International Journal of
MULTILINGUAL EDUCATION

ISSN: (Print) ISSN 1987-9601
(Online) E ISSN 1512-3146
Journal homepage: http://multilingualeducation.org/

Diacritics as Key Predictors: Enhancing Word Recognition in Qurʾānic Texts Despite Contextual Constraints

Amina Abdelhadi
Ibn Khaldoun University of Tiaret, Algeria

Email: amina.abdelhadi@univ-tiaret.dz

To cite this article:Amina Abdelhadi (2026), Diacritics as Key Predictors: Enhancing Word Recognition in Qurʾānic Texts Despite Contextual Constraints: International Journal of Multilingual Education, volume 27, issue 1. pp. 120-133.

To link to this article:
https://doi.org/10.22333/ijme.2026.11581

Amina Abdelhadi[1]

Ibn Khaldoun University of Tiaret, Algeria

https://orcid.org/0009-0006-8024-7960

Diacritics as Key Predictors: Enhancing Word Recognition in Qurʾānic Texts Despite Contextual Constraints

ABSTRACT

It is known that diacritics *(ḥarakāt)* are essential to reading Arabic but their importance for accurate recognition of words has not been fully understood across text types and levels of reader expertise. This mixed-methods study explores the role of diacritics in how readers at different stages of lexical development decode Modern Standard Arabic (MSA) and Qur'ānic passages across two phases. In the first phase, a total of forty-five native Arabic-speaking readers from three educational levels were presented with ten unpracticed, undiacritized MSA sentences. In the second phase, after recording accuracy, speed, and errors, the top fifteen participants read ten Qur'ānic verses selected for complexity of context and orthographic similarities. Both phases used homographs and complex sentences to examine whether word recognition could be predicted from context. A retrospective verbal protocol was also used to assess how diacritics and context affected reading accuracy. It is found that, in MSA texts, reliance on diacritics decreases but remains important when contextual cues are weak while, in Qur'ānic texts, diacritics strongly enhance word recognition because contextual cues are limited. These findings support a simple Arabic word disambiguation model, highlighting that accurate initial reading of classical texts, including non-memorized Qur'anic passages, relies heavily on diacritics regardless of the reader's level of skill.

Keywords: *Arabic reading, word recognition, diacritics, contextual disambiguation, Qur'ānic text, reading proficiency*

Introduction

Imagine reading many isolated words that appear identical but can convey completely different pronunciations and meanings depending on tiny vowel sounds *(diacritics: ḥarakāt)* that are usually not written. This is the everyday reality of Arabic literacy. In Arabic, words are typically written as '*rasm'*, a consonantal skeleton lacking short vowels, which creates lexical and phonological ambiguity and makes word recognition challenging (Bashir et al., 2023). At the core of reading research is word recognition**,** the cognitive process of extracting a word's sounds and meanings from its written form. It involves converting letters into sounds, accessing stored meanings, and integrating them into the text for comprehension (Koda, 2005). Relatedly, Arabic word recognition refers to the cognitive processes

[1] Corresponding Author: *Amina Abdelhadi, amina.abdelhadi@univ-tiaret.dz*

that transform printed letter strings into stored lexical representations and their associated sounds and meanings, involving grapheme-to-phoneme mapping, morphological processing, and access to lexical entries (Abu-Rabia, 1995; Alhussein, 2017). For instance, the sequence كتب can be read as كَتَبَ /kataba/ ("he wrote"), كُتِبَ /kutiba/ ("it was written"), or كُتُب /kutub/ ("books"). Because short vowels are generally absent, a single unvowelled string may correspond to multiple pronunciations and meanings, producing heterophonic homographs in which orthography alone does not determine pronunciation, and pronunciation alone does not determine meaning. Readers must therefore rely on key predictors, including diacritics, morphology, syntax, semantic context, and prior knowledge of Arabic orthography, to accurately recognize words (Ammar, 1997; Alhussein, 2017).

Arabic diacritics are commonly grouped into three sets: short vowels, nunation (*tanwīn*), and syllabification marks. The short vowels '*fatha*, *damma*, and *kasra*' represent the sounds /a/, /u/, and /i/, while *tanwīn* adds final nasalized endings such as /an/, /un/, and /in/. Additional marks such as *sukūn* and *shaddah* indicate syllable structure by marking the absence of a vowel or consonant gemination (Bashir et al., 2023). These diacritics serve as powerful predictors and orthographic cues (Ryding, 2005; Abu-Rabia & Haj, 2021), helping map written Arabic to correct pronunciation and reduce ambiguity. When diacritics are absent, heterophonic homographs create ambiguity in both pronunciation and meaning, whereas homophonic homographs share spelling and pronunciation but differ in meaning even when diacritics are present (Peleg & Eviatar, 2009, as cited in Alhussein, 2017). Fully diacritized texts scaffold pronunciation and support word recognition, particularly for beginning readers. With experience, readers increasingly rely on contextual, morphological, and syntactic cues, which reduces but does not eliminate their dependence on diacritics (Boudelaa, 2014). Nevertheless, challenges persist in texts where vocabulary is rare and sentence structures are complex, and where phonological processing poses a major obstacle, often leading to spelling errors, and to ambiguity in word recognition and interpretation (Abu-Rabia & Taha, 2006; Maroun & Hanley, 2019; Hermena et al., 2021).

