

An Evaluation of the Conceptual Validity of Vowel Blindness as an Explanation for Differences in Arabic Readers' L2 Word Recognition



Ndrew Allmark

Abstract In an influential study, Ryan and Meara (Read Foreign Lang. 7(2):531–540, 1991) posited that errors in the L2 reading and writing of L1 Arabic learners could be due to vowel blindness: a reduced sensitivity to written vowels deriving from the learners' L1 Arabic reading experience. Vowel blindness is frequently cited as a cross-linguistic effect influencing the L2 reading and writing outcomes of Arabic learners, yet its conceptual validity has rarely been scrutinised. This article evaluates the validity of the theory as an explanation for differences in L1 Arabic readers' written word recognition of languages with alphabetic writing systems. The empirical studies included in this review were identified in a systematic scoping review of Arabic L2 word recognition of alphabetic writing systems (reported in Allmark N. A systematic scoping review of evidence pertaining to L2 word recognition among L1 readers of Arabic and its implications for the validity of vowel blindness. Unpublished master's thesis, University of Oxford, Oxford, UK, 2019) whose methodology is summarised in this chapter. The review demonstrates that the current, published evidence is too limited and conflicting to validate vowel blindness. Furthermore, the included studies have methodological weaknesses that limit the overall trustworthiness of their findings. Further research is therefore needed before this phenomenon can be accepted as a factor in poor word recognition among Arabic learners. Recommendations are made for future avenues of research that could improve our understanding of either word recognition or the validity of vowel blindness, and for raising the methodological standards of research in this field.

Keywords L2 reading · Word recognition · Cross-linguistic effects · Vowel blindness

N. Allmark (✉)
Sultan Qaboos University, Muscat, Oman

University of Oxford, Oxford, UK
e-mail: ndrew@squ.edu.om

© The Author(s), under exclusive license to Springer Nature
Switzerland AG 2022

R. Al-Mahrooqi, C. J. Denman (eds.), *Individual and Contextual Factors in the English Language Classroom*, English Language Education 24,
https://doi.org/10.1007/978-3-030-91881-1_7

1 Introduction

Arabic is the fourth most widely-spoken language in the world and the second most widely used phonemic script after Roman (Saiegh-Haddad & Joshi, 2014). L1 Arabic speakers constitute a large population of L2 language learners. However, it has been observed that L1 Arabic learners exhibit particular difficulty in developing L2 reading skills including slower and less accurate word recognition (e.g. Masrai & Milton, 2018; Saigh & Schmitt, 2012). As accurate and efficient word recognition is a prerequisite for successful reading comprehension (Grabe, 2009; Nassaji, 2014), identifying the causes of this pattern should be a priority for educators working with Arabic learners.

It is widely believed that cognitive processes used in L1 reading transfer to L2 reading resulting in systematic differences in the reading development and outcomes of L1 language groups (Koda, 2005). The weak L2 word recognition observed among Arabic learners could therefore be related to features of the L1, its writing system and the learners' L1 reading experience. As a result, an improved understanding of cross-linguistic transfer deriving from L1 Arabic reading experience could make a valuable contribution to our understanding of Arabic learners' L2 reading development.

There are many features of the Arabic language that could affect second language reading. However, in the English Language Teaching (ELT) literature it is the lack of written short vowels that has received the most attention. Arabic is a consonantal writing system whose letters only represent consonants and long vowels; any indication of short vowels is typically omitted. This led Ryan and Meara (1991) to propose that Arabic learners may be less sensitive to L2 written vowels due to their L1 reading experience. They labelled this phenomenon vowel blindness and suggested that it could explain a range of receptive and productive errors, including slower and less accurate word recognition. Vowel blindness has since been frequently cited (e.g. Koda, 1996, 2005; Ryding, 2013); however, its validity remains largely untested, and most educators of Arabic learners are unaware of the extent or strength of the empirical evidence supporting the phenomenon.

This chapter aims to examine the empirical evidence base for vowel blindness as an explanation for Arabic learners' L2 English word recognition performance. It describes key features of the Arabic language and writing system drawing upon Saiegh-Haddad's (2018) Model of Word Reading in Development (MAWRID) to identify factors affecting L1 reading development. Dual-route models of word recognition (Coltheart et al., 2001) and the Orthographic Depth Hypothesis (ODH) (Katz & Frost, 1992) are used to highlight the potential impact of linguistic and orthographic distance on L2 reading development and outcomes. These theories frame a discussion of vowel blindness (Ryan & Meara, 1991) and its theoretical and empirical foundations. Published empirical evidence pertaining to vowel blindness was identified through this author's systematic scoping review of Arabic L2 word recognition of alphabetic writing systems (Allmark, 2019). The methodology of the scoping review is summarised, and the findings that pertain to vowel blindness are presented, evaluated and discussed.

2 Literature Review

The Arabic alphabet contains 28 letters which correspond to the 28 consonant phonemes of Modern Standard Arabic (MSA), the formal, standardised register of the language. The mapping of phonemes and letters in MSA is highly regular; the primary exceptions being three letters which act as *matrēs lectionis*, letters that can represent either a consonant or an associated long monophthong vowel (Daniels, 2013). As MSA has only three long monophthong vowels, the 28 letter alphabet is sufficient to represent all the consonants and long, monophthong vowels of the phonemic system in a very regular manner. MSA also has three short vowels which do not have associated letters; diacritics can be used to represent these short vowel sounds, but the symbols are omitted from the vast majority of texts. There are also two diphthongs which occur due to the blending of adjacent vowels and therefore do not require their own graphemes.

Despite the regularity of the sound-spelling relationship, L1 Arabic word recognition is typically slower than word recognition in other scripts, even among skilled native speaking readers of Arabic (Ibrahim & Eviatar, 2012). Myhill (2014) further observes that the basic literacy rate of every Arab nation is lower than would be predicted from its GDP. Arabic nations are typically 47 places lower in UNESCO's global rankings of basic literacy than countries with similar size economies.

A number of explanations have emerged to explain this low level of reading attainment, including those based on features of the Arabic language and observations of L1 Arabic reading development. In her MAWRID framework, Saiegh-Haddad (2018) identifies three aspects of the Arabic language and its orthography that influence L1 Arabic reading development: vowelisation, morphological structure, and diglossia. Furthermore, a number of writers identify the visual complexity of the Arabic script as a cause of slower L1 word recognition (Eviatar & Ibrahim, 2014; Jordan et al., 2011).

2.1 Vowelisation

Arabic texts can be written in one of two forms: the vowelised form features diacritic symbols which are added to indicate short vowels, gemination (consonant lengthening) and the vocalic and consonantal case-endings of formal MSA, while the unvowelised form omits these symbols. Children learn to read using vowelised script before moving onto the unvowelised script in the third or fourth grade of primary school (Fender, 2008). In later stages, Arabic readers rarely encounter vowelised texts as the vast majority of authentic texts for adults are unvowelised with the notable exception of certain religious texts (Alhawary, 2011).

