing ambiguity from words, but it also positively impacts their reading and pronunciation performance in general.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/modl.12642](https://doi.org/10.1111/modl.12642).

This article is protected by copyright. All rights reserved.

The study findings and pedagogical implications are also relevant to other languages, such as Farsi and Hebrew, which exhibit similar orthographic features.

<END ABSTRACT>

Keywords: Arabic L2; reading; recognition; diacritics; implicit and explicit learning; orthography; SLA

Arabic script without (short) vowel diacritics displays Arabic words written in a short form. To understand the implications for the presence or absence of Arabic diacritics as symbols representing internal short vowels, consider examples (1)-(2).

(1)

Without Diacritics	Pronunciation	Meaning
كتب	Ktb	ambiguous
كتب	Ktb	ambiguous
كتب	Ktb	ambiguous
كتب	Ktb	ambiguous

(2)

With Diacritics	Pronunciation	Meaning
كَتَبَ	Kataba	‘he wrote’
كُتُبَ	Kutub	‘books’
كُتِبَ	Kutiba	‘was written’
كَتَبَ	Katb	‘combining’

The four Arabic words in (1) represent deep orthography, reflecting word morphology (i.e., in particular root and underlying pattern morphemes). In absence of internal short vowels, all four words are identical in shape, consisting of three consonants *ktb* and nothing else. They are ambiguous when they occur in isolation both in terms of pronunciation and meaning. On the other hand, the Arabic words in (2) represent shallow orthography, reflecting the surface phonology of Arabic words, where, in presence of symbols for internal short vowels, each word is unambiguous phonologically and semantically. Even in their early learning of Arabic, Arabic L2 learners encounter a number of homographs which gradually increase as they progress in their exposure to Arabic. Example of early encountered homographs include: عِلْم *ʿilm* ‘he knew’ versus عِلْم *ʿilm* ‘science’ versus عَلَم ‘flag’; أَكَلَ *ʾakala* ‘he ate’ versus أَكْل *ʾakl* ‘food’; كَتَبَ *kataba* ‘he wrote’ versus كُتُبَ *kutub* ‘books’; كِتَاب *kitāb* ‘a book’ versus كُتَّاب *kuttāb* ‘writers’; and دَرَسَ *darasa* ‘he studied’ versus دَرَّسَ *darrasa* ‘he taught.’ Hence, diacritics disambiguate meaning and clarify the pronunciation of a given word. Even when words appear in context, diacritics make word recognition (including retrieval of the exact *meaning* of words and their proper *pronunciation*) more immediately apparent.¹

Current instructional practices of Arabic as an L2 are divided on which of the two options (i.e., suppliance or non-suppliance of short vowel symbols) constitutes best practices. Some believe that diacritics should be considered in teaching due to their utility in clarifying word pronunciation and meaning and facilitating reading development. Others believe that diacritics are not only useless but may also hinder learners’ progress in reading skill development. In essence, the difference between the two camps of supplying versus non-supplying short vowels throughout the visual input amounts to whether they believe diacritics can best be learned explicitly or implicitly. However, neither camp relies on empirical evidence to justify their stance, as no evidence one way or the other is available to date

Based on our observations as Arabic language practitioners and having used different textbooks in our Arabic language instruction, we have noticed a difference in students' learning and pronunciation when diacritics are supplied versus when they are not supplied. When confronted with words missing diacritics for short vowels, the Arabic L2 learner has no recourse but to guess what those vowels are, often unreliably so. Frequently, the rationale offered by opponents of the non-suppliance of diacritics is the need to get Arabic L2 learners to function as native Arabic speakers when dealing with authentic texts, since most (non-religious) authentic texts such as books and newspapers are displayed without short vowels. However, *educated* native speakers do go through a stage of exposure to diacritics at the very inception of their literacy development in their elementary school, which often extends to middle and high school (see Al Midhwah [2018] for a review of treatment of diacritics in native speakers' school textbooks in a number of Arabic speaking countries). It is not at all clear why native speakers (when they first start their schooling) need texts with diacritics for short vowels supplied (with the added advantage of having knowledge of the language/dialectal variety by the time they start their literacy development), whereas L2 learner do not. Therefore, providing empirical evidence of how L2 learners perform under short vowel suppliance versus non-suppliance condition is significant at this juncture to validate either stance, especially in absence of any clear research-based evidence one way or the other.²

<A>ARABIC L1 RESEARCH ON DIACRITICS

Most existing studies on diacritics (vowelization) were conducted on Arabic native speakers. Abu-Rabia (1996, 1997, 1999, 2001, 2007) conducted a series of studies to examine the impact of vowelization on word recognition and reading processes among low- and high-skilled readers, either as an individual factor or in combination with other variables, such as

context, text type, and reader type. Similarly, the reading of vowelized and un-vowelized isolated words and sentences was tested by Abu-Rabia and Siegel (1995) to determine differences in vowelized and un-vowelized isolated words between low- and high-skilled readers. The findings indicated that both levels of readers made fewer errors in isolated vowelized words compared to un-vowelized isolated words and participants at both skill levels improved their reading accuracy when reading words in context (i.e., in sentences). When vowels and context were combined, reading became optimal; conversely, when words were un-vowelized and isolated, the number of errors was highest, suggesting that diacritics (as well as context) facilitate word recognition as well as reading comprehension (e.g., Abu-Rabia, 1996, 1997, 1999).

Other studies produced mixed evidence (e.g., Abu-Hamour, Al-Hmouzb, & Kenanac 2013; Abu-Rabia, 1998, 1999; Taha, 2016) and found skilled and unskilled native Arabic readers read un-vowelized words more accurately than vowelized words. Similarly, in terms of reading speed, skilled readers read unvowelized words more quickly than vowelized words (e.g., Saiegh-Haddad & Schiff, 2016). Taha (2016) suggested that vowelization caused a visual load and could be considered redundant information for native speakers. Maroun and Hanley (2017) conducted two experiments to investigate whether diacritics improved the comprehension of all written words and whether the effects were confined to heterophonic homographs. The results of one experiment showed diacritics significantly increased the accuracy of the participants' semantic decisions about ambiguous words, but no effects for diacritics were found on unambiguous words. The same results were obtained in a follow-up experiment which relied on sentences rather than words and employed Arabic speakers who were university students and who conducted their reading primarily in English or French.

One limitation of the study was that the Arabic proficiency of the participants was not measured.

<A>ARABIC L2 RESEARCH ON DIACRITICS

In the Arabic L2 context, few studies were conducted on the role of Arabic script and diacritics. One such study was conducted by Khaldieh (2001) who investigated the role of knowledge of *ʔiʔrāb* ‘case and mood’ endings (i.e., inflectional endings involving vowels in final word position)³ and vocabulary on reading comprehension among American learners of Arabic as an L2. The study employed an expository text, vocabulary and *ʔiʔrāb* tasks, and an immediate recall protocol conducted in the learners’ native language. The findings revealed that participants relied on knowledge of vocabulary more than that of *ʔiʔrāb*. Khaldieh (2001) suggested further research, using texts employing classical Arabic prose, may provide evidence for the importance of *ʔiʔrāb* endings in reading comprehension according to text type.

Another Arabic L2 study was conducted by Hansen (2010) who examined the effects of suppliance and non-suppliance of internal short vowels on reading speed and reading comprehension and whether learners rely on knowledge of roots and patterns to compensate for lack of internal short vowel information. The study employed participants who were L1 speakers of Danish, English, and German at three proficiency levels in addition to a control group of native speakers. The results revealed that vowelization improved neither reading time nor reading comprehension for levels 1 and 2 learners and, in fact, vowelization seemed to slow reading speed when learners read voweled text in comparison to unvoweled text. Only when reading voweled text did learners’ reading speed improve significantly from level 2 to 3, while reading unvoweled text stalled progression after level 2. However, Hansen (2010) reported that learners at level 3 and native speakers read the two texts (voweled and

unvoweled) in approximately the same amount of time, which contradicted the finding of stalled progression after level 2.

Hansen's (2010) study has a number of limitations. In terms of reading speed, the study overlooked measuring the accuracy of participants' performance. Additionally, reading aloud would have provided more useful information about reading processes using vowelized or unvoweled text compared to silent reading. Regarding comprehension, Hansen (2010) used a single five-question multiple-choice task for each text and acknowledged that such a "test design, which use[d] only five questions in a multiple-choices task, [was] too narrow to allow for statistical measures" (p. 577). Other limitations included lack of control for the type of texts to which the participants were exposed during their formal instruction (i.e., textbook and classroom input). It is not clear whether the participants used fully vowelized, partly vowelized, or un-vowelized textbooks. Comparing two sets of participants (i.e., one group who learned from a vowelized textbook and another who learned from an un-vowelized textbook) would have allowed for a more sound research design (Alhawary, 2017). Finally, Hansen (2010) acknowledged another possible confounding variable in the study design:

it might have been because the texts created for this purpose are rather simple. When readers are proficient, and texts are easy, readers probably pay less attention to vowels because they rely on contextual clues. Thus, the added vowel information becomes redundant. (p. 578)

<A>IMPLICIT AND EXPLICIT LEARNING

The distinction between implicit and explicit learning and associated theoretical underpinnings may best characterize the difference between learning words with short vowel diacritics supplied versus learning words missing short vowel diacritics. Whereas implicit learning is assumed to take place naturally or incidentally without conscious attention ,

explicit learning is stipulated to occur with conscious attention (for a detailed review and rationale of the two modes of learning, see Ellis, 1994; Hulstijn, 2005). Thus, on one hand, presenting diacritics once upon first exposure to words in a vocabulary list (i.e., via written input) and withholding diacritics in all subsequent occurrences of words in subsequent drills and texts assumes the learner would pick up the proper pronunciation (i.e., retrieve the hidden vowels) from aural input alone. By teaching diacritics explicitly only initially and soon afterwards switching to teaching diacritics less explicitly (i.e., implicitly), this approach seems to assume that learning of diacritics takes place incidentally. On the other hand, supplying diacritics in words in vocabulary lists as well as in all subsequent drills and texts seems to maintain that learning of diacritics can take place only explicitly by offering the L2 learner more chances to notice and practice the forms in the written input (i.e., with focus on the input visually).

A number of studies investigated the effects of implicit versus explicit learning. DeKeyser (1995) found that explicit learning led to significantly greater short-term learning of simple artificial grammar rules than implicit learning and that implicit instruction of complex rules, when paired with a focus on meaning, resulted in some but not significant learning. Other studies based on semi-artificial and natural languages found evidence that meaning-focused learning can result in implicit knowledge (e.g., Robinson, 1995; Williams, 2004, 2005; Rebuschat and Williams, 2012; cf. Hama and Leow, 2010). For example, Rebuschat and Williams (2012) found adult L2 learners able to develop implicit knowledge of word order rules of a semi-artificial language following a meaning-oriented task.

