USING DIACRITICS IN THE ARABIC SCRIPT OF MALAY TO SCAFFOLD ARAB POSTGRADUATE STUDENTS IN READING MALAY WORDS

1Khazriyati Salehuddin & 2Heather Winskel
1School of Language Studies & Linguistics
Universiti Kebangsaan Malaysia
2School of Health and Human Sciences
Coffs Harbour Campus, Southern Cross University, Australia

1Corresponding author: khazudin@ukm.edu.my

ABSTRACT

Purpose – This study aims to investigate the use of diacritics in the Arabic script of Malay to facilitate Arab postgraduate students of UKM to read the Malay words accurately. It is hypothesised that the Arabic script could facilitate the reading of Malay words among the Arab students because of their earlier exposure to the Arabic script in comparison to the Romanised script.

Method – Twelve Arabic first language speakers participated in a reading experiment that used DMDX, a Win 32-based display system for psychological experiments, to investigate whether or not Arabic vowel diacritics can facilitate Arabic first language speakers to read Malay words accurately. A total of 100 Malay bi-syllabic words were used as stimuli in three different forms: 1) Arabic script without diacritics; 2) Arabic script with diacritics; and 3) Romanised script. The participants’ responses and reaction times were recorded to analyse accuracy and speed.

Findings – Arabic first language speakers were more accurate when reading words in Arabic script of Malay with diacritics and when reading Romanised script than when reading Arabic words without diacritics. Arabic speakers read Malay words faster in Arabic script without diacritics and in Romanised scripts than when reading words in Arabic script with diacritics.
Significance – This study shows that the use of a more familiar script to a certain extent does facilitate language learners to produce the target language more accurately compared to using a less familiar script. Hence, educators should explore any possible means to scaffold learners in their learning process.

Keywords: Arabic script, cognitive complexity, DMDX, Malay, psycholinguistics, reading.

INTRODUCTION

Scaffolding is a concept that is mentioned every time references are made to Vygotsky’s ‘Zone of Proximal Development’ (ZPD) theory (Vygotsky, 1961, cited in Harris, 2008). Although this concept was never spelled out literally as “scaffolding” by Vygotsky, the notion of “scaffolding” is indeed the foundation of the theory. This is so due to the fact that ZPD refers to the imaginary zone between what learners can perform without help from others and what they cannot perform without help. In other words, it is “the region between what children already know and what they are capable of learning under guidance” (Schaffer, 2008, p. 125). For a learner to be able to move from what they cannot perform without help to what they can perform without help, a special platform is needed to temporarily support them so that they will be able to proceed to a more difficult task. This temporary platform will be removed when the learner is able to perform the once-a-difficult task without help. The act of providing the ‘temporary platform’ is what ‘scaffolding’ is (Moll, 2014).

Although ZPD was initially commonly applied in child development, the notion of ZPD has now been expanded to the adult learning process. Lebak and Tinsley (2010), for example, show that the ZPD and scaffolding concept (which they refer to as ‘Zone of Reflective Capacity’) work well with adults. With this in mind, we would like to apply the notion of scaffolding in the process of developing novice readers into more competent readers.

Reading is a highly complex process (Rayner, Pollatsek, Ashby, & Clifton, 2012) because it involves many different processes that are related to one another in many different ways. Because of this,
reading is one area that is constantly studied within the field of
cognitive psychology (Plaut, 1997) and psycholinguistics. Reading
is studied to investigate how texts are perceived, comprehended,
remembered and produced. In order to probe into what actually
happens in the cognition while reading is taking place, the
‘empirical experimentation’ approach is often adopted by cognitive
psychologists and psycholinguists to explain the processing
mechanisms involved in reading (Rayner et al., 2012).