The present study focuses on how diacritics function as the primary cue for resolving phonological ambiguity in Arabic reading, particularly when context provides limited guidance. Morphology and syntax were disregarded in this study due to concerns about superficial data and the need for tight control over the tasks. More precisely, it examines how reliance on Arabic diacritics as a predictive cue shift depending on the reader's developmental stage and the type of text. Cue weighting can be summarized as:

- ***Modern Arabic reading:*** Context > Diacritics
- ***Classical/Qurʾānic reading:*** Diacritics > Context

In other words, the greater the phonological ambiguity of a word and the weaker the contextual cues, the more critical diacritics are for correct recognition. Building on this conceptualization, the present study addresses the following questions:

1. How do diacritics support word recognition across developmental stages of Arabic readers?
2. To what extent does reliance on diacritics decrease as readers increasingly depend on contextual cues in modern texts?
3. When reading Classical or non-memorized Qurʾānic passages, how does sentence- or verse-level embedding influence the use of diacritics for accurate word recognition?

This study highlights a model of Arabic reading in which diacritics play a central role in resolving lexical ambiguity, especially when contextual cues are limited. It clarifies how diacritics interact with phonology and linguistic context and provides a framework for supporting accurate word recognition in both modern and classical Arabic texts 'Qur'ānic passages in particular'.

Literature review

Guided by the present study's focus on Arabic word recognition, this literature review outlines the Arabic language and Qurʾānic Arabic, with emphasis on diacritics in different reading contexts, while highlighting a gap in how Qurʾānic features are treated in existing research.

The Arabic Language: Alphabet, Varieties and Diacritics

Arabic is a Semitic language written from right to left. It uses an abjad (consonantal) writing system, where the symbols usually represent consonants, and the reader adds the vowels when reading. It has 28 letters, including 25 consonants and three long vowels, most of which have up to four forms depending on their position (Saiegh-Haddad & Joshi, 2014). The script also uses thirteen diacritical marks for short vowels, nunation, and syllables. They help the reader know the correct pronunciation and grammatical meaning (Azmi & Alsaiari, 2014). Arabic first appeared between the first and fourth centuries CE (Al-Azami, 2020) and is generally divided into Classical Arabic (CA), Modern Standard Arabic (MSA), and Dialectal Arabic (DA). CA is prescriptive, while MSA preserves classical structures and adds modern vocabulary, which allows it to be used for both traditional and contemporary contexts (Azmi et al., 2019).

The Arabic script has no capitalization, which makes proper names difficult to identify. In addition, the absence of diacritics creates lexical ambiguity, since a word such as عقد /ʕaqd/ can mean "contract," "knot," "tie," "necklace," or "to make complex" (Azmi & Aljafari, 2018). To reduce this ambiguity, Arabic uses diacritics that function as orthographic cues indicating short vowels and correct pronunciation. Research shows that these marks help readers select the intended meaning of ambiguous words, although alternative meanings may remain partially active (Maroun & Hanley, 2017; Maroun,

2018). At the same time, fully vowelized texts can slow down early reading and cause longer eye fixations, but they usually help readers understand the text better later, showing a speed–accuracy trade-off (Hermena & Reichle, 2020; Hermena et al., 2021). Readers may also process diacritics parafoveally before directly looking at a word, which allows them to anticipate pronunciation and meaning (Hermena et al., 2021). When diacritics are absent, readers tend to rely on the most frequent interpretation, whereas their presence guides them toward context-specific meanings. That is to say, the positive role of diacritics in terms of Arabic word recognition and reading performance suggests that adding them to instructional texts not only helps Arabic L2 learners resolve word ambiguity but also enhances their overall reading and pronunciation skills (Midhwah & Alhawary, 2020). Because of these benefits, selective use of diacritics on ambiguous words is often recommended in teaching to support learning without increasing visual load for skilled readers. However, since Modern Standard Arabic often omits diacritics and depends on context to interpret meaning, ambiguity persists**.** This for example increases the difficulty of tasks like translation and Named Entity Recognition (NER) in natural language processing systems (Azmi & Almajed, 2015 as cited in Bashir et al., 2023).