As Saiegh-Haddad (2018) describes, early readers learn to decode a very complete written representation of a word using both letters and diacritics as part of a grapheme-based phonological recoding mechanism. To adjust to the less complete

unvowelised text, the learners must adopt a letter-based morpho-orthographic recoding mechanism in which knowledge of morphology is used to support word recognition.

By not representing the short vowels of words, unvowelised Arabic has greater scope for homography, in which a single orthographic form can represent more than one meaning and pronunciation. For instance, the written form ك can represent the pronunciation /ʕilm/ (knowledge) or /ʕalam/ (flag). This prevalence of homographs is thought to be problematic for L1 Arabic reading affecting both skilled and poor readers (Abu-Rabia, 1997a). To resolve homography, readers are thought to draw upon the broader context (Fender, 2008) and apply knowledge of morphology and word frequency (Hansen, 2010). Arabic vowelisation therefore affects the degree to which readers attend to different types of information while reading leading to processing preferences that could affect how Arabic learners read in an L2.

2.2 Morphology

The morphology of Arabic is also believed to affect how Arabic learners read. Arabic has a root-based system of morphology in which Arabic content words are formed from two bound morphemes: a root and a word pattern (Abu-Rabia & Taha, 2006). The root is a fixed sequence of three, or occasionally four, letters which imply a core, general semantic meaning associated with a family of words. The root combines with a word pattern to complete the word's phonological form. An advantage of encoding only consonants and long vowels is that it facilitates the identification of the word root which L1 Arabic readers are thought to utilise during word recognition (Saiegh-Haddad, 2018).

The word pattern provides a prosodic template which includes short vowels, syllabification and any required gemination. It also indicates word-class, and the person, number, gender and tense of verbs. Word patterns may be solely comprised of features of pronunciation not encoded in unvowelised Arabic; two words with different pronunciations may therefore share the same spelling. This contributes to the homograph phenomenon described above. Patterns may also include consonantal or long vowel affixes, including prefixes, infixes and suffixes (Boudelaa, 2014; Milin et al., 2018) which are always written as letters. The addition of consonantal or long vowel affixes never affects the order of the root letters, only their proximity to one another. It seems clear that Arabic learners utilise morphological information in L1 reading in a manner that is qualitatively different from readers of non-Semitic languages, and cross-linguistic approaches would assume that this affects their processing of L2 texts.

2.3 *Diglossia*

Diglossia has been implicated as a cause of relatively low patterns of L1 literacy observed throughout the Arab world (Saiegh-Haddad & Spolsky, 2014). Diglossia is a situation in which a so-called high (or H) register and at least one low (or L) spoken register co-exist (Ferguson, 1959). Arabic has numerous spoken varieties which differ from MSA in terms of grammar, lexis and phonology. These spoken Arabic varieties are acquired naturally by Arabic speakers (Saiegh-Haddad, 2018). They are unstandardised in both speech and writing, and they are extremely diverse.

MSA is the H-register of Arabic; it is a highly formalised variety, derived from Classical Arabic, and it has a standardised writing system. MSA must be explicitly taught, particularly in its most formal forms which apply case-endings to content words to mark grammatical function. It is also the language of reading instruction for Arabic children, despite these readers not being orally fluent in the register (Eviatar & Ibrahim, 2014).

Saiegh-Haddad (2018) estimates that only around 20% of an Arabic child's spoken vocabulary is identical to MSA, 40% are nonstandard words with no established written form, and 40% are cognates with related but divergent pronunciations. Diglossia may also complicate parents' decisions when trying to support the development of their children's L1 literacy (Korat et al., 2014).

2.4 *Visuo-Graphemic Complexity*

Several writers stress that the visuo-graphemic complexity of Arabic script may negatively affect letter recognition and L1 Arabic reading fluency (Eviatar et al., 2004; Jordan et al., 2011). Although there are 28 Arabic letters, these are formed from 18 letter shapes which can only be distinguished by the addition of dots above, within or below the letters (Daniels, 2013). The script is always written cursorily with most letters joining to their neighbours within a word; as a result, letters can appear in a word-initial, word-medial, word-final or independent position. Letters can thus have up to four allographs determined by their position within a word (Daniels, 2013; Simon et al., 2006), resulting in around 97 contextually determined forms within the alphabet as a whole (Cook & Bassetti, 2005). Furthermore, the cursive text leads to crowding, which can further complicate letter recognition (Jordan et al., 2011).

In summary, reading development in Arabic learning is shaped by features of the language and its writing system. This is made slower and more difficult by the diglossic registers of the language and the orthographic complexity of the script. Arabic children must also adjust from reading vowelised text to reading unvowelised text, consequently supplementing or modifying their cognitive reading processes.

2.5 *L2 Literacy Development*

It is generally assumed that similarity between the orthographies of two languages facilitates L2 reading development, while greater orthographic distance has a negative effect (Han, 2015). Katz and Frost's (1992) ODH is widely used to describe and compare writing systems. The authors consider orthographic depth in terms of three qualities of a writing system: regularity, consistency and completeness. In regular writing systems the written form of words follows grapheme-correspondence rules; consistency relates to whether graphemes have unique phoneme-correspondence (i.e. whether spelling patterns can represent single or multiple phonemes); while completeness relates to the amount of phonological information encoded in a written word form (Brown & Haynes, 2005; Frost, 2005).

The ODH assumes a dual-route model of word recognition; such models constitute the dominant paradigm in visual word recognition research (Frost, 1998). Dual-route models presuppose two procedures (or routes) through which written words may be recognised: the non-lexical procedure and the lexical procedure (Coltheart et al., 2001). The non-lexical procedure decodes a word's phonological form by applying rules of letter-to-sound correspondence to visual input, while the lexical procedure uses the whole visual sample of a word to identify its phonological form in the mental lexicon (Coltheart et al., 2001).

These two procedures are effectively in competition with one another, and there are a number of factors that influence which of the routes is utilised. Irregularly spelled words require use of the lexical procedure, as the application of phonological rules would lead to misidentification of the word's phonological form. As the lexical procedure can only identify words whose orthographic forms are familiar to the reader, unknown words and pronounceable letter-strings could not be processed through this route and would require non-lexical decoding. The lexical procedure is believed to be faster, and it is assumed that frequently encountered words are more likely to be recognised through the lexical route, and reading experience is therefore thought to lead to greater use of the lexical procedure.

The ODH claims that shallow writing systems, those which are regular, consistent and complete, allow for greater and more efficient use of phonological decoding, and hence the non-lexical route is used to a greater extent. In deep orthographies, the irregularity, inconsistency and incompleteness of the written form prevents the efficient use of non-lexical processes, leading to greater and more effective use of the lexical route. As a result, distinct processing preferences emerge during L1 reading development, and these are thought to influence the development of L2 reading (Katz & Frost, 1992; Koda, 2005).

2.6 *Arabic and English*

All studies of vowel blindness known to this author focus on L2 English reading. Both English and unwowelled Arabic can be labelled as deep orthographies (Perfetti & Verhoeven, 2017), but they differ in their consistency, regularity and completeness.