In a recent study, Kim and Godfroid (2019) challenged the evidence presented in artificial and semi-artificial grammar learning studies (including Rebuschat and Williams, 2012) that auditory processing is more effective than visual-sequential processing for pattern learning and implicit knowledge development, in particular if such evidence is extended to

natural languages. Among other things, Kim and Godfroid (2019) examined the role of audio versus written modality in L2 grammar and implicit knowledge development in a study where English speakers engaged in meaning-focused practice on German word order rules. The study revealed participants gained explicit knowledge in both (aural and visual) modalities. However, only those participants who received natural written (visual) input exposure also developed robust implicit knowledge of word order rules. The authors accordingly conclude that “the findings illustrate the importance of stable visual input for developing L2 grammatical knowledge” (Kim and Godfroid, 2019, pp. 661-2). The rationale for the advantage of the visual mode over the aural mode in natural language is that “the untimed nature of the written mode affords repeated exposure to input, as determined by the reader, whereas the timed nature of the aural mode limits such opportunities” (p. 648).

If the present study is relevant to the explicit-implicit learning divide, the study findings will contribute further to this ongoing debate. If this is the case, two possibilities obtain. First, given previous research findings (e.g., DeKeyser, 1995), explicit learning of vowel diacritics is predicted to be more robust and have more advantages for learning, since diacritics do not involve complex grammatical rules (i.e., by maintaining their presence visually in the written input as well as aurally). Second, when comparing between the two (visual and aural) modalities with respect to vowel diacritic exposure (with and without, respectively) and if this applies to lexical learning, learning diacritics (when supplied in written, visual input) is predicted to develop more robust (implicit) knowledge of diacritics (as in Kim & Godfroid, 2019).

<A>RESEARCH QUESTIONS

As mentioned before, most studies relating to diacritics representing internal short vowels (a.k.a. *taškīl* ‘vowelization’) were conducted on Arabic L1 speakers and very few studies were conducted on Arabic L2 learners. Many studies in Arabic as an L1 support the claim

that diacritics facilitate word recognition and reading, though findings from other studies suggest the contrary. Studies in other languages exhibiting similar orthographic features, such as Hebrew and Farsi, reported a similar facilitative role for diacritics (Baluch, 1992; Schiff, 2012; Shimron & Siyan, 1994). In the Arabic L2 context, Khaldieh (2001) addressed the role of diacritics in terms of *ʔiʕrāb* ‘inflectional’ endings only, whereas Hansen’s study (2010) investigated the more crucial internal short vowels. Both claim that *ʔiʕrāb* endings (Khaldieh, 2001) and internal short vowels (Hansen, 2010) aid the Arabic L2 learner in neither reading speed nor reading comprehension. However, given the different foci of these studies as well as the apparent methodological limitations, further research with tighter control of relevant variables is needed to produce any conclusive evidence. The present study investigates the role of diacritics in word recognition and their impact on Arabic L2 learners’ reading speed, accuracy, and comprehension at different stages of Arabic L2 acquisition. It does so by comparing the performance of two types of Arabic L2 learners: those who were exposed to instructional materials containing diacritics, and those who were exposed to instructional materials not containing diacritics. The study aims to address the following questions:

RQ1. Do diacritics play a role in reading speed and accuracy of isolated words by learners who rely on vowelized textbooks versus those who rely on un-vowelized textbooks at different stages of Arabic L2 acquisition?

RQ2. Do diacritics play a role in reading speed and accuracy of words in context by learners who rely on vowelized textbooks versus those who rely on un-vowelized textbooks at different stages of Arabic L2 acquisition?

RQ3. Do diacritics play a role in reading comprehension by learners who rely on vowelized textbooks versus those who rely on un-vowelized textbooks at different stages of Arabic L2 acquisition?

<A>METHOD

Since the present study was mainly based on comparing the performance of two main groups of Arabic L2 learners and in order to address the research questions reliably, many criteria were considered in participant and program selection, most important of which is the suitability and compatibility of instructional input to which both groups of participants were exposed. It was also necessary to determine the appropriate content of vocabulary and texts for use in designing the study tasks in terms of prior exposure, length, and difficulty level.

Program Selection

Two Arabic language programs at two East Coast universities in the United States were selected as participant recruitment sites: Vowelized Text (VT) program and Unvowelized Text (UVT) program. Certain compatibility measures were taken into account. The two programs belonged to two large public research universities in the United States. Both programs were similar in size and offered instruction in Modern Standard Arabic (MSA) with focus on all four language skills along three stated proficiency levels: beginning, intermediate, and advanced. The length of the semester in both programs was 16 weeks. Crucially, each program adopted a distinct textbook series. The UVT program used *Al-Kitaab* series (Brustad, Al-Batal, & Al-Tonsi, 2004, 2007) which starts by presenting a list of new words at the beginning of each unit with diacritics and then drops diacritics from subsequent drills and texts. In other words, diacritics appear in the textbook only one time at the beginning of each unit. The VT program used *Ahlan wa Sahlan* series (Alosh & Clark 2013, 2015) which presents the new words as well as subsequent drills and texts with diacritics.

Participants

Fifty-four Arabic L2 learners along three proficiency levels (beginning, intermediate, and advanced) participated in the study. Half of them came from the UVT and the other half came from the VT program. Table 1 displays the demographic information of the participants. The participants were all undergraduate students except for five who were graduate (MA) students: three were from the UVT group (1 beginner, 1 intermediate, and 1 advanced) and 2 were from the VT group, 1 beginner and 1 intermediate. To control for prior language exposure and proper textbook exposure, the following types of learners were excluded: heritage learners, learners who had enrolled in another program and had used the other type of textbook or a mixed textbook, and learners who had lived or studied in an Arab country for 3 months or more.

<INSERT TABLE 1 ABOUT HERE>

TABLE 1
Participants

Group & Proficiency Level	Arabic Credits/ Weekly Hours	Gender F/M	Age Range	Age Mean
UVT				
Beginner (n = 12)	4	6/6	18/26	20.75
Intermediate (n = 10)	4	4/6	18/25	20.9
Advanced (n = 5)	4	3/2	19/26	20.6

VT				
Beginner (n = 12)	4	7/5	18/27	20.66
Intermediate (n = 10)	4	7/3	18/23	20.8
Advanced (n = 5)	4	1/4	20/23	21.2

Note. F/M = Total Females/Total Males.

Participants were initially placed in their respective proficiency levels according to their placement by their UVT and VT programs. The placement test had a listening comprehension and reading comprehension component. Teachers of prospective participants within their respective programs were also consulted on the exact proficiency level of each participant. Those who were deemed at a lower or higher level than their placement at the time of the study were excluded based on the teachers' assessment. The results indicated there was good homogeneity between the UVT and VT groups at each level, based on the mean scores and standard deviation (See Table 2).

< <INSERT TABLE 2 ABOUT HERE>

TABLE 2
Proficiency/Placement Mean Scores and Standard Deviations

LEVEL	GROUP	N	MEAN	STD. DEVIATION
Beginner	UVT	12	23.85	5.69
	VT	12	24.58	3.34
Intermediate	UVT	10	50.88	7.80

	VT	10	48.50	6.71
Advanced	UVT	5	66.50	9.72
	VT	5	67.92	9.01

Finally, to ascertain further homogeneity existed between the UVT and VT groups, two additional measures were followed to determine the quantity and quality of Arabic language use in and outside of the classroom. The first measure was use of a slightly modified version of the Language History Questionnaire (LHQ) (Li et al., 2014) which all participants completed. In particular, the LHQ included the two following questions pertaining to language use by participants in and outside of the classroom: (a) How often do you use the Arabic language in the classroom? and (b) How often do you use the Arabic language with people outside the classroom? The responses, which were based on a Likert 1–7-point scale (never, rarely, sometimes, regularly, often, usually, and always, respectively), suggest the two groups were homogenous, though the UVT groups seem to have had slightly more exposure to Arabic input than their VT counterparts (see Table 3). To obtain more specific responses, a third relevant question was included in the LHQ: On average, how many hours per day do you spend reading in the Arabic language? The responses showed that all the respondents in both groups spent less than 3 hours reading in Arabic outside the classroom. Thus, participants' responses to the LHQ questions indicate that both groups were homogeneous overall in terms of Arabic language use in and outside of the classroom.

<INSERT TABLE 3 ABOUT HERE>

TABLE 3
Means of Using Arabic in the Classroom and Outside the Classroom

Level	Using Arabic in the Classroom		Using Arabic Outside the Classroom	
	UVT	VT	UVT	VT
Beginner (n = 12)	5	4.3	2.2	1.7
Intermediate (n = 10)	5	4.5	2	2.1
Advanced (n = 5)	5.4	5.2	2.2	2

The second measure for determining participants' language use was in the form of classroom visitations to determine the extent to which each group was different from the other in particular in receiving implicit versus explicit instruction on diacritics. Based on the outcome of the visits, on one hand, the UVT teachers were observed to not place emphasis on or enforce use of diacritics in their teaching. They all wrote words for their students on the board without diacritics, except for one teacher who wrote some words (though not all) with *šadda* 'geminate consonant symbol' such as the words *مدّه* 'period,' *مطاط* 'rubber,' and *الله* 'God.' The VT teachers, on the other hand, seemed to have placed emphasis on and enforced use of diacritics. They wrote words with short vowel diacritics on the board consistently, more so in the beginner and intermediate levels than in the advanced level. Overall, instructors in the UVT group consistently enforced use of diacritics with their students including in words containing a *šadda* 'geminate consonant symbol' (i.e., placing a

short vowel with geminate symbol). This means that teachers in the UVT group relied on implicit instruction of diacritics (i.e., beyond the initial presentation of new vocabulary items) whereas teachers in the VT group followed a more explicit instruction of diacritics throughout.

Materials

Three lists of isolated Arabic words were designed for this study, one for each proficiency level, and each containing words vowelized and un-vowelized. Six texts were designed, two for each level: one vowelized and one un-vowelized. The lists of words served as the experimental materials for two isolated word reading tasks. The same words were then used as target words in context (i.e., within texts) for two text reading tasks. Comprehension questions on the target words in context were also included in two comprehension tasks. A number of criteria were taken into account in the selection of the word lists and construction of texts, such as difficulty level, comparability of prior exposure, and length as elaborated subsequently.