Qurʾānic Considerations and Diacritics

Qurʾān is the central religious text of Islam. Muslims believe it was revealed in Arabic to Prophet Muhammad (peace be upon him) over 23 years, ending in 632 CE. The word *Qur'an* appears around 70 times in the text and comes from the Arabic verb *qara'a*, meaning "to read" or "to recite" (Britannica, 1999). Because it is recited in Arabic, the language is the liturgical language for about 1.8 billion Muslims**,** most of whom are non-Arabs. Therefore, knowing Arabic can help readers understand the Qur'an's true message. The text has 114 chapters and 6,236 verses, totaling 157,935 words, of which 5,277 are unique. Its chapters cover daily life, social conduct, history, and future events and are written in a concise, almost poetic style (Bashir et al., 2023).

The Qur'an is fully vowelized, which ensures accurate recitation. In this text, diacritics guide pronunciation more than reading speed or continuous comprehension. They help learners read correctly without changing the traditional script (Boulhrir & Chekayri, 2025). Over 1,400 years, thousands of scholars have studied the Qur'an and produced extensive commentary. However, previous studies rarely examine how diacritization affects real-time word recognition in Qur'ānic texts. Research on selective diacritics and eye-tracking is also limited. This study addresses that gap, highlighting the moderating role of diacritics in reading Qur'ānic texts.

Methodology

The study at hand uses a mixed-methods approach, combining quantitative and qualitative data. In Phase One, quantitative analysis measures reading performance in terms of accuracy, speed, and errors. It compares beginner, skilled, and advanced readers using unpracticed ambiguous words, homographs,

and complex sentences, which were intentionally structured to focus on the role of context. In Phase Two, qualitative observation examines how advanced readers interpret ambiguous words and homographs in undiacritized Qur'anic verses**.** A retrospective verbal protocol **(**Ericsson & Simon, 1993) was used in this phase. It helps to see whether context is present or absent, which reduces reliance on diacritics for word recognition and whether adding diacritics improves reading accuracy**.**

Research Design

This study is grounded in a constructivist–cognitive framework, which treats reading as an active process of meaning construction (van den Broek et al., 2005; Khadka, 2024). The present study acknowledges that, in Arabic, readers can often correctly read undiacritized text by using syntax, morphology, and sentence context, but diacritics still improve comprehension of ambiguous words and semantic decisions for them (Maroun & Hanley, 2017; Al-Samarraie et al., 2020). In other words, reliance on diacritics may decrease as readers develop syntactic and morphological awareness and contextual inference skills. Yet it addresses the gap of whether this reliance persists when reading classical texts and non-memorized Qur'anic passages in particular.

Participants

The study used purposive sampling to select 45 native Arabic speakers from three educational levels. Primary students were in Grades 3–5, middle students in Grades 1–4, and secondary students in Grades 1–3. Each group had 15 participants. All reported no reading impairments or difficulties with diacritized texts. Participants were grouped by educational stage, reflecting different reading development levels. **Table 1** shows the characteristics of each group and the corresponding phase.

Table 1. Participant Groups, Reading Levels, and Assigned Phases

Group	Reading Level	Description	N Participants	Phase
Group 1	Beginner Readers	Primary school pupils at an early stage of learning to read Modern Standard Arabic (Grades 1–2 excluded)	15	Phase One
Group 2	Developing Readers	Middle school students with developing proficiency in reading Modern Standard Arabic	15	
Group 3	Proficient Readers	Secondary school students with high proficiency in reading Modern Standard Arabic and familiarity with undiacritized texts	15	
Top 15 performers	Advanced Readers	Selected from Phase One based on mean accuracy scores; read contextually complex Qur'anic verses	15	Phase Two

Corpus Selection

The reading materials were divided into two phases. Phase One included 10 sentences in Modern Standard Arabic, presented without diacritics, for all 45 participants across beginner, developing, and proficient reader groups. Sentences ranged from 9 to 15 words and contained homographs and ambiguous words. Each participant read all 10 sentences, resulting in a total of 450 readings. This phase measured baseline reading performance**,** recording errors, attempts, hesitation and self-corrections. Phase Two involved the top 15 performers (represented as a group) from Phase One and used 10 Qur'anic verses selected for orthographic complexity. Each participant read all 10 verses, giving a total of 150 readings.