The alphabetic orthography of English encodes both long and short vowels and thus has a higher level of completeness from an ODH perspective than unvowelised Arabic; however, Arabic is both more regular and consistent than English. While the relationship between phonemes and letters in Arabic is generally one-to-one, in English the spelling of many words conflicts with established sound-spelling conventions. One phoneme can be represented by several graphemes (such as the spelling of /k/ as <c> or <k>), and one grapheme can indicate several phonemes (such as the sounds represented by <ow> in 'snow' /əʊ/ and 'how' /aʊ/). However, the ODH does not explicate the relative contributions of these three elements in shaping L1 reading development.

In addition to the elements of orthographic depth, the two languages differ in terms of their phonemic systems, and in the lengths of the graphemes used to encode written words. English has a larger overall phonemic inventory than Arabic due to a greater number of vowels. MSA has eight vowels including its two diphthongs; while Standard British English has 20 vowels consisting of 12 monophthongs and 8 diphthongs (Perfetti & Verhoeven, 2017). Phonological awareness is a predictor of reading success in both languages (Perfetti & Verhoeven, 2017), and a lack of L2 English phonological knowledge and awareness is likely to exacerbate the difficulties caused by the irregularity and inconsistency of English spellings.

Finally, while Arabic graphemes are virtually all single letters, English utilises many multigraphs and these can include the use of consonant letters in the spelling of vowel sounds (for example, <oy> in 'toy', or <igh> in 'night'). Consequently, L1 English word recognition utilises a number of processing strategies, such as multiple-letter analysis and awareness of neighbour-frequency effects, which Arabic learners are unlikely to have developed through L1 reading experience (Hansen, 2014).

This discussion has identified a range of features of L1 Arabic that could affect L2 reading, including, but not limited to, the readers' experience of unvowelised script. The comparison of the two languages illustrates some key differences, while highlighting some of the limitations of the ODH. We should now turn our attention to the claims and theoretical foundations of vowel blindness.

2.7 *Vowel Blindness*

Ryan and Meara's (1991) theory of vowel blindness attributes a range of reading and writing errors among Arabic learners to their L1 reading experience. It purports that L1 reading experience of unvowelised Arabic and its consonantal writing system leads to a decreased sensitivity to vowels in alphabetic writing systems, including poorer word recognition. However, despite providing a speculative theoretical outline of vowel blindness, Ryan and Meara's seminal article did not provide empirical evidence that could adequately support the theory. Their report contained anecdotal discussion of productive spelling errors made by the researchers' Arabic students, and it reported an empirical word-matching experiment.

The experiment focused on the performance of three groups of participants: English for Academic Purposes (EAP) students of L1 Arabic backgrounds, EAP students of non-Arabic L1 backgrounds, and a group of English native speaker teachers. The Arabic students were described as being of lower-intermediate to intermediate level proficiency, and their counterparts were described as being of a “comparable” proficiency level. Participants were instructed to decide whether two spellings, displayed consecutively, were identical. One vowel was removed from each erroneous spelling; however, the correct consonants were preserved in all stimuli. The speed and accuracy of the responses were taken as indicative of the participants’ L2 word recognition ability. The Arabic group demonstrated the slowest and least accurate responses suggesting a general weakness in L2 word recognition. However, in the absence of a consonant error condition, the results could not demonstrate a difference in the groups’ processing of vowels and consonants. Ryan and Meara (1991) acknowledged this in their discussion and called for future research to explore vowel blindness further. However, they proceeded to draft a diagnostic test to identify vowel blindness (Ryan & Meara, 1996).

Vowel blindness was postulated before extensive research had been carried out to explore the role of short vowels in L1 Arabic reading development (Alghamdi, 2015). Since then, a relatively extensive body of L1 Arabic research has emerged that explores L1 vowel processing. Abu-Rabia (1996, 1997b, 1998; Abu-Rabia & Siegel, 1995) has consistently observed that the presence of vowel diacritics facilitates word reading accuracy among L1 Arabic readers. Furthermore, the addition of incorrect vowel diacritics lowered participants’ accuracy in an Arabic read-aloud task (Abu-Rabia, 1998). This suggests that Arabic participants are not only capable of making use of short vowel information in L1 reading, but they also struggle to inhibit their processing of this information, despite years of experience reading unvowelised texts. As Alghamdi (2015) suggests, if a desensitivity to short vowel information cannot be demonstrated in the L1, it may be unreasonable to assume its presence in L2 reading.

Vowel blindness and Ryan and Meara’s (1991) study are frequently cited in discussions of Arabic L2 learners (see Koda, 1996, 2005; Ryding, 2013); however, there has not been adequate discussion of its validity. This is partly because published studies exploring word recognition among L1 Arabic readers are spread over numerous journals and publications including many with low impact, or the ideas are explored in theses and unpublished literature that are less easily accessible. Educators of Arabic learners tend to be unaware, therefore, of the extent or strength of the empirical evidence supporting the phenomenon.

3 Identification of Included Studies

The studies discussed in this chapter were identified in a recent systematic scoping review (reported in Allmark, 2019). Systematic reviews follow rigorous, replicable and transparent methodologies to reduce bias and ensure that relevant evidence is not excluded arbitrarily (Siddaway et al., 2019). Scoping reviews are conducted

when reviewers aim to describe the extent and nature of a field of research (Gough et al., 2012). The recent review aimed to provide a map of the published evidence pertaining to L2 word recognition among Arabic learners reading in alphabetic writing systems, guided by the following questions:

1. What is the published evidence pertaining to the word recognition processes of L1 Arabic readers engaged in L2 reading of non-consonantal, alphabetic writing systems?
2. According to the literature identified in RQ1, what factors are identified as influencing L1 Arabic readers' L2 word recognition processes, and what is the nature and extent of their contributions to word recognition?

Studies identified in the scoping review that pertained to the relative processing of vowels and consonants in L2 reading were then used to answer the third research question:

3. To what extent does existing evidence support the conceptual validity of vowel blindness?

The primary method of identifying studies was through electronic searching of four online bibliographic databases covering a broad range of journals and dissertations in the field of education and linguistics:

- ProQuest Linguistics Collection
- MLA International Bibliography
- Web of Science Core Collection
- Scopus

However, several additional methods were also used to overcome any unforeseen limitations to the electronic search strategy (as recommended by Brunton et al., 2012). These were:

- backwards citation of included studies: i.e. searching the reference lists of studies selected for inclusion in order to find potentially relevant articles;
- forward citation of included studies: i.e. using additional electronic searches to identify studies which cite the articles already selected for inclusion;
- studies that were known, or made known, to the author.

A search string was piloted and refined iteratively to help ensure that it was both sensitive enough to locate the maximum number of relevant records, and precise enough to reduce the number of irrelevant studies to a manageable amount. The final search string contained three fields: (1) words related to word recognition, including terms related to vowel-blindness, component sub-processes of word recognition, and empirical measures of word recognition; (2) words beginning with 'Arab*'; and (3) terms related to second language learning.