<C>Word Lists. To neutralize effects that may result from the words' difficulty levels, all target words of the word lists were extracted from the participants' textbook lessons according to their level. To control for comparability of prior exposure by the two groups, the frequency of occurrence of each word was considered. Studies on L2 vocabulary learning have shown different numbers for the minimum level of frequency that an L2 learner needs in order to learn a new word. Some studies suggested that a new word needs to occur 6 to 12 times to be acquired (Al-Batal, 2006; Ryding, 2013; Saragi, Nation, & Meister, 1978). Following this cut off criterion, target words comprised words to which participants had been exposed fewer than 12 times, to control for high-frequency exposure effects on word recognition.⁴

To apply the preceding criterion a number of steps were followed. First, the target words of the study were extracted from the last few lessons ending with the target lesson (i.e., the lesson during which the study took place) at the end of the Spring (second) semester of 2017 in both the UVT and VT groups, as listed in Table 4. The study was conducted at the end of the second semester of the school year. In particular, lists of new vocabulary that were given in each of the target lessons for each level in both textbooks were inserted in Excel tables. Then common words between the UVT and VT column sets were extracted for each level, beginner, intermediate, and advanced. Subsequently, the frequency of each common extracted word at each level was counted in each textbook to confirm that students were exposed to the words fewer than 12 times and ensure that both groups were exposed to them almost equally. The final step was determining the type of each extracted word in terms of whether it was homographic or non-homographic. This was done in two phases; the first phase was determining the original type of word in Arabic, while the second was deciding whether this word is homographic or non-homographic based on the learners' knowledge. For example, the original word شعر in Arabic is a homographic word because it can be شَعَرَ *šaḥar* 'he felt,' شِعْر *šīr* 'poetry,' or شَعْر *šar* 'hair' as determined by the diacritic; at the same time, it can be a non-homographic word, based on the L2 learners' knowledge, if the learner only learned one pronunciation and meaning of this word in their textbook such as شَعَرَ *šaḥar* 'he felt.'⁵

<INSERT TABLE 4 ABOUT HERE>

TABLE 4

Target Lessons and Target Words

	VT (Alosh & Clark 2010)	UVT (Brustad et al. 2004)
Beginner Level		
Lesson participants were at during the study	12	15
Total of final common words (used in the study)	17	17
Intermediate Level		
Lesson participants were at during study	24	3
Total of final common words (used in the study)	18	18
Advanced Level		
Lesson participants were at during the study	16	10
Total of final common words (used in the study)	22	22

Accordingly, seventeen common words were extracted from the UVT and VT for the beginner level; they appeared almost equally in the un-vowelized and vowelized beginner-level textbooks. The range of frequencies of the extracted words at this level was 1 to 12. Based on the original type of these words in Arabic, there were 13 homographic words and 4 non-homographic words. However, based on the learners' L2 knowledge in both groups, the words comprised 5 homographic and 12 non-homographic words. Examples of homographic words at this level include عَلِمَ *ʿalim* 'he knew,' عِلْمٌ *ʿilm* 'science,' أَكَلَ *ʾakal* 'he ate,' أَكْلٌ *ʾakl* 'food' and non-homographic words يَعْلَمُ *yaʿlam* 'he knows,' بِطَاقَةٌ *bitʿāqah* 'card,' and مَجَلَّةٌ *madzallah* 'magazine' (see Online Supporting Information Appendix A).

The list of common extracted words at the intermediate level contained 18 words, which exhibited almost equal occurrence in the intermediate lessons in the UVT and VT

textbooks. The frequency of these words ranged from 1 to 10. The words were divided into the homographic and non-homographic types based on the original type in Arabic. Originally, there were 14 homographic and 4 non-homographic words; however, based on the participants' L2 knowledge, the list included 7 homographic and 11 non-homographic words. Examples of homographic words at this level include جَمَعَ *dzamaʕ* 'he combined/gathered,' جَمَعَ *dzamʕ* 'plural,' شَعَرَ *šaʕar* 'he felt,' *šaʕr* 'hair' and non-homographic words مُهَنْدِس *muhandis* 'engineer' and عِدَّة *ʕiddah* 'several' (see Online Supporting Information Appendix B).

For the advanced level, 22 common words were extracted; these words exhibited almost equal occurrence in the advanced lessons in the UVT and VT. The range of frequency of these words in each textbook was 1 to 7. The words were divided into the homographic and non-homographic types based on their original types in Arabic. Originally, there were 14 homographic words and 8 non-homographic words, and the same numbers were found based on the participants' L2 knowledge. Examples of homographic words at this level include حَزِنَ *hazin* 'he saddened,' حُزْن *huzn* 'sadness,' مَثَّلَ *maθθal* 'he represented,' مِثْل *miθl* 'like' and non-homographic words مَرْحَلَةٌ *marḥalah* 'stage' and مَرْكَز *markaz* 'center' (see Online Supporting Information Appendix C).

<C>Texts. In designing the six texts (two for each level, one vowelized and the other un-vowelized), many criteria were considered to ensure that all the texts were suitable for participants' levels in terms of difficulty level, length, and prior exposure. Difficulty level was taken into account by relying on (a) participants' proficiency levels corresponding to ACTFL (2012) rubrics/descriptions of functional reading abilities, and (b) participants' familiarity with topics through exposure via their textbooks. According to participants' scores on the proficiency/placement test (Table 2), the three groups placed at end of the beginner, end of the intermediate, and advanced levels, corresponding roughly to ACTFL's

intermediate low, intermediate high, and advanced low, respectively. According to ACTFL (2012) standards, readers at the intermediate low level are able “to understand some information from simple connected texts dealing with a limited number of personal and social needs.” Accordingly, the two beginner texts were on topics related to normal daily life activities and had already been discussed in the participants’ textbooks (see Online Supporting Information Appendices D-E). ACTFL (2012) rubrics for the intermediate high level state that readers can understand texts (including description and narration) related to personal and social topics based on their own interests and knowledge from their textbooks. Accordingly, the two texts designed for the intermediate groups were descriptive and narrative texts that related to “reading” and “holidays” (see Online Supporting Information Appendices F-G). As for ACTFL (2012) standards for the advanced low, readers can deal with topics that are new to them, can comprehend the main idea and supporting details of narrative and descriptive texts related to real-world topics, and can fill gaps in their lexical and structural knowledge by using contextual clues and background knowledge. Hence, the two texts chosen for this level related to the United Nations and to the International Arabic Language Day (see Online Supporting Information Appendices H-I).

Length of texts was also taken into account according to participants’ proficiency level. The two beginner texts were the shortest. Both comprised 120 words (one vowelized containing 774 characters and the other unvowelized containing 485 characters). The two intermediate texts were longer, each consisting of 141 words (one vowelized containing 972 characters and the other unvowelized text containing 653 characters). The two advanced texts were the longest: one consisting of 150 words (for the vowelized text containing 1067 characters) and the other comprising 149 words (for the unvowelized text containing 735 characters).⁶

More crucially, the target words selected for the word lists were used in the texts in the respective levels. As previously discussed, the target words were extracted from the participants' UVT and VT textbooks based mainly on comparability of participants' exposure in the two textbooks. Other (non-target) words which acted as fillers in the texts were selected based on general ACTFL proficiency guidelines concerning the range of vocabulary expected in each level; an attempt was also made to avoid biasing either group in terms of prior textbook exposure of such vocabulary items.

<C>Comprehension Questions. In addition to the word lists and texts, six sets of comprehension questions were included in the study, two for each level. The questions were related to the target words in each text, and participants were asked to write (in English) the correct meaning of each word based on its occurrence in the text (see Online Supporting Information Appendices J-N).

The foregoing task contents were further assessed by three professional instructors of Arabic as a foreign language whose suggestions for slight modifications were incorporated. In addition, the materials were piloted before they were used in the study.⁷

Data Collection Procedure

The study was conducted in a quiet, private room. The participants sat in front of a laptop (with a 15-inch screen), received detailed oral instructions for the first reading task, and then practiced on four words. Each participant was requested to wear a headset attached to the laptop, and then the list of isolated words was presented under two conditions (vowelized and un-vowelized) randomly. Each participant was asked to read each word aloud. Each word appeared on a slide individually, and was shown twice—once with diacritics and once without—at random. Randomization occurred at two levels: (a) with respect to the order of all words in the list in general, and (b) whether a word appears with diacritics or without. The

words appeared in Lotus Arabic Linotype font (size 18) with a black text on a white background. After finishing the first reading task of isolated words, each participant was requested to move to another seat, and he/she was given detailed oral instructions about the next task which consisted of reading two types of (un-vowelized and vowelized) texts. The participants were asked to read out loud the two texts separately. Half of them started with the vowelized text and the other half with the un-vowelized text. Each text was printed on one page (Lotus Linotype font; size 18; black color on a white background). All readings were recorded using Audacity software. After having read the two texts, participants were given a sheet with comprehension questions and a new copy of the first text in which the target words were highlighted. Participants were asked to write the meaning of each target word based on its position in the text. The time for completing this part of the task was limited to 15 minutes. The same procedure was repeated for the second text.⁸

Data Analysis and Coding

To detect the statistical significance of certain factors (i.e., the textbook group, conditions, and interaction between group and conditions), a two-way repeated-measures ANOVA and linear mixed model were run in SPSS software. The results were obtained for both the vowelized un-vowelized condition.

To determine word recognition (i.e., word reading speed and word accuracy) in the isolated word reading task, all audio recordings of words were transcribed using ELAN software, and then the duration of recognition for each word was measured using ELAN and PRAAT. The duration of word recognition was measured from the word's initial appearance until the participant stopped reading the word in milliseconds.⁹

The word recognition accuracy in the isolated word reading task was evaluated in terms of two main scales, namely a general scale and a detailed scale. The general scale used the following criteria: correct pronunciation of the word = 1 point, incorrect pronunciation = 0 points, and incorrect followed by correct pronunciation (i.e., hesitation) = 0.5 points. As for the detailed scale, each word was divided according to the number of characters and a following short vowel/diacritic (or lack thereof) in each word. The final short vowel (following the last character/consonant) of each word was excluded, as this position represents a grammatical case or mood marking in Arabic and case and mood markings are not among the target forms of the present study. The word accuracy percentage was measured using the following equation:

$$\frac{(\text{Number of characters in the word} - 1) - (\text{Total number of character errors}) \times 100}{(\text{Total numbers of characters in the word} - 1)}$$

For example, the word **يَبْلُغُ** *yabluy* ‘reaches’ has three parts, namely **يَ** [ya], **بُ** [b], and **لُ** [lu]; the last part, **غُ** [ɣ] was not calculated, because the vowel following it represents mood ending, not an internal diacritic. Thus, the scale of this word is 3. Accordingly, if participants pronounced this word with a pronunciation error in a vowel following a character, such as **يَبْلِيغُ** *yabliy*, the accuracy of this word in the detailed scale would be 66.66% (see also Figure 1).