Phase One Process

In *Phase One*, a structured grid was used to record participants' performance while reading ten undiacritized sentences. For each sentence, the criteria recorded included Reading Time (s), Errors**,** and Accuracy (%)**.** First, reading time was noted, as longer times may indicate hesitation or difficulty. Next, errors on the first attempt were counted, and accuracy was calculated based on these errors. Finally, participant-level accuracy was computed as the average across all ten sentences.

Table 2: Phase One Reading Assessment Grid for Word Recognition

Errors (1 Attempt)	Reading Time (s)	Base Accuracy (%)
0	≤10	100
0	10–13	95
0	>13	90
1	≤10	85
1	10–13	80
1	>13	75
2	≤10	70
2	10–13	65
2	>13	60
3	≤10	60
3	10–13	55
3	>13	50

Note: G*iven that each sentence contains only 9 to 15 words**, Sentences with 4 or more errors are excluded from scoring as they represent extreme difficulty.*

The study examines how readers rely on context to determine meaning, disregarding syntax and morphology, although these cues can sometimes help. Accuracy decreases when errors occur, and longer reading times indicate hesitation. For instance, in the undiacritized sentence

"كتب المؤرخ علم عن عقد المدينة القديمة وسلم النتائج للمكتبة" (S6), **كتب** may initially be read as *kataba* ("he wrote") or *kutub* ("books"), **علم** as *ʿalima* ("knew") or *ʿallama* ("taught"), **عقد** as *ʿaqd* ("knot/contract/necklace"), and **سلم** as *handed/greeted* or *ladder*, but the narrative context clarifies the

intended meanings.

In **"سلم المعلم اللوح المكتوب و عقد للطالبة المتفوقة"** (S9), **سلم, اللوح,** and **عقد** could each be interpreted in multiple ways, yet context indicates the correct readings.

Phase Two Process

Proficient readers were asked to read selected Qur'anic verses aloud. The verses were first presented without diacritics to examine how contextual cues alone support disambiguation. For example, **وجوه يومئذ ناضرة إلى ربها ناظرة** (75:22–23) illustrates contextual differentiation between **ناضرة** and **ناظرة**. Similarly, **ولا تزر وازرة وزر أخرى** (6:164) and **إنما يخشى الله من عباده العلماء** (35:28) show that readers rely mainly on context to determine meaning. Another example is **فكأين من قرية أهلكناها وهي ظالمة فهي خاوية على عروشها وبئر معطلة وقصر مشيد** (22:45), where interpretation may depend on contextual inference when diacritics are absent. Additional examples include **فلما رأينه أكبرنه وقطعن أيديهن** (12:31), **فما استطاعوا أن يظهروه وما استطاعوا له نقبا** (18:97), and **قال أرأيتك هذا الذي كرمت علي لئن أخرتن إلى يوم القيامة لأحتنكن ذريته إلا قليلا** (17:62), illustrating how lexical ambiguity can be resolved through context. After the initial reading, the same words were presented with full diacritics and read again. A retrospective verbal protocol was then used, in which participants described any hesitations, misreadings, or self-corrections. This allowed the researchers to examine whether diacritics improved reading accuracy and how readers relied on contextual cues.

In Phase One, participants read unpracticed Modern Arabic texts, which were generated by the researchers specifically for this study and not taken from existing sources. Mean scores were calculated for each group to allow for ANOVA analysis. In Phase Two, advanced readers were observed reading ambiguous Qur'anic verses, and researchers recorded correct word recognition and whether diacritics improved accuracy. Participation in the study was voluntary**,** all data were anonymized**,** and no prior memorization of Qur'anic texts was emphasized.

Results

This section presents the main findings. Phase One shows reading performance of Modern Standard Arabic without diacritics by proficiency, and Phase Two examines recognition of Qur'anic words without diacritics.

Phase One – Quantitative Results

The first phase measured reading performance of 45 participants across three proficiency levels: Beginner, developing, and proficient. Three dependent variables were analyzed: Accuracy (%), Reading Time (s), and Errors.

Table 3. ANOVA results for the dependent variables across groups in phase one

Dependent Variable	df (Between)	Mean Square	F	Sig.
Time	2	148.062	8558.252	< 0.001
Errors	2	135.576	544.789	< 0.001
Accuracy%	2	49,760.056	1062.602	< 0.001

The ANOVA analysis showed that the three groups differed significantly on reading time, errors committed, and level of accuracy. This indicates that group membership in Phase One had a strong effect on performance. Overall, performance improved progressively from the Emergent group to the Developing group, and from the Developing group to the Proficient group, as reflected in **Figure 1**, which highlights the mean differences between groups based on post hoc analysis.