Although reading experiments have been widely conducted in
various parts of the world, particularly those that investigate reading
texts in English, Greek, Roman, Hebrew, Japanese and Chinese,
such experiments are scarce when it comes to reading Arabic texts
in the Arabic script. Other than studies by Abu-Rabia (2002),
Abu-Rabia & Siegel (2003) and Taouk & Coltheart (2004), to the
knowledge of the authors, no other studies have been conducted to
investigate the cognitive processes that are involved in reading texts
in the Arabic script. Both Abu-Rabia and Siegel (2003) and Taouk
and Coltheart (2004), for example, studied the reading of Arabic
by Arab children. Abu-Rabia and Siegel’s (2003) study, which was
conducted in the form of written tests, showed that being multilingual
does not affect children’s reading development; whereas Taouk and
Coltheart (2004) found that children develop their reading of Arabic
(in Arabic script) in the same stages as they develop their reading
of English (in Romanised script). However, while these studies
investigated children’s reading accuracy, none of them examined
readers’ reaction time in their data analysis. In addition, neither of
them examined reading the Arabic script in languages other than the
Arabic language.

The Arabic morphological system is a “root-based” system (Abu-
Rabia, 2002, p. 300), formed via a combination of consonants that
make up the basic meaning of words (Abu-Rabia & Siegel, 2003).
The derived meanings are pronounced differently depending on
the different vowels that accompany the consonants. For example,
the combination of Arabic consonants /f/, /t/ and / ř / which
semantically refers to ‘open’, are read as /fataňa/ (‘he opened’),
and /fataňa/ (‘opening’) in different contexts. Readers will have
to decide beforehand how to read the words based on the context
of the occurrence of the words, making the Arabic script a ‘deep’
orthography. In order to solve this problem, the vowel diacritics (tashkeel) are introduced into the Arabic writing system, particularly for novice readers, to make the script a more transparent or ‘shallow’ orthography. The vowel diacritics are later gradually removed from the Arabic writing system (to become more opaque or ‘deep’) for more advanced readers. This can be considered as a form of “scaffolding” that is used in Arabic literacy.

The Arabic script is not only used in the Arabic language. Urdu, Persian, Sindhi and Malay are among the other languages in the world that use the Arabic script in the written form of their languages. However, in Malaysia, the Romanised script of Malay is used more widely than the Arabic script of Malay. The usage of the Arabic script for written Malay is somewhat limited to religious texts since the script is relatively more complex for the Malay language in comparison to the Romanised script (Salehuddin, 2012). In particular, reading the various forms of vowels that exist in the Malay sound system is a challenging task for the readers since what readers may see in the Arabic script of Malay may not map readily onto the Malay vowel phonemes.

To illustrate, while there are 10 vowel phonemes in the Malay phonological system, there are only three vowel letters in the Arabic script, namely ‘alif’ (‘ا’), ‘ya’ (‘ي’) and ‘wau’ (‘و’) (See Table 1 for the mapping of Arabic vowel letters onto the Malay sound system). From the perspective of Contrastive Analysis, this phenomenon is known as a ‘Divergent Phenomenon’ - a phenomenon that has the highest degree of difficulty for learners to learn or acquire a language (Gass & Selinker, 2008).

Table 1

How the Arabic Vowel Letters are Mapped onto the Malay Sound System (Adapted from Salehuddin, 2012, p. 1188)

<table>
<thead>
<tr>
<th>Arabic vowel letters</th>
<th>Malay vowel phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘ا’</td>
<td>/a/</td>
</tr>
<tr>
<td>‘ي’</td>
<td>/i/, /u/, /e/, /e/, /o/</td>
</tr>
<tr>
<td>‘و’</td>
<td>/u/, /o/, /a/</td>
</tr>
<tr>
<td>No symbol</td>
<td>/a/, /i/, /u/, /e/, /e/, /o/, /u/, /o/, /a/</td>
</tr>
</tbody>
</table>
With the aim of reviving the popularity of the Arabic script of written Malay, Salehuddin (2013) proposed several transformations in the Arabic script to facilitate the process of reading the Arabic script of Malay. One of the transformations proposed was the introduction of vowel diacritics into the Arabic script of written Malay – a feature that is not currently present in the script for the Malay language. A reading experiment on a group of Malay readers has shown that this transformation has benefitted the readers in terms of reading accuracy (Salehuddin & Winskel, 2014).