<INSERT FIGURE 1 ABOUT HERE>

FIGURE 1

Example of the Detailed Accuracy Scale

Word	س	ل	ر	ر	Detailed scale accuracy equation	Accuracy
Scale	س (1)	ل (1)	ر (1)	ر (1)	$\frac{(4 - 1) - (\text{errors}) * 100}{3}$	100%
No errors	—	✓ ل (1)	✓ ر (1)	✓ ر (1)	$\frac{(4 - 1) - (0) * 100}{3}$	100%
One error	—	✗ ل (0)	✓ ر (1)	✓ ر (1)	$\frac{(4 - 1) - (1) * 100}{3}$	66.66%

To address the speed of the text reading tasks, each reading was recorded and analyzed in PRAAT, which was used to measure the duration of each text reading (from when the participant started reading the text until he/she stopped reading) in milliseconds. In terms of text reading accuracy there were two accuracy measures that were adopted, namely the accuracy of target words in each text and accuracy of the whole text. The accuracy of target words was determined based on the two accuracy scales (the general scale and detailed scale) used in the previous reading word list tasks. The whole text accuracy measure relied on a general scale which scored the correct pronunciation of each word in the text with 1 point, incorrect pronunciation with 0 points, and incorrect followed by correct pronunciation (i.e., hesitation) with 0.5 points. This scale was used to obtain a broad understanding of the effect

of diacritics on the reading accuracy of texts, though only the target words are tightly controlled for comparability of exposure by the two groups while other words occurring in the texts are not. To measure the final accuracy score of the whole text, the following equation was used:

$$\frac{(\text{Total number of text words} - \text{Total number of errors in text}) \times 100}{\text{Total number of text words}}$$

Finally, the comprehension questions were related to the target words in each text, and the participants at each level were asked to write the correct meaning of each word based on its position in the text. Then, their scores were measured for each question using the following scale: correct answer = 1 point, incorrect answer = 0 points.

<A>RESULTS

Isolated Word Reading Task

The analysis of isolated word reading speed indicated that participants of the VT group significantly read isolated words at a faster speed than participants in the corresponding UVT group in all proficiency levels (Table 5). Furthermore, the results showed that the performance of participants in all proficiency levels of the VT group was nearly stable under both V and UV conditions, whereas the performance of participants in all proficiency levels of the UVT group was unstable and their reading speed under the V condition was considerably slower than that under the UV condition (see also Figure 2). An additional observation here is that overall the existence of vowel diacritics seems to cause participants of both groups to take more time to pronounce words. However, the difference in the performance of the VT group on vowelized versus non-vowelized words is much smaller,

than that of their UVT counterparts, and is almost negligible (see Online Supporting Information Table 1).

<INSERT TABLE 5 ABOUT HERE>

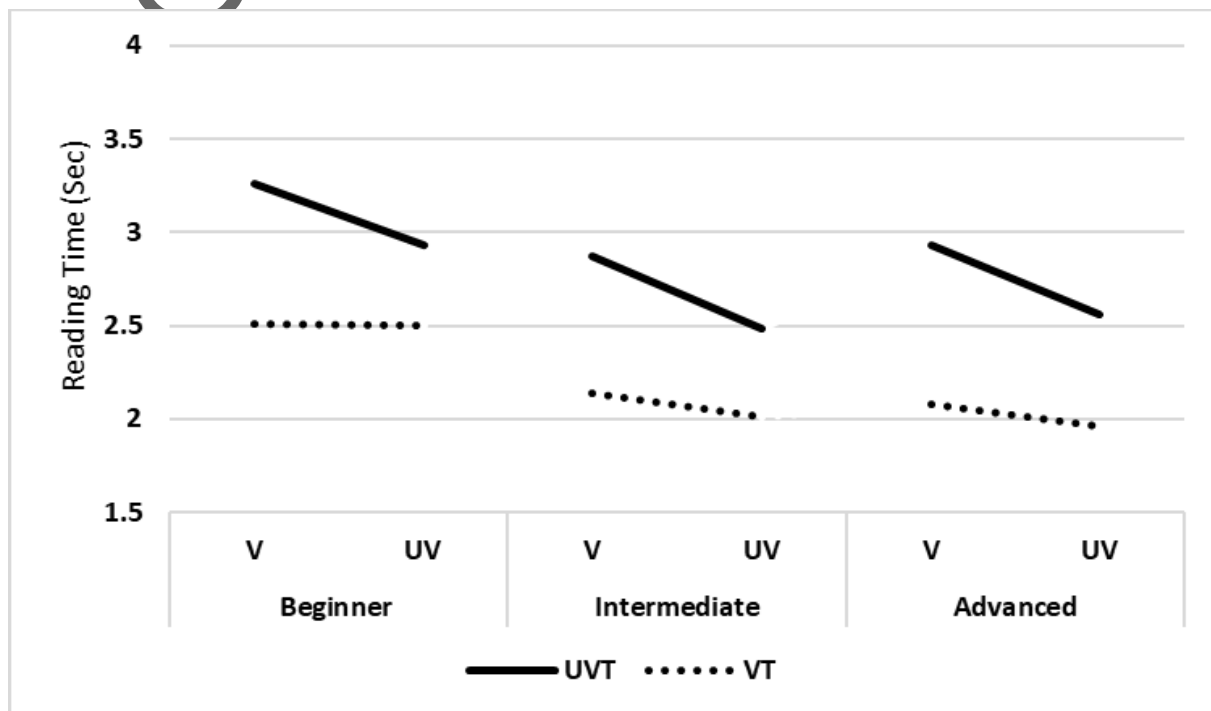
TABLE 5
Reading Time of Isolated Words – Test Effects

Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	11.544	.003
	Condition (V vs UV)	1	22	6.010	.023
	Group * Condition	1	22	5.579	.027
Intermediate	Groups (VT vs UVT)	1	18	13.965	.002
	Condition (V vs UV)	1	18	19.547	.000
	Group * Condition	1	18	4.696	.044
Advanced	Groups (VT vs UVT)	1	8	6.334	.036
	Condition (V vs UV)	1	8	20.464	.002
	Group * Condition	1	8	5.026	.055

<INSERT FIGURE 2 ABOUT HERE>

FIGURE 2

Reading Time of Isolated Words at Three Proficiency Levels



Based on the detailed accuracy scale of the participants' performance in the isolated word reading task, the result showed a remarkable advantage of the VT participants over their UVT counterparts, especially in the beginner and intermediate VT groups (a similar pattern obtained using the general accuracy scale). The findings also confirm that participants in the VT group showed a stable performance under both V and UV conditions, whereas participants in the UVT group showed an unstable performance under both conditions (see Table 6; Figure 3; see also Online Supporting Information Table 2).

<INSERT TABLE 6 ABOUT HERE>

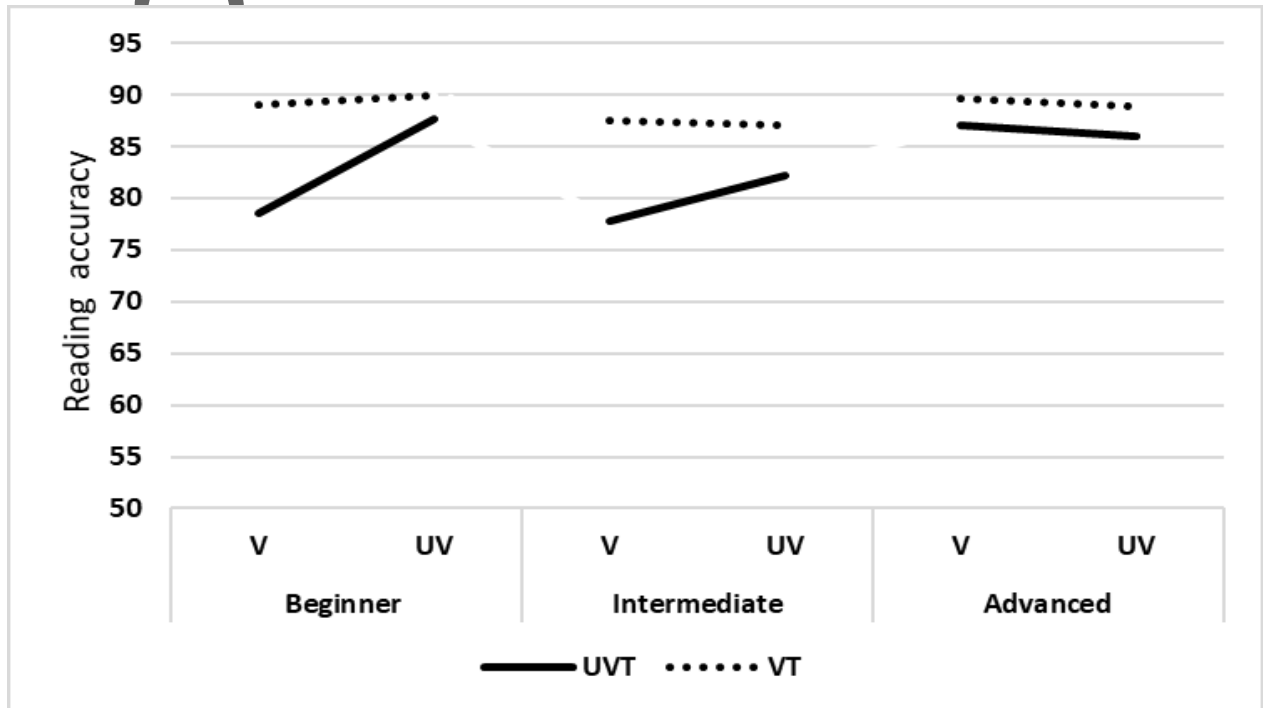
TABLE 6
Reading Accuracy of Isolated Words - Tests effects (Detailed Scale)

Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	4.205	0.052
	Condition (V vs UV)	1	22	11.094	0.003
	Group * Condition	1	22	7.201	0.014
Intermediate	Groups (VT vs UVT)	1	18	4.340	0.052
	Condition (V vs UV)	1	18	5.196	0.035
	Group * Condition	1	18	7.713	0.012
Advanced	Groups (VT vs UVT)	1	8	0.615	0.455
	Condition (V vs UV)	1	8	7.264	0.027
	Group * Condition	1	8	0.308	0.594

<INSERT FIGURE 3 ABOUT HERE>

FIGURE 3

Reading Accuracy of Isolated Words at Three Proficiency Levels (Detailed Scale)



Text Reading Tasks

Results for reading speed of texts showed that participants in the beginner and intermediate VT groups read texts at a significantly faster speed than their counterpart UVT groups. Moreover, the advanced VT group maintained an advantage of reading speed over their UVT counterparts. Additionally, the result also showed that participants in the VT group maintained a stable reading speed under both V and UV conditions. The difference between their reading speed under both conditions was small. However, the UVT group exhibited an unstable performance since the difference between their reading speed under both the V and UV conditions was considerably wide in all proficiency levels (see Table 7; Figure 4; see also Online Supporting Information Table 3).