The electronic database search identified 1812 candidate sources whose bibliographic data were uploaded to the online systematic review platform Rayyan (Ouzzani et al., 2016). Rayyan identified potential duplicate entries, of which 150 were confirmed and excluded by the reviewer leaving 1662 studies. The reviewer

and a research assistant double-screened 10% ($n = 166$) of the titles and abstracts by applying the eligibility criteria detailed in Table 1 using a cautious, over-inclusive approach as recommended by Petticrew and Roberts (2006) to prevent the arbitrary exclusion of relevant studies. The inter-rater reliability level was 90.4% and any disagreements were resolved through discussion. The remaining titles and abstracts were then screened by the main reviewer.

Ultimately, 1531 entries were excluded and 131 potential entries were retained. The author attempted to obtain full-texts for these reports using the Bodleian library online system, inter-library requests, inter-library loans and direct requests to authors. Double-screening was undertaken for 10% of the retrieved full-texts; however, this led to an unacceptably high level of disagreement (23%). The majority of the disagreements related to whether or not the study was exploring L2 word recognition. To overcome this, an addendum was created that stipulated that studies focusing on the speed, accuracy or nature of the processing of words, pseudowords, word-parts or lexical chunks were to be included; while studies were excluded that solely focus on the comprehension of whole clauses, sentences or texts, or which use surveys to indirectly explore reading processes without an accompanying reading task. All of the full versions of the texts retrieved were then double-screened and the disagreement rate was below 12%. This was considered adequate, and disagreements were resolved in a meeting. Any remaining texts that had not yet been retrieved were later single-screened by the main reviewer.

Table 1 Eligibility criteria

	Include/Exclude criteria
Bibliographic information	Include 1: Complete or sufficient bibliographic information. Exclude 1: Insufficient bibliographic information for text retrieval.
Focus of study	Include 2: Explores L2 word recognition of alphabetic (and non-Abjad) writing systems. Exclude 2: Research that does not explore L2 word recognition, or only explores L2 word recognition of syllabic, logographic, or alphabetic-abjad writing systems.
Outcomes	Include 3: Includes direct measures of L2 word recognition. Exclude 3: Solely based on the analysis of secondary data, or does not include measures of L2 written word recognition.
Participants	Include 4: Includes analysis of one or more distinct groups of Arabic learners. Exclude 4: Participants do not include L1 Arabic learners. Include 5: Adult (18 years) language learners/users. Exclude 5: Non-adult L2 language learners/users. Include 6: Participants are literate in L1 Arabic. Exclude 6: Participants are explicitly described as not being literate in their L1, or as having low L1 literacy. Include 7: Participants are typically developing (i.e. the entire sample does not have learning difficulties, language disorders, or sensory disorders). Exclude 7: Participants are described as having specific learning difficulties, language disorders, or sensory disorders.

A further 92 texts were identified through other means: backwards citation, forward citation searching using the ‘cited references search’ function of Web of Knowledge, or by being known by, or made known to, the author. The full texts of six studies could not be retrieved, and alternative write-ups were retrieved for a further two texts. Ultimately, 49 studies were included in the final synthesis, and the review revealed that of these only seven explored the relative processing of vowels and consonants by Arabic participants or attempted to address behaviours associated with vowel blindness. The full process of identifying and screening articles is summarised in the PRISMA diagram below (see Fig. 1).

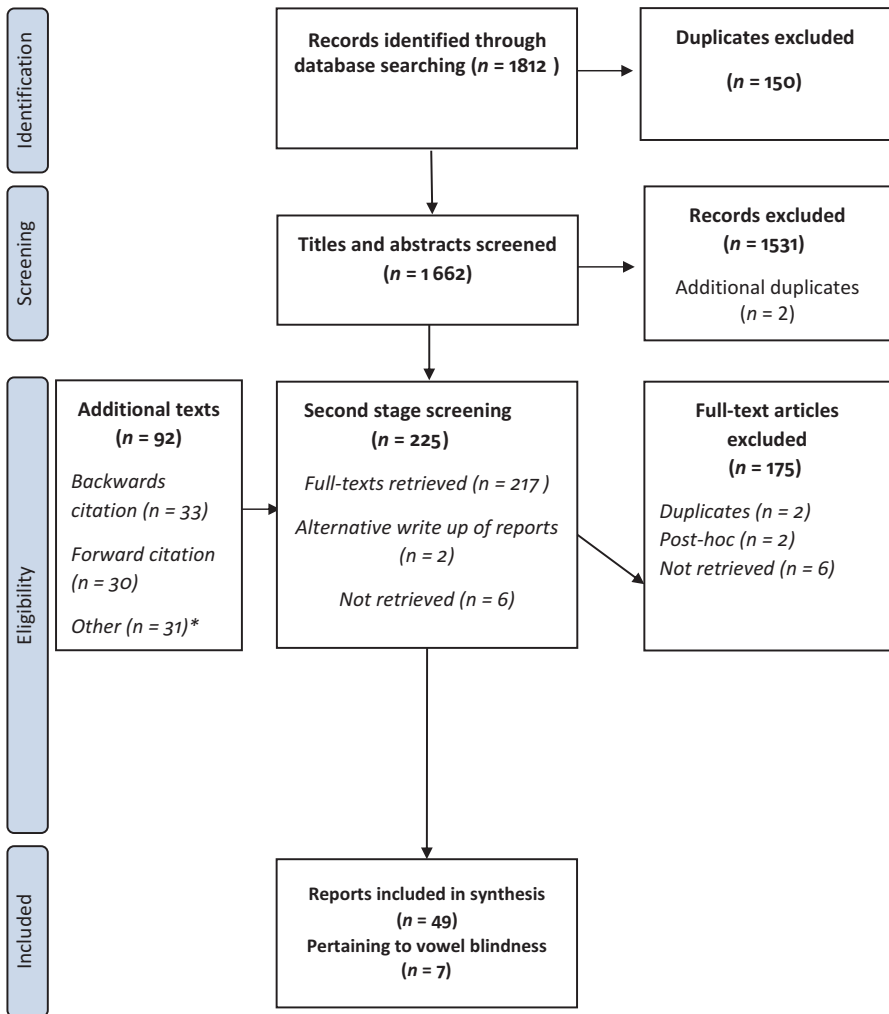


Fig. 1 PRISMA diagram of searching and screening

*The *Other* figure includes the two additional texts discussed in the chapter, bringing the total to 94

4 Quality Assessment

The Quality Assessment Tool for Quantitative Studies (Effective Public Health Practice Project, 1998) (hereafter the EPHPP instrument) was used to ensure a standardised and rigorous approach to the critical appraisal of the methodological standards of the quantitative studies included in the scoping review (Liabo et al., 2017). The instrument allows six categorical scores to be calculated based on closed questions, with these scores then used to allocate a global rating. Table 2 shows the EPHPP results for the seven studies exploring vowel blindness or vowel and consonant processing. This includes the seminal study by Ryan and Meara (1991), even though its methodology did not allow for the comparison of vowel and consonant processing.