Following earlier studies on the Arabic script of written Malay, this paper presents findings of a study that was conducted to investigate if the introduction of vowel diacritics into the Arabic script of written Malay can act as a form of scaffolding to help novice readers of Malay (in this case Arabic first language speakers) to read Malay words more accurately. It was hypothesised that Arabic first language speakers would not face difficulty when reading Malay words in the Arabic script and that these speakers will read Malay words more accurately with faster reaction times when the words are presented in the Arabic script with diacritics than without diacritics. It was also hypothesised that the same group of readers will read Malay in the Arabic script more accurately with faster reaction times in comparison to reading Malay in the Romanised script and that preference in script types will influence the accuracy and the reaction times when reading Malay words in the Arabic script. Arabic first language speakers read the Arabic scripts earlier and more frequently than they read the Romanised script; hence, despite the divergence in the Malay phonology in comparison to Arabic, familiarity with the Arabic script will probably make it easier for Arabic first language speakers to read Malay in the Arabic script. In addition, the use of vowel diacritics in the Arabic script of Malay may help reduce the difficulty in deciding which vowels to use when reading the Malay words.

Being the national language of Malaysia, the Malay language is made a compulsory subject to be learned by non-Malaysians who wish to pursue their studies at Universiti Kebangsaan Malaysia (UKM). Because the Romanised script is the standard script for the Malay language, the Malay language is taught to all foreign students using the Romanised script.
To the first language speakers of Arabic, the Malay language is a foreign language; hence, learning to read in that language may be a difficult task. This is because from the Contrastive Analysis perspective, there are quite a number of differences between Arabic and Malay, among others, in their phonological, morphological and semantic systems. Where the learning of the Malay writing system is concerned, Arabic first language speakers have to rely on the writing system of their second language, i.e., the English language, in order to learn to read Malay. The question is, would it be easier for first language speakers of Arabic to read Malay in the Arabic script? Would it also be easier for the first language speakers of Arabic to read Malay in the Arabic script if vowel diacritics are introduced in the Arabic script of Malay?

**METHODOLOGY**

Twelve readers (nine male, three female) who are first language speakers of Arabic participated in this study. They are postgraduate students at UKM. All of them regard the Malay language as a foreign language and English as their second language. They only had very limited words in their Malay vocabulary list and they could only form simple Malay sentences. They were quite proficient in the English language as all of them had scored above 6.0 in their TOEFL examination; a pre-requisite to be a postgraduate student at the School of Language Studies and Linguistics (SoLLs), UKM. They all had been exposed to the Romanised script since they were in their early teens, i.e., since they learned English in schools in their homeland. All participants had normal or corrected-to-normal vision. They were between 26 and 39 years old and were recruited after an invitation was circulated to several postgraduate students at SoLLs. Prior to the session, each participant answered a questionnaire designed to gauge his or her fluency in reading Arabic, with and without diacritics. The participants’ speed when reading a sentence in Arabic with the vowel diacritics was between 2.4 seconds and 5.0 seconds whereas their speed when reading a sentence in Arabic without the vowel diacritics (similar length) was between 1.4 seconds and 5.8 seconds. Six of the participants preferred to read texts in the Arabic language without the vowel diacritics.
All participants were tested individually. Each of them sat approximately 50 cm in front of a *Multi-touch Full HD All-in-One* computer. Prior to the experimental session, each participant was briefed that Malay words written in Arabic scripts (printed in cursive) would be presented on the computer monitor and their task was to read the words aloud. Eighteen (18) bi-syllabic Malay words of various syllable structures (see Table 2 for the Malay bi-syllabic structures and frequency of occurrence) were presented to the participants as practice trails using the DMDX, a Win 32-based display system for psychological experiments that records reaction times to visual and auditory stimuli. Reaction times in reading experiments give information on the reading speed; hence, the speed recorded in the experiment is a good indication with regards to the degree of difficulty of particular reading materials. Eighteen words were given to the participants in the practice session to familiarise them with the procedure. After reading all the words in the practice trials, participants were asked to proceed to the experimental session by pressing the ‘space bar’ key. All instructions were given to the participants in English.