<INSERT TABLE 7 ABOUT HERE>

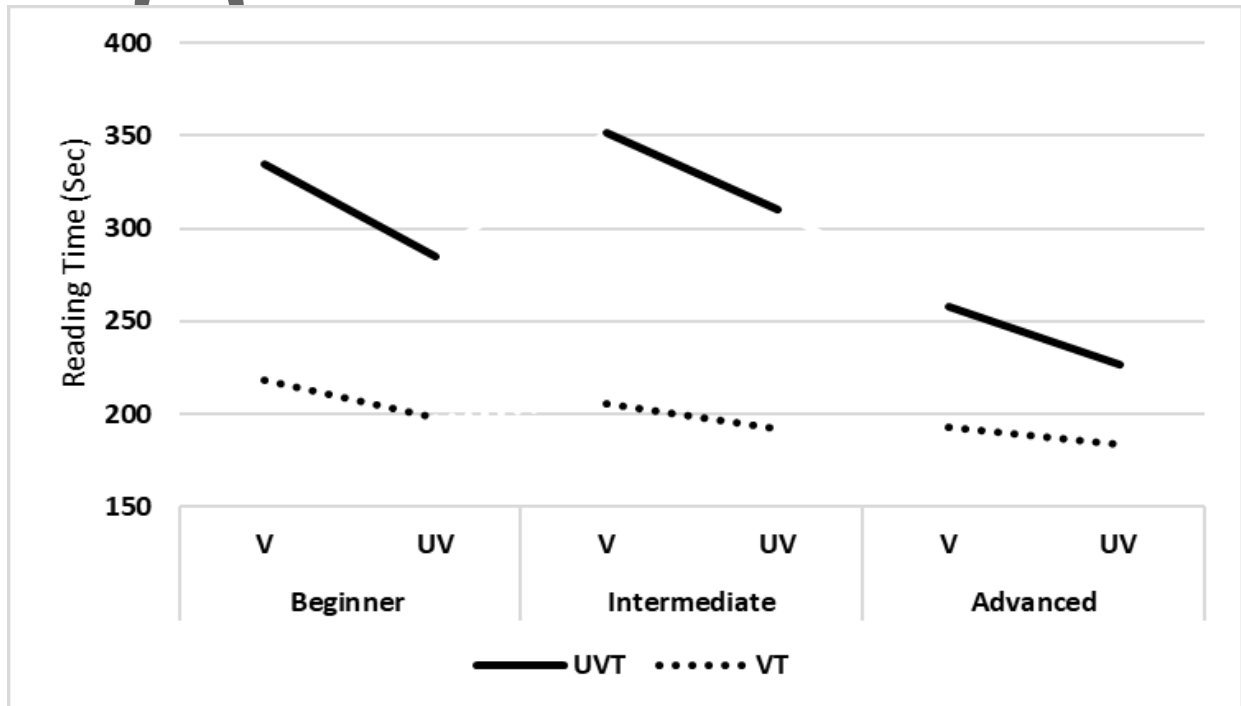
TABLE 7

Reading Time of Texts – Tests Effects

Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	19.474	.000
	Condition (V vs UV)	1	22	25.488	.000
	Group * Condition	1	22	4.669	.042
Intermediate	Groups (VT vs UVT)	1	18	29.425	.000
	Condition (V vs UV)	1	18	17.833	.001
	Group * Condition	1	18	4.723	.043
Advanced	Groups (VT vs UVT)	1	8	1.443	.264
	Condition (V vs UV)	1	8	23.139	.001
	Group * Condition	1	8	6.901	.030

<INSERT FIGURE 4 ABOUT HERE>

FIGURE 4
Reading Time of Texts at Three Proficiency Levels



Although as in the word list reading tasks, both groups took understandably more time reading vowelized texts perhaps reflecting their conscious attention to pronounce texts accurately, the difference in the performance of the VT group on vowelized versus non-vowelized texts is much smaller, than that of their UVT counterparts, here too.

Accuracy in the text reading tasks was measured in two ways: (a) target word accuracy, and (b) whole-text accuracy. Target word accuracy was analyzed by using the two main (detailed and general) accuracy scales used in the previous isolated word task. Results of target word accuracy obtained using the detailed scale showed that participants in both the UVT and VT groups were more accurate in reading V target words than UV target words. However, participants in the VT group were more accurate in reading target words than participants in the UVT group (a similar pattern of findings emerged using the general scale).

Furthermore, the differences between UVT and VT groups were significant among participants at the beginner and intermediate proficiency levels (see Table 8; Figure 5; see also Online Supporting Information Table 4).

<INSERT TABLE 8 ABOUT HERE>

TABLE 8
Reading Accuracy of Target Words - Tests Effects (Detailed Scale)

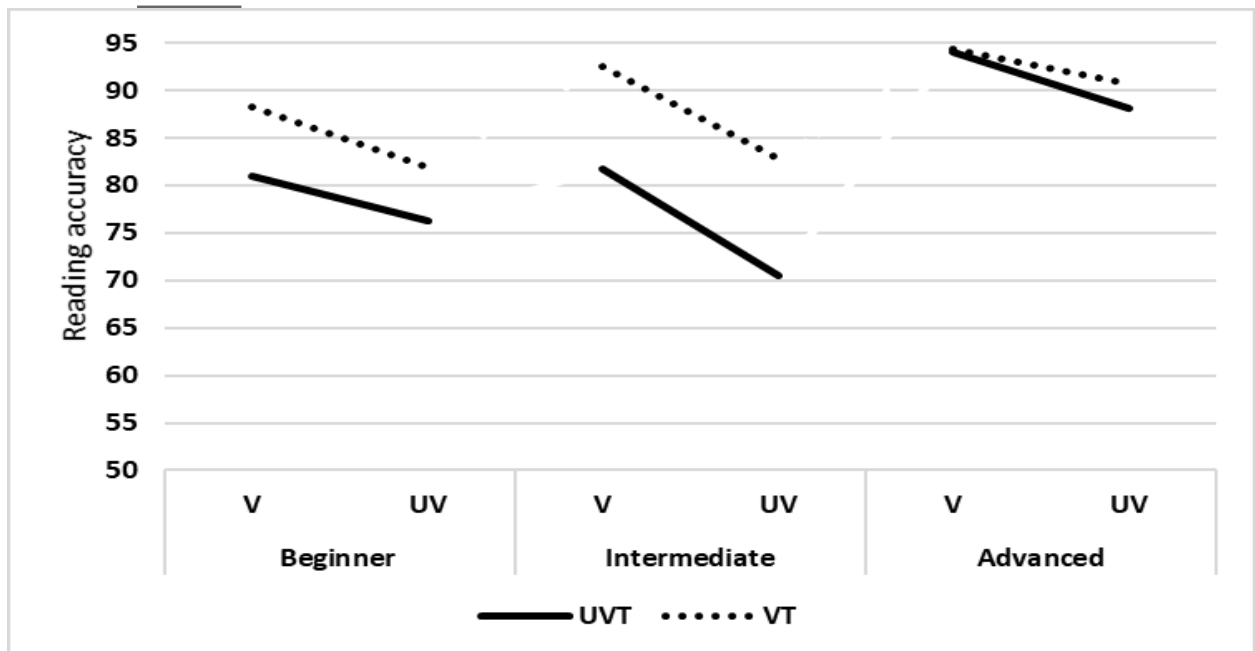
Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	5.396	0.030
	Condition (V vs UV)	1	22	6.148	0.021
	Group * Condition	1	22	0.143	0.709
Intermediate	Groups (VT vs UVT)	1	18	11.538	0.003
	Condition (V vs UV)	1	18	43.128	0.000
	Group * Condition	1	18	0.196	0.663
	Groups (VT vs UVT)	1	8	0.654	0.442

Advanced	Condition (V vs UV)	1	8	9.657	0.014
Group * Condition		1	8	0.617	0.455

<INSERT FIGURE 5 ABOUT HERE>

FIGURE 5

Reading Accuracy of Target Words at Three Proficiency Levels (Detailed Scale)



Results of the whole-text accuracy showed that participants in both groups read V texts more accurately than UV texts. However, participants in the VT group read both V and UV texts more accurately than participants in the UVT group, especially participants at the intermediate and advanced proficiency levels. Moreover, participants in the intermediate and advanced VT groups showed higher stability in reading under V and UV conditions than

participants in the corresponding UVT groups (see Table 9; Figure 6; see also Online Supporting Information Table 5).