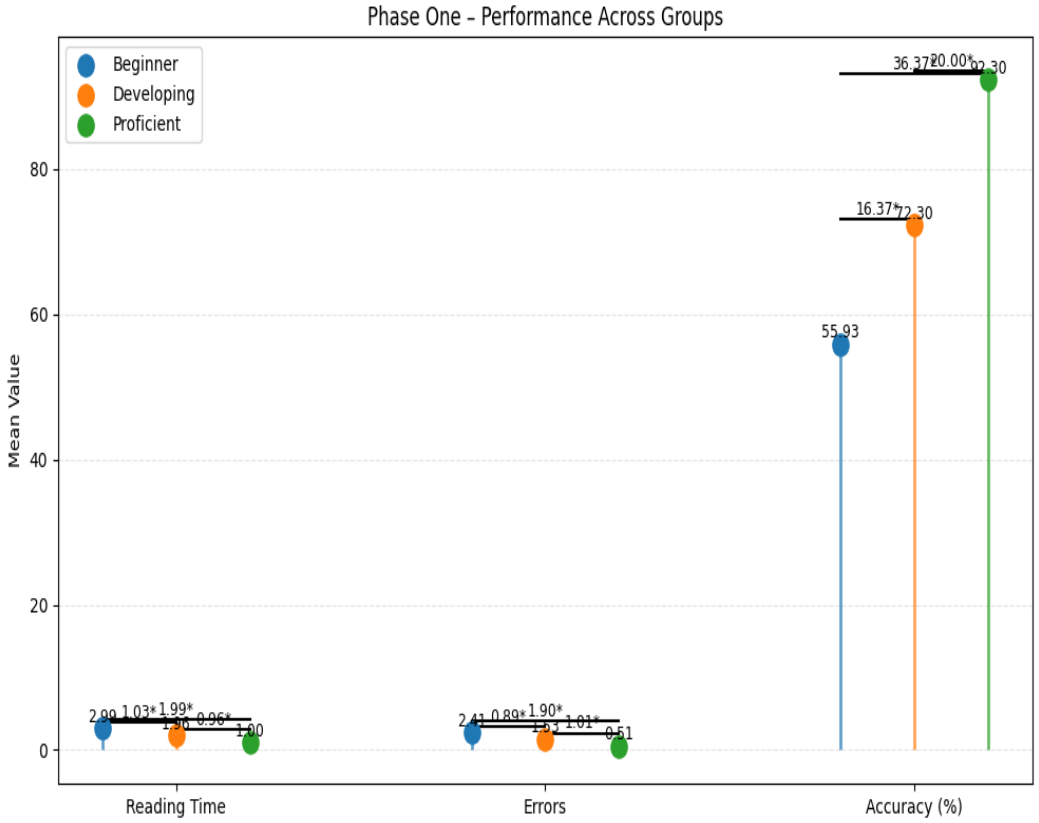

Figure 1. Mean performance differences across groups in phase one

The three groups differed significantly on reading time, errors committed, and accuracy when reading sentences without diacritics in Phase One. For reading time, the Beginner group read slower than the Developing group by 1.03 seconds and slower than the Proficient group by 1.99 seconds. The Developing group also read slower than the Proficient group by 0.96 seconds. Errors decreased progressively, with the Beginner group committing 0.89–1.90 more errors than the other groups. Accuracy improved, with the Beginner group scoring 16.37–36.37% lower than the other groups. All

differences were statistically significant according to Tukey HSD post hoc tests. Performance improved clearly from Beginner to Developing to Proficient, with the Proficient group consisting of 15 top performers who advanced to the second phase.

Phase Two – Qualitative Results

Phase Two analyzed advanced readers reading Qur'anic verses, initially without diacritics. Hesitations, the need for diacritics, and self-corrections were recorded. Sentences were then provided with diacritics for ambiguous or complex words after hesitation, followed by the retrospective technique.

Table 4 . Word Recognition from Context in Undiacritized Qur'anic Texts

Verse	Word	Phonetic Transcription	Recognized?	Cue Used	Diacritics Needed?	Self-Correction?
75:22–23	ناضرة	nāḍirah	Yes	Context	No	No
75:22–23	ناظرة	nādirah	No	Limited context	Yes	Yes
35:28	العلماء	al-ʿulamāʾ	No	Minimal context	Yes	Yes
6:164	وزر	wizr	Yes	Context	No	No
12:31	أكبرنه	akbarnahu	No	Limited context	Yes	Yes
18:97	نقبا	naqba	No	Minimal context	Yes	Yes
17:61	أأسجد	a-asjudu	No	Minimal context	Yes	Yes
17:62	أرأيتك	araʾaytuka	No	Limited context	Yes	Yes
22:45	مشيد	mashid	No	Minimal context	Yes	Yes
22:45	معطلة	muʿaṭalah	Yes	Context	Minimal	No
2:260	فصرهن	faṣorhunna	No	Minimal context	Yes	Yes
80:31	فاكهة	fākihatin	Yes	Context	Minimal	No