The scores for the ‘selection bias’ category were affected by the apparent use of convenience samples in all these studies. Convenience sampling is common in linguistics for practical and financial reasons; however, this approach reduces the confidence with which findings can be generalised to wider populations. Studies that defined their populations broadly (e.g. Arabic or Arabic English as a Second Language (ESL) learners) and yet included narrow samples were marked low. Moderate scores were allocated to studies that defined their populations more precisely and avoided premature generalisations to wider populations.

In the study design category, case-controlled and quasi-experimental studies were ranked as moderate, while within-subjects studies were ranked low as the lack of a control group reduces certainty that the performance of participants derives from the causal factors being investigated. The EPHPP instrument identifies

Table 2 EPHPP instrument results

	Selection bias	Study design	Confounders	Blinding	Data collection methods	Withdrawals/ Dropouts	Global rating
Al Juhani (2015)	3	3	N/A	1	2	N/A	Weak
Alhazmi et al. (2019)	2	2	2	1	1	N/A	Strong
Alsadoon and Heift (2015)	3	2	3	1	3	3	Weak
Hayes-Harb (2006)	3	2	2	1	2	N/A	Moderate
Ryan and Meara (1991)	3	2	3	1	3	N/A	Weak
Saigh and Schmitt (2012)	2	3	N/A	2	2	N/A	Moderate
Stein (2010)	2	2	2	2	3	1	Moderate

randomised controlled trials as the strongest study design; however, none of the included studies used this design.

The case-controlled and quasi-experimental studies were also graded for how well relevant confounders were controlled for. The following potential confounders were identified based on the researcher's background research: gender, proficiency/L2 experience, reading comprehension, age, regional background or spoken Arabic variety, education level, L1 print experience, language of prior education or primary language of literacy, length of residency in host country, and pre-intervention scores. For each study, a percentage was calculated based on the proportion of relevant confounders from this list that were controlled. No studies controlled for the 80% of relevant confounders required for the high score. Studies that controlled for less than 60% of relevant confounders were allocated low scores.

Blinding scores were calculated based on the assessor's reported awareness of the participants' status as experimental or control, and the participants' awareness of their own status and of the goals of the research. Computerised data collection instruments help to ensure objectivity in the collection of data such as speed and accuracy of responses or eye-movements; hence, the use of such instruments positively affected a study's score. Studies were also given more positive scores if there was no evidence that participants were told the purpose of the experiment, and the instructions to participants were clearly reported.

The data collection scores are based on the evidence for the validity and reliability of the data collection instruments. The validity of instruments was typically justified through references to the literature. Reliability, however, was rarely discussed or evidenced thus weakening the trustworthiness of the studies' findings. Scores for withdrawals and dropouts were only assigned to those studies which required participants to attend more than one session of data collection or learning.

The EPHPP instrument also includes questions related to intervention integrity and the statistical analyses used, though these answers do not affect the scores calculated by the instrument. Statistical analyses were generally appropriate; the main exception was Stein (2010) who decided to exclude non-Egyptian Arabic participants from her analysis and yet drew conclusions pertaining to Arabic speakers in general.

The EPHPP instrument results indicate that we should exercise caution when drawing conclusions from the included studies, especially when generalising to the large and diverse global population of Arabic language speakers learning English. There remains considerable scope for improvement in the methodological rigour of vowel blindness studies. The scores also show the variation in the trustworthiness of the studies' findings. It is important that these are considered as we discuss the specific methodologies and findings of the individual studies.

5 Evaluation of Individual Studies

The included studies used a range of methodological instruments to explore vowel and consonant processing. However, the word-matching task was the most commonly used instrument. Such tasks make intuitions regarding word recognition

processing based on the speed and accuracy of participants' judgements of the similarity between two stimuli. Ryan and Meara's (1991) seminal investigation used a word-matching task; however, as it only included vowel-error stimuli, the researchers were unable to draw conclusions regarding the participants' relative sensitivity to vowels and consonants.

In a partial replication, Hayes-Harb (2006) added a deleted consonant error condition to Ryan and Meara's (1991) task. She controlled the frequency of word stimuli more tightly by only including words with a frequency of over 100 per million. Her three groups of participants mirrored the groups of the original study: L1 Arabic, non-Arabic ESL and native speakers of English. All ESL participants were described as intermediate level, slightly higher than Ryan and Meara's lower intermediate participants, though neither study clearly describes how proficiency descriptions were ascertained.

It was hypothesised that the presence of vowel blindness would lead to faster and more accurate processing of consonant error conditions among Arabic participants. However, all three groups demonstrated faster responses to the deleted vowel condition than to the deleted consonant condition ($F(1,29) = 9.416, p < .01$, partial $\eta^2 = .25$) with no significant effect of language group or stimulus type on the accuracy of the ESL participants' responses. The experiment could therefore not provide evidence of vowel blindness. The researcher attributed this surprise result to the small sample size and the subsequent weaker statistical power of the analysis.

Hayes-Harb (2006) then conducted a second experiment, a letter detection task with 45 participants using the same three group types as the previous experiment. Four short texts were administered, and participants had 50 seconds to read for comprehension while circling all instances of either the letter 't' or 'o', before turning over the text to answer comprehension questions. It was hypothesised that the presence of vowel blindness would lead to a weaker performance in the 'o' letter condition. The Arabic participants' performance was similar when detecting target vowels and consonants; while the other groups were significantly more accurate when detecting the vowel 'o' than the consonant 't' ($p < .01$). The author stated that the findings could not be attributed to differences in general English word processing and suggested that this could be due to the cross-linguistic transfer deriving from the less prominent role of written vowel information in Arabic.

Hayes-Harb (2006) argued that the second experiment more closely resembled natural reading, and she queried the validity of the word-matching task for detecting vowel blindness as it does not require the reader to access the form or meaning of the word. However, the author acknowledged that the letter detection task had not been used previously to evaluate the amount of visual attention given to particular types of letters. Furthermore, the task required participants to complete two activities concurrently which may disadvantage readers with slower and less automatic word recognition, qualities commonly associated with Arabic participants.

Al Juhani (2015) also conducted a partial replication of Ryan and Meara's (1991) word-matching task with a larger sample of Arab participants ($N = 35$). Like Hayes-Harb's (2006) Arab participants, Al-Juhani's subjects' response times were faster in the vowel condition; however, their performance was significantly less accurate ($p < .001$) with a large effect size (partial $\eta^2 = .54$). Her results therefore partially

contradict the surprise findings of Hayes-Harb. The study lacked a non-Arabic comparison group. We therefore cannot be certain that results were not affected by confounding variables introduced through the stimuli or test administration.