<INSERT TABLE 9 ABOUT HERE>

TABLE 9
Reading Accuracy of Texts - Tests Effects (Whole Texts)

Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	.254	.619
	Condition (V vs UV)	1	22	13.986	.001
	Group * Condition	1	22	.278	.603
Intermediate	Groups (VT vs UVT)	1	18	5.029	.038
	Condition (V vs UV)	1	18	24.612	.000
	Group * Condition	1	18	.350	.561
Advanced	Groups (VT vs UVT)	1	8	.503	.498
	Condition (V vs UV)	1	8.000	.492	.503

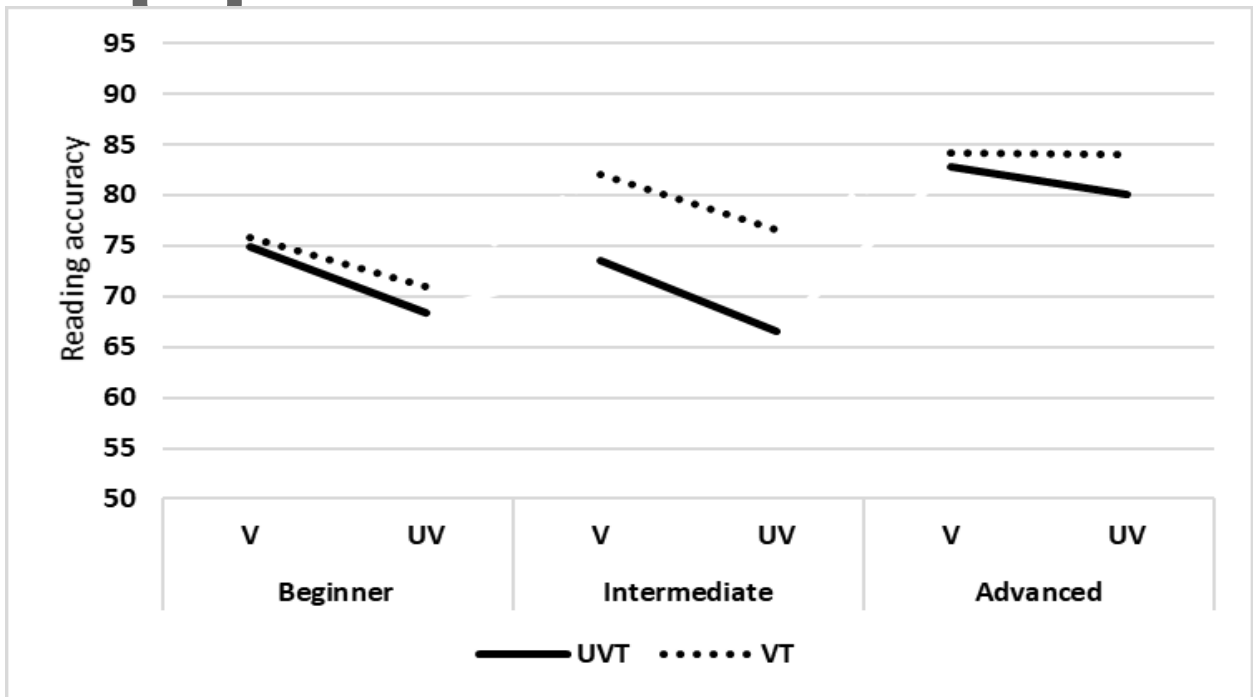
UV)

Group * Condition	1	8.000	.403	.543
-------------------	---	-------	------	------

<INSERT FIGURE 6 ABOUT HERE>

FIGURE 6

Reading Accuracy of Texts at Three Proficiency Levels (Whole Texts)



Word Comprehension Task

Finally, results of the comprehension analysis of target words in texts showed that participants in the beginner, intermediate, and advanced VT groups maintained an advantage of target word comprehension over their UVT counterparts. In addition, the intermediate and advanced levels in both the UVT and VT groups comprehended V target words slightly better than UV, but the beginner participants in both groups comprehended V target words significantly better than UV (see Table 10; Figure 7; see also Online Supporting Information Table 6)

<INSERT TABLE 10 ABOUT HERE>

TABLE 10

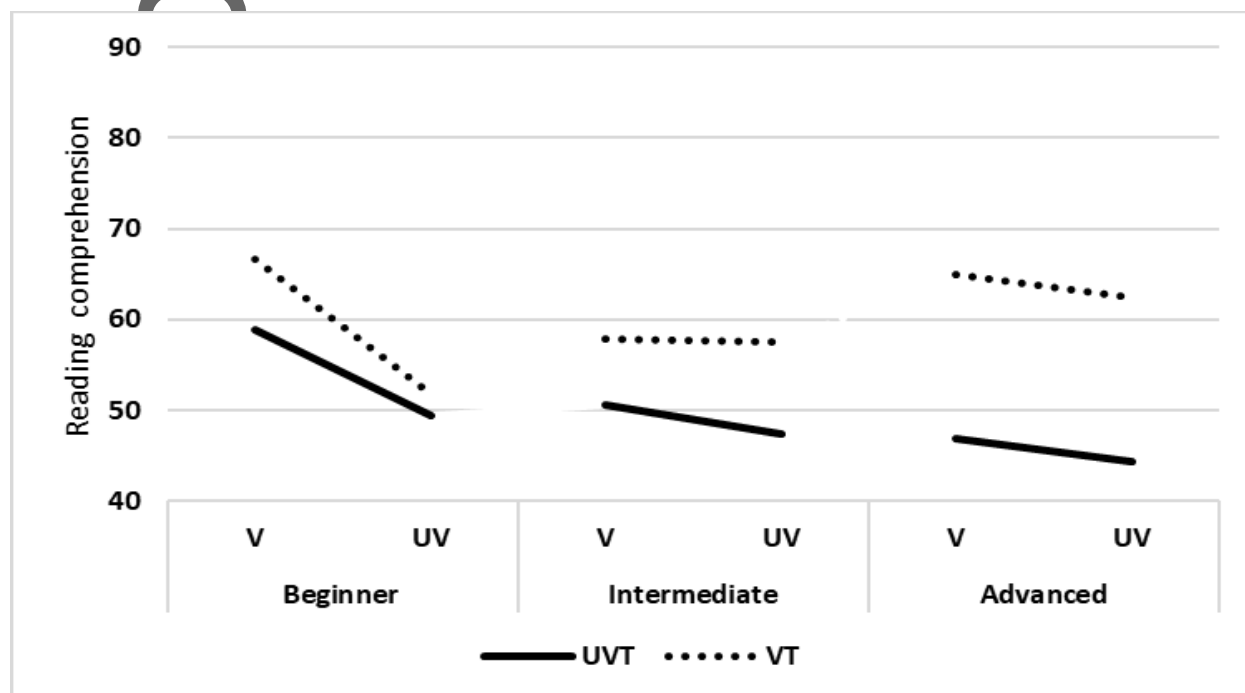
Reading Comprehension of Target Words - Tests Effects

Level	Source	Numerator df	Denominator df	F	Sig.
Beginner	Groups (VT vs UVT)	1	22	.001	.972
	Condition (V vs UV)	1	22	14.508	.001
	Group * Condition	1	22	1.478	.237
Intermediate	Groups (VT vs UVT)	1	18	.678	.421
	Condition (V vs UV)	1	18	.117	.736
	Group * Condition	1	18	.078	.783
Advanced	Groups (VT vs UVT)	1	8	85.246	.000
	Condition (V vs UV)	1	8	2.341	.165
	Group * Condition	1	8	.660	.440

<INSERT FIGURE 7 ABOUT HERE>

FIGURE 7

Reading Comprehension of Target Words at Three Proficiency Levels



<A>DISCUSSION AND CONCLUSION

Previous studies ignored input effects, especially those related to input frequency as well as the nature of diacritics exposure in the instructional input of participants. Input frequency refers to participants' (prior) repeated exposure to each word during their learning process. The present study attempted to control for this important factor by noting the frequency of each word in each level of textbook (which should be fewer than 12 times) to reduce the effect of high frequency exposure of word recognition as well as by including target words with comparable exposure by the two groups of participants. The nature of diacritics

exposure is related to the textbooks used by the two groups of participants, which either included or excluded diacritics from words.

The results of the isolated word reading task demonstrated that all three proficiency levels of participants who used a vowelized textbook read significantly faster than those who used the un-vowelized textbook when reading both vowelized and un-vowelized words. The results also showed the performance of the learners who relied on the vowelized textbook in all three proficiency levels was stable under both vowelized and un-vowelized conditions. In other words, even though vowelized textbook learners read vowelized words more slowly than un-vowelized words, the time difference was negligible between reading the vowelized and un-vowelized words (for example, two seconds to read vowelized word versus 1.95 seconds to read un-vowelized word). On the other hand, un-vowelized textbook learners at all three proficiency levels demonstrated greater discrepancy between reading speeds under both conditions (i.e., vowelized words were read in 2.9 seconds whereas un-vowelized words were read in 2.5 seconds). Accordingly, the un-vowelized textbook group seems to have encountered considerable difficulties in terms of reading speed when reading vowelized words. Thus, the results of the vowelized textbook group contradict the assumption that supplying diacritics in text when teaching Arabic as a foreign language might impede learners' reading when a word is encountered without diacritics. On the contrary, based on the findings of this study, it appeared that excluding diacritics from text when teaching Arabic as a foreign language might impede un-vowelized textbook learners' reading speed in both vowelized and un-vowelized conditions, especially when reading vowelized text. In other words, teaching Arabic words and texts without diacritics might negatively affect learners' reading speed when they later encounter words with diacritics, and it also could delay improvement in reading speed of un-vowelized words.

The text reading task revealed similar findings to those yielded by the isolated word reading task. Although the results demonstrated that the speed of reading vowelized text was slower than reading un-vowelized text by both groups in all three proficiency levels, participants at the beginner and intermediate levels of the vowelized textbook groups read both vowelized and un-vowelized texts at a significantly faster speed than their un-vowelized counterpart groups. In addition, participants at the advanced level in the vowelized textbook group maintained an advantage of reading speed over their un-vowelized counterparts. Furthermore, the results showed that participants in the vowelized textbook group maintained a stable reading speed under both vowelized and un-vowelized conditions, with very little difference in speed between the two conditions. On the other hand, participants in the un-vowelized textbook group showed an unstable performance, because the difference between their reading speeds under both vowelized and un-vowelized conditions was considerably wide at all proficiency levels.

Hence, with respect to whether diacritics play a role in reading speed and accuracy of isolated words and words in context by learners who rely on vowelized textbooks versus those who rely on un-vowelized textbooks at different stages of Arabic L2 acquisition, diacritics do seem to play a role in reading speed of *isolated words* and *words in context* by learners who rely on vowelized textbooks versus those who rely on un-vowelized textbooks at different stages of Arabic L2 acquisition. This observation is congruent with findings in previous studies that examined the effect of shallow orthography on the reading process in different languages and found that learners who relied on transparent orthographies achieved reading fluency earlier than their counterparts who relied on deeper orthographies (e.g., Frith, Wimmer, & Landerl, 1998; Goswami, Gombert, & de Barrera, 1998; Seymour, Aro, & Erskine, 2003). Similarly, the study results indicated that diacritics (i.e., in shallow

orthography) appeared to assist in the development of reading speed in learners of Arabic as an L2 under both vowelized and un-vowelized conditions.