The retrospective verbal protocols showed that readers sometimes struggle with Qur'anic words without diacritics. They reported hesitations, misreadings, and self-corrections when context was limited. In 75:22–23, ناضرة was recognized correctly because the context helped. But ناظرةcaused hesitation. Readers had to pause and self-correct. In 17:61 (أأسجد) and 17:62 (أرأيتك), the words were

difficult to read without diacritics. Readers said the orthography and limited context made them unsure of the correct reading. Words like العلماء (35:28) and مشيد (22:45) also caused difficulty. Participants reported that the limited context did not give enough clues, so diacritics were necessary for correct reading. In contrast, words such as وزر (6:164) and فاكهة (80:31) were read correctly because the context was clear.

Participants noted four main reasons for difficulty:

1. Semantic ambiguity (e.g., ناضرة vs. ناظرة)
2. Synatactic structure:العلماء
3. Limited context
4. Complex spelling (e.g., أرأيتك ,أأسجد)
5. Rare or less common words (e.g.عجاب , نقبا ,مشيد)

So, these results show that diacritics are crucial when context alone is not enough. The verbal reports helped identify where readers struggled and when they relied on extra cues to read correctly.

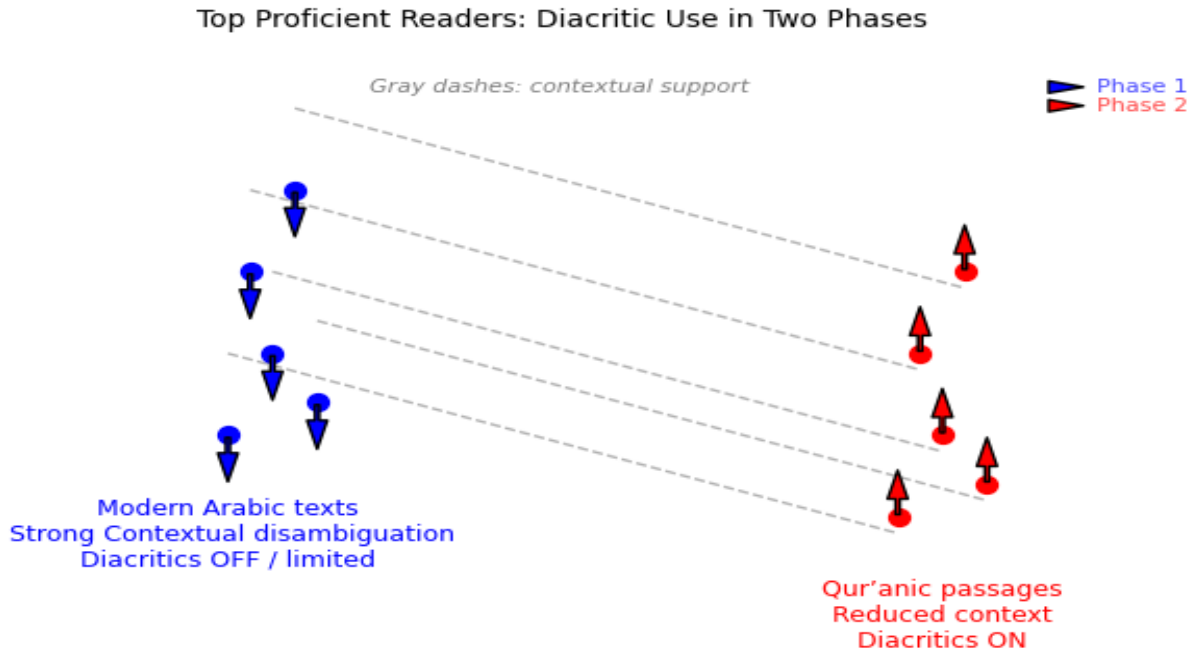

Figure 2 . Proficient readers' reliance on diacritics across two reading contexts

Figure 2 illustrates the reading strategies of fifteen highly proficient readers across two phases. In Phase 1, involving unpracticed Modern Arabic texts, readers rely primarily on contextual cues which result in low diacritic usage. In Phase 2, corresponding to Qur'anic passages with limited context, readers increase their reliance on diacritics. Gray dashed lines represent contextual support, which decreases as readers move to Phase 2 because Qur'anic passages provide less contextual

disambiguation, leading to greater diacritic usage. These findings indicate that even highly proficient readers adjust their approach according to text type: contextual cues alone may suffice in modern texts, whereas precise diacritic information becomes essential in non-memorized Qur'anic passages.