Table 2

*Malay Bi-syllabic Structures and their Frequency of Occurrence in the Experiment*

<table>
<thead>
<tr>
<th>Malay Bi-Syllabic Structures</th>
<th>0 VL</th>
<th>1 VL (1)</th>
<th>1 VL (2)</th>
<th>2 VL (1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CV.CV (e.g., bumi)</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>2. CV.CVC (e.g., sakit)</td>
<td>8</td>
<td>11</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>3. CV.CVV (e.g., ramai)</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. CV.CV (e.g., rimba)</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CVC.CVC (e.g., timbun)</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6. CVC.CVV (e.g., bantai)</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>40</td>
<td>30</td>
<td>22</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note:* “C” refers to consonant sound and “V” refers to vowel sound. Hence, “CV” refers to a syllable with one consonant sound and one vowel sound. “VL” refers to Vowel letters; hence, “0 VL” means no vowel letter in the word whereas “1 VL (1)” means 1 vowel letter in the first syllable of the word.
The first part of the experiment involved reading aloud 100 Malay bi-syllabic words in two different forms. The first form consisted of Malay bi-syllabic words written without vowel diacritics, whereas the second form consisted of the same Malay bi-syllabic words but written with vowel diacritics. The words were taken from the same list of words that was used in a reading experiment that compared the difference in reading Malay words in the Arabic script with or without diacritics by Malay participants (Salehuddin & Winskel, 2014). Since there are Malay words that contain vowels that are non-existent in Arabic, careful selection was made to ensure that all the Malay words have only the three vowel sounds that exist in Arabic, namely, /a/, /i/ and /u/. The words include “silau” (سلاى), “rambai” (رامبى), and “santan” (سانطن). The presentation of Malay words in the Arabic script of Malay with diacritics was based on the proposition put forward by Salehuddin (2013), where vowel diacritics were added to all syllables in the bi-syllabic words, regardless of whether the respective syllables have a vowel letter or otherwise; for example, “silau” (سلاى), “rambai” (رامبى), “tukar” (تىكار) and “santan” (سانطن). The presentation of stimuli was randomised and mixed for both with and without diacritics, with different order for each participant. After the 100th stimulus, a page with the instruction “Press the space bar key to continue” written on it appeared on the monitor. The experimental session resumed once the participants pressed the ‘space bar’ key. The entire experimental session lasted for about 20 minutes for each participant.

The second part of the experiment involved the same 100 Malay bi-syllabic words, but now, written in the Romanised script. There was no break in between the first and second parts of the experiment. This part lasted for about 10 minutes for each participant.

RESULTS

Accuracy of reading was measured through the participants’ mean proportion of correct responses; whereas the speed of reading was measured through the participants’ reaction times. The mean proportion of correct responses and reaction times are presented in Table 3.
Correct Responses (Mean Proportion) and Reaction Times (in Milliseconds) for Reading Malay in Arabic (With and Without Diacritics) and Romanised Scripts

<table>
<thead>
<tr>
<th>Preference</th>
<th>Arabic Script (Without Diacritics)</th>
<th>Arabic Script (With Diacritics)</th>
<th>Romanised Script</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct Response</td>
<td>Reaction Times</td>
<td>Correct Response</td>
</tr>
<tr>
<td>Without</td>
<td>.70 (.05)</td>
<td>792 (72)</td>
<td>.84 (.08)</td>
</tr>
<tr>
<td>With</td>
<td>.62 (.08)</td>
<td>672 (35)</td>
<td>.77 (.06)</td>
</tr>
</tbody>
</table>

Note: (a). 0.70 means that for every stimuli that appeared without diacritics, those who prefer Arabic script without diacritics than with diacritics gave .70 correct responses.
Note: (b). Standard Deviations are in Parentheses.