The other studies in the review all used original designs. For instance, Saigh and Schmitt (2012) administered a pen and paper lexical decision task to Arab ESL students ($N = 24$) with an IELTS score of at least 4.5. Lexical decision tasks require participants to indicate whether a letter-string is a word. In this experiment, the stimuli incorporated two conditions: vowel length (long or short) and error type (correct vowel, incorrect vowel and missing vowel). This was the only study which distinguished between long and short vowels, reflecting differences in how these phonemes and graphemes are encoded and processed by speakers of Arabic as a first language.

The participants were significantly more accurate when recognising long vowel errors ($p < .05$). However, the authors did not consider the increased visual saliency of long vowels in written English. Long vowels are typically written with multi-graphs which are more noticeable. This may have contributed to the improved performance. Furthermore, the instrument could have been testing vocabulary knowledge and spelling, rather than word recognition. Words were selected from a list of the 4000 most frequent words in English. For students with the minimum English proficiency of IELTS 4.5, some of these words may have been unfamiliar. There was also no time limit for the test, and the pen and paper format disallows the collection of response times. The study also followed a within-subjects design with no non-Arabic comparison group. We cannot, therefore, be confident in the trustworthiness or generalisability of its findings.

Alhazmi et al. (2019) used eye-tracking technology to explore the validity of vowel blindness. Such technology allows researchers to measure the length of time spent fixating on individual words, letters or areas of a text, and it is assumed that longer fixation times indicate greater processing time (Stevenson, 2015). This methodology can be combined with natural, self-paced reading activities that are similar to real-life reading activities and which allow participants to approach texts using their individual reading strategies (Witzel et al., 2012).

Upper-intermediate and advanced Arab learners of English ($n = 30$) and a comparison group of native English speakers ($n = 20$) were administered a silent English word-reading task. As they read, the number and length of fixations on consonants, vowels, and words were measured. It was assumed that the presence of vowel blindness would lead to a lower amount of time fixating on vowels relative to consonants. The researchers observed that the Arab participants read at around half the speed of their English counterparts; however, the two groups spent similar proportions of their reading time fixating on consonants and vowels. The Arab participants actually demonstrated a slightly longer, yet statistically significant, fixation time for the vowels ($t(29) = 3.284, p = 0.003$). It was therefore concluded that there was no evidence for the transfer of an L1 desensitivity to vowel information. However, the Arab participants' high proficiency could have made the symptoms of vowel blindness less salient. L1 processing habits that inhibit L2 reading development are believed to diminish with L2 reading experience and proficiency (Koda, 2005).

Stein (2010) used a 'silent pronunciation task' to explore how consonantal context affected Arab participants' processing of vowels. Participants were instructed to silently read a target word before hearing the experimenter read two words aloud; they had to choose which rhymed with the written word. The spelling patterns investigated included onset-to-vowel associations at the beginnings of words and vowel-to-coda association at the ends. At the beginning of the study, there were 108 volunteers including 20 Arab speakers. However, an auditory discrimination task was used to filter out students whose phonological awareness was too weak for the main task. Only six Arab participants passed this test, of which five were Egyptian, and only the Egyptian Arab participants' data were used in the final statistical analysis. The author observed a non-statistically significant sensitivity to the consonantal constraints imposed by onset-to-vowel consonantal constraints, but no similar sensitivity for vowel-to-coda associations. However, it was ultimately acknowledged that the lack of data prevented a firm conclusion.

Finally, Alsadoon and Heift (2015) explored whether textual input enhancement could counter the supposed symptoms of vowel blindness. Thirty beginning level female students from Saudi Arabia completed a reading task as their eye-movements were recorded. For the experimental group, the target words and their vowels were made more salient through textual enhancement (bold text, underlining and red font). This group demonstrated significantly longer fixation times for target words and re-read them more frequently ($p < .001$). This correlated with improved performance in post- and delayed post-tests featuring the target words. The results suggest that the receptive and productive errors associated with Arab learners' L2 English literacy can be reduced or overcome with appropriate teaching and learning interventions. However, the study cannot show that these symptoms are the result of a desensitivity to vowel information, and do not, therefore, provide evidence regarding the validity of vowel blindness.

6 Discussion

The studies featured in this review support the observation that Arab L2 readers typically demonstrate slower and less accurate word recognition compared to other L1 groups. However, the evidence presented in the studies is too conflicting and limited to validate or refute the phenomenon of vowel blindness.

The most commonly used instrument was the word-matching task. However, the studies utilising this instrument (Al Juhani, 2015; Hayes-Harb, 2006; Ryan & Meara, 1991) provided conflicting results. Ryan and Meara's original study did not include a deleted consonant condition and could not therefore provide evidence of differences in the processing of written vowels and consonants, while the two replication studies, which featured separate vowel and consonant error conditions, failed to provide unambiguous or concordant evidence supporting vowel blindness.

Hayes-Harb's (2006) second experiment, involving the letter detection task, appeared to support the validity of vowel blindness. However, while the observed difference between the Arab and non-Arab participants' results is hard to explain,

we must remain cautious of these findings until sufficient evidence of the validity and reliability of this use of the instrument is available. Saigh and Schmitt's (2012) lexical decision task was the only study which included separate conditions based on the vowel length. Vowel length could prove to be an important factor affecting L1 Arab word recognition processes and should be explored in future studies. However, operationalising vowel length objectively may pose a number of challenges due to differences in how short and long vowels are encoded in English.

Alhazmi et al. (2019) and Alsadoon and Heift (2015) used eye-tracking to explore participants' online reading behaviours. Most notably, Alhazmi et al.'s vowel blindness study suggested that the major difference between the Arab and non-Arab participants was the overall speed of processing regardless of letter type. However, their participants were upper-intermediate and advanced level students, and we should be cautious in generalising these findings, particularly to lower-level learners, as cross-linguistic transfer is likely to diminish with increased L2 reading experience (Koda, 2005). Eye-tracking could also be a promising avenue of research for vowel and consonant processing, including its use to triangulate the outcomes of reading tasks. This technology is useful in L2 reading studies as it provides data on the unfolding process of reading, rather than its outcomes, without affecting the main characteristics of the task (Dussias, 2019).

Stein (2010) identified consonantal context as a promising line of research. However, several methodological factors in the study design prevented it from providing trustworthy findings. A more focused replication of the study with a smaller number of conditions could produce useful insights into written vowel processing.

Considering the limited and conflicting empirical evidence presented in this review, there is certainly scope for further research to explore Arab participants' processing of vowels and consonants and to test the validity of vowel blindness. Such research could comprise replications or original study designs. Replication studies serve an important role in confirming and disconfirming the results of previous studies and for directing further exploration of psychological processes (Brandt et al., 2014). However, replication studies that pertain to vowel blindness have so far been limited to word-naming tasks and these have failed to clearly replicate each other's findings.

Direct replications, those which avoid any intentional changes to the original study design (Marsden et al., 2018), could help to confirm whether the findings of the studies in this review are replicable and provide evidence of the studies' reliability and validity in exploring vowel and consonant processing in word recognition. However, they also run the risk of maintaining the methodological weaknesses of the original studies.