Discussion

The findings highlight the critical role of diacritics in Arabic word recognition, particularly in texts with high lexical ambiguity. In Phase One, which involved modern Arabic texts, beginners made more errors and had lower accuracy than intermediate and proficient readers. This suggests that when developmental stage is low, readers rely more heavily on diacritics to correctly identify words. Intermediate readers performed better but still benefited from diacritics for less frequent or ambiguous words. Proficient readers recognized most words correctly without diacritics, relying primarily on context. These results align with Hermena et al. (2021), who found that diacritics provide vowel information that disambiguates homographs and supports accurate word recognition, particularly in later sentence processing. They also align with Abu Rabia and Siegel (1995), who reported that both skilled and poor Arabic readers improved accuracy when reading words with context, and that skilled readers benefited more from context than poor readers. Importantly, their study shows that diacritics (vowels) improve reading accuracy for all readers, especially when context is limited. This explains why beginners in our study depended more on diacritics, while proficient readers could rely more on context.

In Phase Two, with Qur'anic passages, even proficient readers depended heavily on diacritics. Context alone was often insufficient due to rare, archaic, or specialized vocabulary. This demonstrates that diacritics are essential for accurate first-pass reading in classical texts, regardless of reader skill.

Conclusion and Implications

This study shows that diacritics are essential for accurate word recognition in Arabic, especially in Classical and Qur'ānic texts. Skilled readers can use context to recognize some words without diacritics, but words with high phonological ambiguity still require them. The more ambiguous a word is and the weaker the available cues, the more critical diacritics become. These findings have practical relevance for literacy instruction, curriculum design, assessment, and reading of classical or religious texts.

Key questions for future research include whether diacritics can be selectively introduced to improve reading in different types of texts, such as Qur'anic passages, or whether they are unnecessary for certain readers. It is also important to determine which diacritics are most critical for accurate word recognition: shaddah (consonant doubling) or short vowels (ḥarakāt). Another question concerns those

who memorize the Qur'an: do they rely on diacritics, and if so, are diacritics a learned habit rather than a functional necessity? This is particularly relevant because the Qur'an exists in only one fixed version, unlike other texts that can be modified or rewritten. Answering these questions will allow future studies to move beyond single-word recognition and explore how selective use of diacritics interacts with context and semantic expectations in authentic Qur'ānic passages. This is to guide educators, curriculum developers, and NLP designers. For Classical and Qur'ānic texts, diacritics are essential. This benefits both non-native and native speakers to support accurate word recognition.

Limitations

The current study has several limitations. First, the reported percentages of cue usage and recognition rates for advanced readers are not directly supported by previous research. Second, reading in actual Qur'ānic or classical texts was not fully measured, which limits generalizability. Third, the contribution of syntax, morphology, and other factors to word recognition is not clearly quantified. Future studies could investigate which cues are most influential across text types:

- Classical/Qur'ānic texts: Diacritics > Morphology > Syntax > Context
- Modern texts: Context > Morphology > Syntax > Diacritics

Fourth, the retrospective design relies on participants' reflections, which may introduce bias. Finally, purposive sampling, while suitable for targeting specific reader profiles, limits generalization to all Arabic readers.

References

Abu-Rabia, S. (1995). Reading in Arabic orthography: Phonological, syntactic, working memory skills in normally achieving and poor Arabic readers. *Reading Psychology, 16*(4), 351–394. https://doi.org/10.1080/0270271950160401

Abu Rabia, S., & Haj, A. (2021). The impact of diacritic marking on listening comprehension skills in Arabic orthography. *The Journal of Educational Research*, *114*(1), 64–73. https://doi.org/10.1080/00220671.2021.1872476

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), 1-19. https://doi.org/10.1080/0270271950160101

Abu-Rabia, S., & Taha, H. (2006). Phonological errors predominate in Arabic spelling across grades 1-9. *Journal of Psycholinguistic Research*, 35(2), 167–188. https://doi.org/10.1007/s10936-005-9010-7

Al-Azami, M. M. (2020). *The history of the Qur'anic text from revelation to compilation: A comparative study with the Old and New Testaments* (2nd ed.). Turath Publishing.