The study results are also in support of certain findings in other studies (e.g., Hansen, 2010; Ibrahim, 2013; Taha, 2016; Taha & Azaizah-Seh, 2017) in the sense that participants of un-vowelized textbook group took more time to read the vowelized words than the un-vowelized words. Perhaps this is due to participants' conscious attention triggered by the presence of the diacritics to pronounce the words accurately. This result appears comparable with Hansen's (2010) conclusion that "for beginner and intermediate learners of Arabic, the additional graphical information that vowels represent adds a heavy cognitive burden on the already heavily charged decoding system. Due to this 'cognitive overload,' vowel information cannot be utilized" (p. 578). However, it is important to note that this may only be applicable to learners who have relied on un-vowelized textbooks. In contrast, the participants (in the present study) who relied on the vowelized textbook in their learning program from the beginner level seemed able to decode words and use the diacritics advantageously. Their reading speed results showed great stability in all three proficiency levels under both conditions. Namely, they could read vowelized and un-vowelized words with a slight difference in time between them.

The results of the un-vowelized textbook group might also be comparable with Hansen's (2010) conclusion, as the results showed a discrepancy in reading speed of vowelized versus un-vowelized words. This could be noticed in the significant interaction between group and condition at the beginner and intermediate levels and the nearly significant interaction that was found in the results of the advanced level. However, the whole picture of word recognition cannot be complete by an examination of only reading speed. Other elements must be examined, including reading accuracy.

The second part of the first and second research questions focuses on reading accuracy. Recall, accuracy was measured according to two main scales: the detailed scale and general scale. The former took into account the accuracy of every short vowel in each word (except the last vowel representing a grammatical case or mood ending), whereas the latter dealt with the word as a whole unit, such that any error in pronouncing any part of a word was considered incorrect pronunciation of the entire word. Based on the detailed scale the results showed a near-significant advantage achieved by participants using the vowelized textbook at the beginner and intermediate levels. An advantage at the advanced level over their counterpart participants in the un-vowelized textbook group was also maintained. Interestingly, the results also confirmed that participants using the vowelized textbook had a very stable performance under both vowelized and un-vowelized conditions, whereas participants using the un-vowelized textbook seemed to display unstable performance. The general scale yielded a similar pattern of findings (see Al Midhwah, 2018).

In other words, the learners who used the vowelized textbook always read vowelized words more accurately than un-vowelized words, which is congruent with previous studies (Abu-Hamour, et al, 2013; Abu-Rabia, 1997, 1999; Maroun & Hanley, 2017; Seraye, 2004). This benefit of reading with diacritics was consistent with the goal for the historical development of the diacritics system in Arabic orthography, which aimed to improve accuracy and comprehension when reading Arabic texts (Alhamad, 1982; Alhassan, 2003; Framawi, 1978; Ismaeel, 2001; Jum'ah, 1967; Sharshal, 2000). In contrast, the participants who used the un-vowelized textbook appeared to encounter difficulties when reading vowelized words. This fact could be explained by their lack of exposure to diacritics in their textbook, which exposed them to words with diacritics only once (in new vocabulary lists), after which they encountered words without diacritics in the remaining texts and drills.

Therefore, reading words with diacritics may have resulted in the “heavily charged decoding system” (Hansen, 2010, p. 578) and confused them during the reading process.

The results of the reading accuracy for the word in context reading task were consistent with those yielded by the isolated word reading task. The results obtained from the detailed scale indicated that participants in the beginner and intermediate groups of the vowelized textbook read both vowelized and un-vowelized significantly more accurately than their counterparts in the un-vowelized textbook group. At the advanced level, the participants from the vowelized textbook group also maintained advantage in terms of target word accuracy over their counterparts in the un-vowelized textbook group. The general scale produced somewhat similar results (see Al Midhwah, 2018).

Accordingly, and going back to the second half of the first and second research questions related to reading accuracy of isolated words and words in context, the results of both readings tasks indicate that participants who relied on a vowelized textbook in their learning program benefited from the presence of diacritics in words and texts. Moreover, using diacritics to learn Arabic from the earliest stages can assist with improving reading accuracy in both vowelized and un-vowelized conditions. Conversely, not including diacritics in textbooks could delay the acquisition of Arabic reading skills. Those participants who benefited from the use of diacritics demonstrated that changing Arabic text to shallow orthography assists with accelerating and enhancing reading performance. Hence, the results also seem to confirm the role of shallow orthography, which “makes the teaching of phonological recoding relatively straightforward and allows the acquisition of basic reading skills to proceed at a faster pace” (Frith, Wimmer, & Landerl 1998, p. 32), an idea that is at variance with deep orthography, which is a more complex process that relies heavily on other

components in addition to the phonological features of orthography (Frith et al., 1998; Goswami, Gombert, & de Barrera, 1998; Hansen, 2008; Seymour, Aro, & Erskine, 2003).

As for reading comprehension (in relation to the third research question), many L1 studies (e.g., Abu-Hamour, Al-Hmouz, & Kenana, 2013; Abu-Rabia, 1998, 1999, 2001; Abu-Rabia & Siegel, 1995; Maroun & Hanley, 2017; Seraye, 2004) and L2 (Hansen, 2010; Khaldieh, 2001), which attempted to examine the role of diacritics in terms of reading comprehension, had methodological limitations due to lack of controlling for certain reading comprehension factors, such as reader background related to topic exposure, frequency of exposure to words, and text difficulty and length. To avoid these limitations, the comprehension portion of this study, as in the two reading tasks, relates to target words rather than general comprehension questions. Target words were not entirely new to participants. To mitigate the effects of vocabulary difficulty (which would make participants resort to guessing) and to control for high-frequency exposure effects, target words consisted of words to which participants had been exposed fewer than 12 times, mostly between 1-6 times (see Online Supporting Information Appendices A-C). The nature of the design of target words and the congruity of results of the word comprehension task with those of the speed and accuracy of the word and text reading tasks allow us to conclude with some degree of certainty for the role of diacritics in facilitating word recognition and comprehension, although implementing recall protocols or similar measures would have allowed for more reliable evidence for this facilitative role. It is worthy of note here that the results of the homographic versus non-homographic words are compatible with the overall results reported on here (for results which include homographic versus non-homographic target words, see Al Midhwah, 2018).

The study results demonstrated that participants at the beginner level comprehended vowelized target words in text significantly more accurately than un-vowelized target words. The study of Abu-Rabia (1999), who investigated the effect of Arabic vowels on the reading comprehension of native Arabic children, reached a similar conclusion that vowels were a significant facilitator of reading comprehension. This study's results are also congruent with those of Abu-Hamour, Al-Hmouz, and Kenanac (2013), who examined the effect of short vowelization on comprehension in Arabic as an L1. They found that diacritics were a facilitator of oral reading comprehension in poor readers.

In terms of Arabic as an L2, however, the findings of Hansen's (2010) study indicated that short vowels did not significantly facilitate reading comprehension for learners of Arabic as a second language. While there was no significant difference in terms of reading comprehension under vowelized and un-vowelized conditions at the intermediate and advanced levels of both study groups of the present study, the results, nonetheless, indicated that target words in vowelized text were easier to comprehend than reading the same words in un-vowelized text. Likewise, in their study, Abu-Hamour et al. (2013) found that diacritics (in addition to background knowledge and context) were a significant facilitator of reading comprehension for skilled readers as well. Although the results of comprehension did not reveal statistically significant differences between the vowelized and un-vowelized textbook participants, the results nonetheless demonstrated that the vowelized textbook participants maintained an advantage in terms of reading comprehension, under both vowelized and un-vowelized conditions, over the un-vowelized textbook group.

In sum, based on the foregoing study findings, an obvious advantage in reading speed, accuracy, and comprehension can be observed among the participants who relied on the vowelized textbook. Furthermore, the vowelized textbook participants achieved a high level

of stability in their reading speed and accuracy in both vowelized and unvowelized conditions, which supports the claim that shallow orthography (i.e., with diacritics) may assist readers with earlier achievement of reading fluency than deep orthography (i.e., without diacritics) (Frith et al., 1998). In other words, deep orthography, as used in the unvowelized textbook, may pose a challenge to reading performance because this type of orthography contains more ambiguous orthographic-phonological relations than shallow orthography (Goswami et al., 1998). In turn, this difficulty may delay the progress of learning to read Arabic. Hence, further study of the two types of orthographies was worthwhile for detecting the effects of each type on word recognition and reading performance in Arabic as an L2. Accordingly, employing the benefits of shallow orthography could open the door to enhancing teaching of Arabic as a second language.

Finally, if we are right in assuming the learning modes of diacritics for short vowels (as exhibited by the two groups of participants in the study) are those of implicit versus explicit learning, then the study seems to suggest that explicit learning is more robust than implicit learning. Accordingly, the study findings seem to confirm previous findings in two respects. First, since diacritics do not involve complex grammatical rules, the study confirms the finding that explicit learning leads to significantly more learning gains than implicit learning (e.g., DeKeyser, 1995). Second, and more evidently, by receiving a sustained presence of short vowel diacritics in their instructional, written input (in both their textbook and teachers' classroom instructional practices), the VT group seems to have received a more advantageous (visual) input than their UV counterpart, confirming findings reported in Kim and Godfroid's (2019) study. This is justified by the superior performance of the VY group over that of the UV group. Since the present study did not include measures of awareness, the first possibility must be considered tentative pending further research.

To conclude, it is critical to understand the full scope of the effects of the two types of Arabic orthographies on the process of word recognition and reading and the two modes of learning (implicit versus explicit) through which short vowel diacritics are learned. Many studies on different European languages, such as English (e.g., Goswami et al., 2001), French (e.g., Goswami et al., 1998), Greek (e.g., Goswami, Porpodas & Wheelwright, 1997), Spanish (e.g., Goswami et al., 1998), German (e.g., Frith et al., 1998), have suggested that orthography has an effect on the process of reading acquisition. Arabic and other languages that have both deep and shallow orthography including those which employ diacritics like Arabic (or even Romanized script versus logographic script) would benefit from further research. Future replication or extension studies should also avoid some of the limitations of the present study. One limitation is the small sample of participants, especially those of the advanced groups. Another limitation is lack of implementation of additional measures such as recall protocols when investigating the role of diacritics in word comprehension. Research in this area is promising and is likely to lead to crucial findings in how the two types of orthographies of natural languages interact with explicit and implicit learning further under different conditions and settings. Similarly, the role of diacritics in learning Arabic (and similar languages) as an L2 may also provide important findings if examined in terms of other language skills such as writing.

ACKNOWLEDGMENTS

This research was supported by the Deanship of Scientific Research and RSSU at King Saud University and a Rackham Research Grant at the University of Michigan. We express our thanks to the editor, Dr. Marta Antón, and the three anonymous reviewers for their insightful comments.