Two repeated analyses of variance (ANOVA) were conducted for proportion of correct responses and reaction times. Script Type (without diacritics, with diacritics, Romanised) was a within subjects factor and Preferred Script Type (with diacritics, without diacritics) was a between subjects factor.

The ANOVA on the Correct Response data showed a main effect of Script Type \((F(2, 20) = 14.94, p < .001, \eta^2_p = 0.599)\). There was no significant effect of Script Preference \((p > .8)\). This means that those who prefer to read Malay in the Arabic script did not perform differently. However, there was a significant interaction between Script Preference and Script Type \((F(2, 20) = 7.01, p < .01, \eta^2_p = 0.443)\). Paired sampled t-tests were conducted. As can be observed in Figure 1, Arabic speakers were more accurate in reading words in the Arabic script of Malay with diacritics than without diacritics \(t(11) = 7.73, p < .001\). They were also more accurate when reading words in Romanised script than in the Arabic script of Malay without diacritics \(t(11) = 2.70, p < .05\). However, the difference between reading the Arabic script of Malay with diacritics and the Romanised script was not significant \(p > .6\).
The ANOVA on the Reaction Times data showed a main effect of Script Type ($F(2, 20) = 10.97, p = .001, \eta^2_p =.523$). There was a significant effect of Script Preference ($F(1, 10) = 7.96, p < .05, \eta^2_p =.443$); however, there was no significant interaction between Script Preference and Script Type ($p > .7$). Paired sampled $t$-tests were conducted. As can be seen in Figure 2, Arabic speakers were faster in reading words in the Arabic script of Malay without diacritics ($t(11) = 7.74, p < .001$) and in Romanised scripts ($t(11) = 3.65, p < .01$) than reading words in the Arabic script of Malay with diacritics. However, the difference in the reaction times when reading the Arabic script of Malay without diacritics and Romanised script was not significant ($p > .9$).

**DISCUSSION AND CONCLUSION**

The similarities in the form, orientation and presentation of the Malay words in the experiment with the Arabic words were noticeable. As hypothesised, Arabic first language speakers are not expected to face difficulty when reading Malay words in the Arabic script of Malay. However, not all results are as predicted.
First language speakers of Arabic read Malay words more accurately when reading Malay words in the Arabic script of Malay with diacritics than when reading without diacritics. This supports the hypothesis that diacritics facilitate reading of Malay words in the Arabic script of Malay. However, the hypothesis that Arabic first language speakers will read Malay in the Arabic script more accurately in comparison to the Romanised script is not supported. Instead, the Arabic first language speakers performed significantly better when reading Romanised script than when reading the Arabic script of Malay without diacritics, despite the Romanised script being a secondary form of writing (in comparison to the Arabic script) which was learned through the English language, their second language. The Arabic first language speakers’ accuracy in reading Malay in the Romanised script and in reading Malay in the Arabic script with diacritics was not significantly different. Such findings may suggest that reading Malay words in the Arabic script without diacritics, which is the current practice of reading the Arabic script of Malay in Malaysia, is a difficult process, even for Arabic first language speakers, who read the Arabic script much earlier than they read the Romanised script. This also supports the notion that reading without diacritics is cognitively more complex (Salehuddin, 2012) even for first language speakers of Arabic because of its deep orthographic nature.
Nevertheless, the hypothesis that Arabic first language speakers will read Malay in the Arabic script more accurately with faster reaction times in comparison to reading Malay in the Romanised script is not supported. This is because there is a one-to-one mapping between the Malay words they read and the sound they produce in both Romanised and Arabic scripts with diacritics. Arabic first language speakers make more mistakes in reading words without diacritics (e.g., “mata” (مَتا) was read as “mat”; “nazir” (نَازِر) was read as “nazar”) partially because they do not have much access to the Malay lexicon, like they have to the Arabic lexicon.

With regards to the hypothesis that the first language speakers of Arabic will read Malay words with faster reaction times when the words are presented with diacritics than without diacritics, results reveal that Arabic speakers read Malay words in the Arabic script with diacritics significantly more slowly than they read Malay words in the Arabic script without diacritics and in the Romanised script. This suggests that while vowel diacritics facilitate Arabic first language speakers in reading Malay words in the Arabic script more accurately, the presence of the diacritics slows their reading. These contradicting findings from what was hypothesised earlier can be explained.