Alhussein, A. (2017). *The effect of printed word attributes on Arabic reading* (Doctoral dissertation, Lancaster University). Lancaster University. https://doi.org/10.17635/lancaster/thesis/267

Al-Samarraie, H., Sarsam, S. M., Alzahrani, A. I., & Alalwan, N. (2020). Reading text with and without diacritics alters brain activation: The case of Arabic. *Current Psychology, 39*, 1189–1198. https://doi.org/10.1007/s12144-019-00493-6

Ammar, M. (1997). *Les stratégies d'identification des mots écrits en arabe*. (Doctoral dissertation, Université de Nantes, France). Université de Nantes. https://www.sudoc.fr/043841376

Azmi, A. M., & Aljafari, E. A. (2018). Universal web accessibility and the challenge to integrate informal Arabic users: A case study. *Universal Access in the Information Society, 17*(1), 131–145. https://doi.org/10.1007/s10209-017-0522-3

Azmi, A. M., & Almajed, R. S. (2015). A survey of automatic Arabic diacritization techniques. *Natural Language Engineering, 21*(3), 477–495. https://doi.org/10.1017/S1351324913000284

Azmi, A. M., & Alsaiari, A. (2014). A calligraphic-based scheme to justify Arabic text improving readability and comprehension. *Computers in Human Behavior, 39*, 177–186. https://doi.org/10.1016/j.chb.2014.07.003

Azmi, A. M., Al-Qabbany, A. O., & Hussain, A. (2019). Computational and natural language processing-based studies of hadith literature: A survey. *Artificial Intelligence Review, 52*(2), 1369–1414. https://doi.org/10.1007/s10462-019-09692-w

Bashir, M. H., Azmi, A. M., & Nawaz, H. (2023). Arabic natural language processing for Qur'anic research: A systematic review. *Artificial Intelligence Review, 56*(7), 6801–6854. https://doi.org/10.1007/s10462-022-10313-2

Boudelaa, S. (2014). Is the Arabic mental lexicon morpheme-based or stem-based? Implications for spoken and written word recognition. In E. Saiegh-Haddad & R. M. Joshi (Eds.), Handbook of Arabic literacy (Vol. 9, pp. 23–45). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8545-7_2

Boulhrir, T., & Chekayri, A. (2025). From divine revelation to human inscription: Modernizing the Qur'anic orthography for pedagogical and translation purposes. In A. Chekayri (Ed.), *Connecting language, literacy, and cognitive development: Multidisciplinary approaches to education* (pp. 19–52). Al Akhawayn University Press. https://doi.org/10.5281/zenodo.15103570

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). The MIT Press.

Hermena, E. W., Bouamama, S., Liversedge, S.P. & Drieghe, D. (2021). Does diacritics-based lexical

disambiguation modulate word frequency, length, and predictability effects? *PLOS ONE, 16*(11), 1-27. https://doi.org/10.1371/journal.pone.0259987

Hermena, E.W. & Reichle, E.D. (2020), Insights from the study of Arabic reading. *Language and Linguistics Compass, 14* (e12400), 1-26. https://doi.org/10.1111/lnc3.12400

Khadka, B. K. (2024). Interplay of learning theories in the development of reading skills. *Journal of Tikapur Multiple Campus*, *7*(1-2), 121–141. https://doi.org/10.3126/jotmc.v7i1-2.63185

Koda, K. (2005). Word recognition. In *Insights into second language reading: A cross-linguistic approach* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9781139524841.005

Maroun, M. (2018). *Diacritics and the resolution of ambiguity in reading Arabic* (Doctoral dissertation, University of Essex). Retrieved from https://repository.essex.ac.uk/22078/

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*(3), 319–335. https://doi.org/10.1007/s11145-016-9677-1

Maroun, M., & Hanley, J. R. (2019). Are alternative meanings of an Arabic homograph activated even when it is disambiguated by vowel diacritics? *Writing Systems Research, 11*(2), 203–211. https://doi.org/10.1080/17586801.2020.1798327

Midhwah, A. A., & Alhawary, M. T. (2020). Arabic diacritics and their role in facilitating reading speed, accuracy, and comprehension by English L2 learners of Arabic. *The Modern Language Journal, 104*(2), 418–438. https://doi.org/10.1111/modl.12642

Peleg, O., & Eviatar, Z. (2009). Semantic asymmetries are modulated by phonological asymmetries: Evidence from the disambiguation of homophonic versus heterophonic homographs. *Brain and Cognition, 70*(1), 154–162. https://doi.org/10.1016/j.bandc.2009.01.007

Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge University Press.

Saiegh-Haddad, E., & Joshi, R. M. (2014). *Handbook of Arabic literacy: Insights and perspectives*. Springer. https://doi.org/10.1007/978-94-017-8545-7

Van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating Memory-Based and Constructionist Processes in Accounts of Reading Comprehension. *Discourse Processes*, *39*(2–3), 299–316. https://doi.org/10.1080/0163853X.2005.9651685