NOTES

¹ Historically, the development of Arabic script and the incorporation of diacritics were attributed to the spread of errors among native and non-native Arabic speaker with the rise of literacy in the late 7th to early 8th century C.E. (Ismaeel, 2001). Prior to that time, Arabic script had included only 26 symbols for consonants and three symbols for long vowels, but none for short vowels. In addition to symbols representing internal short vowels, diacritics representing inflectional endings of case and mood were also developed. Inflectional endings (a.k.a. *ʔiʕrāb* endings) also involve short vowels, but they are syntactic in nature and do not affect word meaning as opposed to internal short vowels. The scope of the present study is confined to internal short vowels and does not include diacritics representing inflectional endings.

² Consider also the motivation behind the historical development of Arabic script incorporating short vowel diacritics, as explained in footnote 1, since, among other things, the change of one vowel may even result in a different word with a different meaning, resulting in possibly stigmatizing errors. This is true today in the Arabic L2 context where, for example, in beginner level classes the high frequency modifier *ʔuḫrā* ‘other’ is often mispronounced as *ʔaḫrā* ‘shittier’ or in later levels the word *quwwād* ‘leaders’ as *qawwād* ‘pimp.’

³ The scope of the present study does not include *ʔiʕrāb* “case and mood” endings.

⁴ Frequency of occurrence in this study is limited to textbook word counts.

⁵ The results reported here do not include a breakdown of homographic versus non-homographic. The results are nevertheless compatible with the overall results reported on here (for results which include homographic and non-homographic target words, see Al Midhwah, 2018).

⁶ Moreover, the two text types (the vowelized and non-vowelized) in each level were homogenized so that they are comparable in terms of frequency of both word token and word type (for more details, see Al Midhwah, 2018). In counting diacritics, each diacritic was counted as one character.

⁷ The pilot study was conducted on six students from three corresponding levels. They followed all steps and performed all the tasks designed for the study. The pilot provided us with valuable information related to the readability of the typeface and font size as well as execution and timing of the tasks. The pilot was also beneficial in terms of adding clarity to task instructions. Most importantly, the pilot study assured us the texts were at the appropriate levels of the participants, not surprising due to the many measures which we relied on in constructing the texts.

⁸ Because both texts contained the same target words used to measure speed and accuracy, both comprehension sets of questions were administered after the reading of the two texts in order to control for participants' awareness of such words, in particular when encountering them again in the reading of the second text.

⁹ An anonymous reviewer commented that it is not clear what exactly is being measured here; suppose the word is hard to pronounce and takes longer to say, does that mean that "word recognition" is slower? We would like to point out that this is not a purely perception or recognition experiment. The task is also intended to assess how the learner verbalizes the perception of a given word, how they learned it, and how they are producing it, where accuracy is another major concern.

REFERENCES

- Abu-Hamour, B., Al-Hmouz, H., & Kenana, M. (2013). The effect of short vowelization on curriculum-based measurement of reading fluency and comprehension in Arabic. *Australian Journal of Learning Difficulties, 18*, 181–197.
- Abu-Rabia, S. (1996). The role of vowels and context in the reading of highly skilled native Arabic readers. *Journal of Psycholinguistic Research, 25*, 629–641.
- Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers in reading paragraphs, sentences, and isolated words. *Journal of Psycholinguistic Research, 26*, 465–482.
- Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type and vowelization. *Reading and Writing, 10*, 105–119.
- Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research, 28*, 93–101.
- Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*, 39–59.
- Abu-Rabia, S. (2007). The role of morphology and short vowelization in reading arabic among normal and dyslexic readers in grades 3, 6, 9, and 12. *Journal of Psycholinguistic Research, 36*, 89–106.
- Abu-Rabia, S., & Siegel, L. S. (1995). Different orthographies different context effects: the effects of Arabic sentence context in skilled and poor readers. *Reading Psychology, 16*, 1–19.
- ACTFL. (2012). Accessed 1 January 2017 at <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>

- Al-Batal, M. (2006). Playing with words: Teaching vocabulary in the Arabic language curriculum. In K. M. Wahba, Z. A. Taha, & L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century* (pp. 331–340). New Jersey, NJ: Lawrence Erlbaum.
- Alhamad, G. (1982). *Rasm al-muṣḥaf: Dirāsa lughawiyya tārikhiyya*. Baghdad, Iraq: al-Lajna al-Waṭaniyya.
- Alhassan, S. (2003). *Al-kitāba al-‘Arabiyya min al-nuqūš ‘ilā al-kitāb al-makhṭūf*. Riyadh: Dār Al-Fayṣal Al-Thaqāfiyya.
- Alhawary, M. T. (2017). Empirical directions in the future of Arabic second language acquisition research and second language pedagogy. In K. M. Wahba, Z. A. Taha, and L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century II* (pp. 408-421). London and New York: Routledge.
- Al Midhwah, A. A. (2018). *The role of diacritics in word recognition and their impact on Arabic L2 learners’ reading speed, accuracy, and comprehension at different stages of Arabic L2 acquisition*. (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, MI.
- Alosh, M., & Clark, A. (2010). *Ahlan wa sahlān: Functional Modern Standard Arabic for beginners, second edition*. New Haven, CT: Yale University Press.
- Alosh, M., & Clark, A. (2013). *Ahlan wa sahlān: Functional Modern Standard Arabic for intermediate learners, second edition*. New Haven, CT: Yale University Press.
- Baluch, B. (1992). Reading with and without vowels: What are the psychological consequences? *Journal of Social and Evolutionary Systems*, 15, 95–104.
- Brustad, K., Al-Batal, M., & Al-Tonsi, A. (2004). *Al-kitaab fii ta‘allum al-‘Arabiyya: A textbook for beginning Arabic, part one, second edition*. Washington, DC: Georgetown University Press.

- Brustad, K., Al-Batal, M., & Al-Tonsi, A. (2007). *Al-kitaab fii ta'allum al-'Arabiyya, part two, second edition*. Washington, DC: Georgetown University Press.
- DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17, 379-410.
- Ellis, N. (1994). Implicit and explicit language learning—An overview. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1-31). London: Academic Press.
- Framawi, A. (1978). *Qiṣṣat al-naqṭ wa al-shakl fī al-Muṣḥaf al-Sharīf*. Cairo: Dār Al-Nahḍa Al-'Arabiyya.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2, 31–54.
- Goswami, U., Porpodas, C., & Wheelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education*, 12, 273-292.
- Goswami, U., Gombert, J. E., & Barrera, L. F. (1998). Childrens orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, 19, 19–52.
- Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2001). Pseudo homophone effects and phonological recoding procedures in reading development in English and German. *Journal of Memory and Language*, 45, 648-664.
- Hama, M., Leow R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, 32, 465-491.
- Hansen, G. (2008). Visual word recognition in Arabic: Towards a language-specific reading model. University of Southern Denmark: Centre for Contemporary Middle East Studies, Working Paper Series No. 13.
- Hansen, G. F. (2010). Word recognition in Arabic as a foreign language (AFL). *Modern*

Language Journal, 94, 567–581.

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second language learning: Introduction. *Studies in Second Language Acquisition*, 27, 129-140.

Ibrahim, R. (2013). Reading in Arabic: New evidence for the role of vowel signs. *Creative Education*, 4, 248–253.

Ismaeel, S. M. (2001). *Rasm al-muṣḥaf wa ḍabṭuh bayna al-tawqīf wa al-iṣṭilāḥāt al-ḥadīthah*. Cairo: Dār Al-Salām.

Jum‘ah, E. (1967). *Dirāsa fī taṭuwwur al-kitābāt al-kūfiyya ‘alā al-’ahjār fī miṣr fī al-qurūn al-khamsa al-’ulā lil-hijra*. Cairo: Dār Al-Fikr Al-‘Arabiyy.

Khaldieh, S. (2001). The relationship between knowledge of icraab, lexical knowledge, and reading comprehension of nonnative readers of Arabic. *Modern Language Journal*, 85, 416–431.

Kim, K. M., & Godfroid, A. (2019). Should we listen or read? Modality effects in implicit and explicit knowledge. *Modern Language Journal*, 103, 3, 648–664.

Leung, J., Williams, J. N. (2012). Constraints on explicit learning of grammatical form-meaning connections. *Language Learning*, 62, 634-662.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, 17, 673–680.

Maroun, M., & Hanley, J. R. (2017). Diacritics improve comprehension of the Arabic script by providing access to the meanings of heterophonic homographs. *Reading and Writing*, 30, 319–335.

Ranking Academic Ranking of World Universities. (2016). Accessed 9 February 2017 at

<http://studychacha.com/discuss/201718-ranking-academic-ranking-world-universities.html>

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33, 829-856.

Robinson, P. (1995). Attention, memory, and the Noticing Hypothesis. *Language Learning*, 45, 283-331.

Ryding, C. (2013). *Teaching and learning Arabic as a foreign language: a guide for teachers*. Washington, DC: Georgetown University Press.

Saragi, T., Nation, I. S. P., and Meister, G. F. (1978). Vocabulary learning and reading. *System* 6, 72-78.

Saiegh-Haddad, E., & Schiff, R. (2016). The impact of diglossia on vowelized and unvowelized word reading in Arabic: A developmental study from childhood to adolescence. *Scientific Studies of Reading*, 20, 311-324.

Schiff, R. (2012). Shallow and deep orthographies in Hebrew: The role of vowelization in reading development for unvowelized scripts. *Journal of Psycholinguistic Research*, 41, 409-424.

Seraye, A. M. (2004). *The role of short vowels and context in the reading of arabic, comprehension and word recognition of highly skilled readers*. (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143-174.

Sharshal, A. (2000). *Al-ṭarrāz fī šarḥ ḍabṭ al-kharrāz*. Madina, Saudi Arabia: King Fahd Complex for the Printing of the Holy Quran.

Shimron, J., & Sivan, T. (1994). Reading proficiency and orthography: Evidence from

Hebrew and English. *Language Learning*, 44, 5–27.

Taha, H. (2016). Deep and shallow in Arabic orthography: New evidence from reading performance of elementary school native Arab readers. *Writing Systems Research*, 8, 133–142.

Taha, H., & Azaizah-Seh, H. (2017). Visual word recognition and vowelization in Arabic: New evidence from lexical decision task performances. *Cognitive Processing*, 18, 521–527.

Taha, H. (2016). The development of reading and spelling in Arabic orthography: Two parallel processes? *Reading Psychology*, 37, 1149–1161.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. N. Williams, S. Rott, M. Overstreet (Eds.), *From meaning connections in second language acquisition* (pp. 203-218). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27, 269–304.

Author Manuscript