Diacritics in Arabic texts are often removed in texts meant for advanced readers. However, the diacritics are maintained in the Qur’an, the Holy Scripture for the Muslims, to ensure accuracy in reading it. Hence, adult Arabic speakers would typically encounter Arabic scripts with diacritics only when reading the Qur’an. Being the Holy Scripture, the act of reading the Qur’an has to be done accurately. Hence, regardless of what the first language of the reciters is, those who read the Qur’an will be more likely to recite it slowly. Since the first language speakers of Arabic in this study are all Muslims, they are more likely to have been conditioned to read the Qur’an carefully. Since the presence of the vowel diacritics in the Malay words in the Arabic script may make the Malay words look more like Qur’anic words rather than Malay words, the first language speakers of Arabic tended to read the Malay words with diacritics more carefully, and hence, more slowly.
The Arabic script for the Malay language is considered to be cognitively complex because of its inconsistency in terms of whether it is a ‘shallow’ orthography or a ‘deep’ orthography within the same reading material. To illustrate, vowel /a/ in the Malay word ‘mama’ is manifested in the letter ‘alif’ (i.e., ‘ا’), in both syllables (i.e., ‘ماما’); yet, the letter ‘alif’ is not manifested in the first syllable of ‘katil’ (i.e., ‘كِتِل’), in the second syllable of ‘jasa’ (i.e., ‘جَاس’), and in neither syllables of ‘tamak’ (i.e., ‘طَمَع’).

The Arabic first language speakers also show that they were able to read Malay words in the Romanised script better than they were able to in the Arabic script without diacritics. This is because as mentioned earlier, the Romanised script of Malay is more transparent than the Arabic script of Malay (Salehuddin, 2012). The Arabic first language speakers also read Malay in the Romanised script as accurately as they read Malay in the Arabic script with diacritics. This is because when vowels are manifested in the form of diacritics in the Arabic script of Malay, reading the Malay words appears to be a more ‘scaffolded’ reading task for the Arabic first language speakers. The presence of vowel diacritics helped them to decide which vowel phoneme should be used to make up a particular syllable, particularly when the words were not familiar to them. The presence of vowel diacritics in this sense may function as a ‘scaffolding’ instrument which helped the Arabic first language speakers to read the Malay words more accurately, particularly when reading words they had never heard of.

The small number of subjects and words in the current study may not justify the claims made in this manuscript. The same goes with the small number of words used in this study. Nevertheless, this study may be a good indicator to educators who teach Malay as a foreign language, particularly when their learners are adult learners. Being mature learners does not mean they need less assistance in becoming more proficient speakers of the target language. Scaffolding, although more likely to be discussed in research on child development, is still needed by mature learners. Educators can scaffold these learners by making use of items that the learners are more familiar with, especially at the early stages of learning.

With regards to the current study, educators teaching Malay to first language speakers of Arabic may introduce new Malay words not
only in the Romanised script, but also in the Arabic script – a script that the learners are more familiar with, to scaffold the learners. After writing the Malay words in the Romanised script, educators could scaffold the learners by writing the same Malay words in the Arabic script; first, with diacritics. These Malay words in the Arabic script with diacritics will act as a special platform that can temporarily support the learners in their initial stage of reading Malay. Later, when the learners are familiar with how the words should be read, the diacritics (the temporary platform) can be removed, so that the Malay words in the Arabic script are now written without diacritics. This more familiar script can be used in lessons until the Malay words are firmly established in the learners’ mental lexicon.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Education Malaysia, for funding the ERGS/1/2011/SSI/UKM/02/5 research entitled Exploring the Cognitive and Perceptual Processes in Reading. Special gratitude also goes to all 12 postgraduate Arabic first language speakers of Universiti Kebangsaan Malaysia who participated in the study.

REFERENCES