Partial replications are also required to explore whether observed findings could extend to different sub-populations of Arab learners. Of particular importance are partial replications with participants of lower- or higher-proficiencies. This could provide insight into the effect of L2 proficiency and reading experience on word recognition and the processing of vowels and consonants. Partial replication of Alhazmi et al.'s (2019) eye-tracking study with lower-level learners would be particularly welcome.

A common weakness across the included studies was the use of convenience sampling and general categories of miscellaneous non-Arab ESL learners. Logographic, alphabetic and syllabic writing systems differ in their encoding of vocalic information, and the performance of readers from each background could contrast in distinct ways with the performance of participants who are speakers of Arabic as an L1. Comparison groups that comprise readers of a single type of L1 writing systems could therefore facilitate more objective comparisons.

Future studies could also compare reading by L1 Arabic and Hebrew users as these represent the two major, consonantal languages with unvowelised standard written forms; if vowel blindness is genuine, we would expect it to be present in readers of other unvowelised consonantal languages. L1 readers of other languages that use Arabic script, such as readers of Urdu or Kurdish, could potentially help researchers corroborate whether observed effects derive from the script or from other causal factors.

Arabic and English writing systems differ in terms of the regularity, consistency and completeness of their written forms. This could directly contribute to differences in the accuracy and speed of vowel processing in L2 English. Future studies of vowel and consonant processing could also explore regularity and consistency effects and readers' preferred processing routes. These could include interventional studies that examine the effect of awareness-raising teaching activities that focus on vowel information and written word recognition, such as the effects of a programme of synthetic phonics. Affective and socio-cultural factors have also not been investigated or reported effectively; it is particularly concerning that no studies controlled for participants' L1 literacy experience given the increasing use of English-medium primary and secondary education. Mixed method studies could also focus on the impact of affective and socio-cultural factors in greater depth.

7 Conclusion and Recommendations

Before considering the implications of this systematic review, several limitations need to be acknowledged. First, the systematic review did not meet Macaro et al. (2018) stipulation that a systematic review have multiple reviewers. Multiple reviewers can help prevent human error or the assertion of individual assumptions that can bias a review. However, steps were taken to reduce any potential bias and ensure that the methodology and its reporting were transparent. This included the use of a research assistant in the screening of potential studies.

In addition, six dissertations could not be retrieved by the review's retrieval deadline for full screening. It is possible that these studies included methodologies and findings relevant to the processing of short and long vowels. Their omission should therefore be considered a weakness of this review. If this review is updated or replicated in the future, effort should be made to locate these studies. Finally, the methodological quality of the individual studies was typically moderate or low, according

to the scores obtained using the EPHPP instrument, and the lack of statistically comparable data limits the extent to which conclusions can be confidently drawn from the review.

When Ryan and Meara (1991) first described their theory of vowel blindness, they acknowledged that the phenomenon required empirical validation. This review suggests that, almost three decades later, the empirical evidence remains limited and the findings of published studies are conflicting. Methodological weaknesses in the published studies limit the trustworthiness of their findings and our ability to generalise them to the broader population of Arabic learners. The existing empirical research is therefore insufficient to either validate or refute the theory of vowel blindness.

Further research is required to explore the causal factors that contribute to systematic differences in the word recognition processes of L1 Arabic language-users, and both original research and replications will support this agenda. It is imperative that methodological rigour is improved to increase the trustworthiness of findings, particularly the refinement of comparison groups, preferably using objective sampling strategies appropriate to the aims of the studies. The inclusion of eye-tracking in original studies and replications would also help to increase objectivity, triangulate the findings of tasks and provide insight into online reading behaviours. Furthermore, the role of contextual and sociodemographic factors has not yet been explored. There is therefore scope for qualitative and mixed method studies to explore the role of such factors, including L1 reading experience and language of education.

A number of recommendations can be made for the teaching and learning of students of English who speak Arabic as an L1. Poor awareness of the phonology of English and its writing system could contribute to the observed weaknesses in word recognition and vowel processing. Reading curricula should provide systematic and incremental development of learners' phonolexical awareness, including knowledge of the phonemes of the target varieties of English and the range of spelling patterns that encode these sounds in regularly spelled words. For instance, L1 English reading synthetic phonics approaches could provide a starting point for the design and development of effective L2 English syllabi.

Learners also need practice to develop fast and efficient recognition of high-frequency, irregularly spelled words. Timed reading activities, particularly those that focus on faster reading of easier texts with known vocabulary, should be integrated into the curriculum to promote the development of reading speed. Likewise, extensive reading resources of an appropriate level can support improved reading speed (Grabe, 2009).

It is clear that many Arabic learners struggle with L2 English reading and written word recognition. Vowel blindness does not seem to adequately explain the difficulties faced by this population, and there remains a lack of empirical findings to explain what makes L2 reading so challenging for them. In the absence of this evidence, the above suggestions could help educators to support Arabic learners in improving the speed and accuracy of their written word recognition.

Acknowledgements The author is deeply indebted to Hamish Chalmers at the University of Oxford's Department of Education for his input and supervision during the systematic review that informed this chapter, and to Ailsa Campbell for her dedicated work as co-screener and research assistant during the project.

References

- Abu-Rabia, S. (1996). The role of vowels and context in the reading of highly skilled native Arabic speakers. *Journal of Psycholinguistic Research*, 25(6), 629–641.
- Abu-Rabia, S. (1997a). The need for cross-cultural considerations in reading theory: The effects of Arabic sentence context in skilled and poor readers. *Journal of Research in Reading*, 20(2), 137–147.
- Abu-Rabia, S. (1997b). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading & Writing: An Interdisciplinary Journal*, 16, 423–442.
- Abu-Rabia, S. (1998). Reading Arabic texts: Effects of text type, reader type and vowelization. *Reading and Writing: An Interdisciplinary Journal*, 10, 105–119.
- 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: An International Quarterly*, 16(1), 1–19. <https://doi.org/10.1080/0270271950160101>
- Abu-Rabia, S., & Taha, H. (2006). Reading in Arabic orthography: Characteristics, research findings, and assessment. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 321–338). Lawrence Erlbaum Associates.
- Al Juhani, E. A. (2015). *The sensitivity to vowels vs. consonants in English words: The case of Arabic ESL learners*. Retrieved from <https://prezi.com/tgjagr0yjbv/the-sensitivity-to-vowels-vs-consonants-in-english-words-t/>
- Alghamdi, H. (2015). Vowel blindness: A critical review. In *Proceedings of the 8th Saudi Students conference*. Imperial College London.
- Alhawary, M. T. (2011). *Modern standard Arabic grammar: A learner's guide*. Wiley-Blackwell.
- Alhazmi, K., Milton, J., & Johnston, S. (2019). Examining 'vowel blindness' among native Arabic speakers reading English words from the perspective of eye-tracking. *System*, 80, 235–245. <https://doi.org/10.1016/j.system.2018.12.005>
- Allmark, N. (2019). *A systematic scoping review of evidence pertaining to L2 word recognition among L1 readers of Arabic and its implications for the validity of vowel blindness*. Unpublished master's thesis, University of Oxford, Oxford, UK.
- Alsadoon, R., & Heift, T. (2015). Textual input enhancement for vowel blindness: A study with Arabic ESL learners. *The Modern Language Journal*, 99(1), 57–79.
- 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: Insights and perspectives* (pp. 31–54). Springer.
- Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grangef, J. A., Perugini, M., Spiess, J. R., & van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50(1), 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Brown, T. L., & Haynes, M. (2005). Individual differences in second language learning. The development of reading skills. *New Directions for Child Development*, 27, 19–34.
- Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2012). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd ed., pp. 93–122). Sage Publications.

- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. <https://doi.org/10.1037//0033-295X.108.1.204>
- Cook, V., & Bassetti, B. (2005). *Second language writing systems*. Multilingual Matters.
- Daniels, P. T. (2013). The Arabic writing system. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 412–432). Oxford University Press.
- Dussias, P. E. (2019). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149–166. <https://doi.org/10.1017/S026719051000005X>
- Effective Public Health Practice Project. (1998). *Quality assessment tool for quantitative studies*. Canadian Health Care.
- Eviatar, Z., & Ibrahim, R. (2014). Why is it hard to read Arabic? In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 77–96). Springer.
- Eviatar, Z., Ibrahim, R., & Ganayim, D. (2004). Orthography and the hemispheres: Visual and linguistic aspects of letter processing. *Neuropsychology*, *18*(1), 174–184. <https://doi.org/10.1037/0894-4105.18.1.174>
- Fender, M. (2008). Arabic literacy development and cross-linguistic effects in subsequent L2 literacy development. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development* (pp. 101–124). Routledge.
- Ferguson, C. A. (1959). Diglossia. *WORD*, *15*(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*(1), 71–99.
- Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). Blackwell Publishing.
- Gough, D., Oliver, S., & Thomas, J. (2012). Commonality and diversity in reviews. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd ed., pp. 35–65). Sage Publications.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Han, F. (2015). Word recognition research in foreign language reading: A systematic review. *University of Sydney Papers in TESOL*, *10*, 57–91.
- Hansen, G. F. (2010). Word recognition in Arabic as a foreign language. *Modern Language Journal*, *94*(4), 567–581. <https://doi.org/10.1111/j.1540-4781.2010.01094.x>
- Hansen, G. F. (2014). Word recognition in Arabic: Approaching a language-specific reading model. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 55–76). Springer.
- Hayes-Harb, R. (2006). Native speakers of Arabic and ESL texts: Evidence for the transfer of written word identification processes. *TESOL Quarterly*, *40*(2), 321–339.
- Ibrahim, R., & Eviatar, Z. (2012). The contribution of the two hemispheres to lexical decision in different languages. *Behavioral and Brain Functions*, *8*(3), 1–7. <https://doi.org/10.1186/1744-9081-8-3>
- Jordan, T. R., Almabruk, A. A. A., McGowan, V. A., & Paterson, K. B. (2011). Evaluating hemispheric divisions in processing fixated words: The evidence from Arabic. *Cortex*, *47*(8), 992–997. <https://doi.org/10.1016/j.cortex.2011.02.012>
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost, L. Katz, & M. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Elsevier Science & Technology.
- Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal*, *80*(iv), 450–460.

- Koda, K. (2005). *Insights into second language reading: A cross-linguistic perspective*. Cambridge University Press.
- Korat, O., Aram, D., Hassunha-Arafat, S., Iraki, H. H.-Y., & Saiegh-Haddad, E. (2014). Mother-child literacy activities and early literacy in the Israeli Arab family. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 323–347). Springer.
- Liabo, K., Gough, D., & Harden, A. (2017). Developing justifiable evidence claims. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (2nd ed., pp. 251–277). Sage Publications.
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher. *Language Teaching*, 51(1), 36–76. <https://doi.org/10.1017/S0261444817000350>
- Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, 68(2), 321–391. <https://doi.org/10.1111/lang.12286>
- Masrai, A., & Milton, J. (2018). The role of informal learning activities in improving L2 lexical access and acquisition in L1 Arabic speakers learning EFL. *The Language Learning Journal*, 46(5), 594–604. <https://doi.org/10.1080/09571736.2018.1520655>
- Milín, P., Smolka, E., & Feldman, L. B. (2018). Models of lexical access and morphological processing. In E. M. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 240–268). Wiley Blackwell. <https://doi.org/10.1002/9781118829516>
- Myhill, J. (2014). The effect of diglossia on literacy in Arabic and other languages. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 197–223). Springer.
- Nassaji, H. (2014). The role and importance of lower-level processes in second language Reading. *Language Teaching*, 47(1), 1–37. <https://doi.org/10.1017/S0261444813000396>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – A web and mobile app for systematic reviews. *Systematic Reviews*, 5(210). <https://doi.org/10.1186/s13643-016-0384-4>
- Perfetti, C., & Verhoeven, L. (2017). Epilogue: Universals and particulars in learning to read across seventeen orthographies. In L. Verhoeven & C. Perfetti (Eds.), *Learning to read across languages and writing systems* (pp. 2018–2028). Cambridge University Press.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.
- Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language*, 7(2), 531–540.
- Ryan, A., & Meara, P. M. (1996). *A diagnostic test for vowel blindness in Arabic speaking learners of English*. Lognostics. Retrieved from <http://www.lognostics.co.uk/vlibrary/arp96.pdf>
- Ryding, K. C. (2013). Second-language acquisition. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 392–411). Oxford University Press.
- Saiegh-Haddad, E. (2018). MAWRID: A model of Arabic word reading in development. *Journal of Learning Disabilities*, 51(5), 454–462. <https://doi.org/10.1177/0022219417720460>
- Saiegh-Haddad, E., & Joshi, R. M. (2014). Foreword. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. v–xiv). Springer.
- Saiegh-Haddad, E., & Spolsky, B. (2014). Acquiring literacy in a diglossic context: Problems and prospects. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 225–240). Springer.
- Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System*, 40(1), 24–36. <https://doi.org/10.1016/j.system.2012.01.005>
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70(1), 747–770. <https://doi.org/10.1146/annurev-psych-010418-102803>

- Simon, G., Bernard, C., Lalonde, R., & Rebaï, M. (2006). Orthographic transparency and grapheme-phoneme conversion: An ERP study in Arabic and French readers. *Brain Research, 1104*(1), 141–152. <https://doi.org/10.1016/j.brainres.2006.05.071>
- Stein, R. E. (2010). *Sensitivity to consonantal context in reading English vowels: The case of Arabic learners*. Unpublished doctoral dissertation, University of Memphis, Memphis, USA.
- Stevenson, M. (2015). Researching reading. In A. Paltridge & B. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 315–334). Bloomsbury.
- Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window and maze. *Journal of Psycholinguistic Research, 41*, 105–128. <https://doi.org/10.1007/s10936-011-9179-x>