The Development of a Measure of Orthographic Knowledge in the Arabic Language:

A Psychometric Evaluation

Sana Tibi, Lisa Fitton, & Autumn McIlraith

Abstract

Although Arabic is an official language in 27 countries, standardized measures to assess Arabic literacy are scarce. The purpose of this research was to examine the item functioning of an assessment of Arabic orthographic knowledge. Sixty novel items were piloted with 201 third grade Arabic-speaking students. Participants were asked to identify the correctly spelled word from a pair of two words. Although the assessment was designed to be unidimensional, competing models were tested to determine whether item performance was attributable to multidimensionality. No multidimensional structure fit the data significantly better than the unidimensional model. The 60 original items were evaluated through item fit statistics and comparing performance against theoretical expectations. Twenty-eight items were identified as functioning poorly and were iteratively removed from the scale, resulting in a 32-item set. A value of 0.987 was obtained for McDonald's coefficient omega from this final scale. Participants' scores on the measure correlated with an external word reading accuracy measure at 0.79 ($p < .001$), suggesting that the tool may indeed measure skills important to word reading in Arabic. The task is simple to score and can discriminate among children with below-average orthographic knowledge. This work provides a foundation to develop Arabic literacy assessments.

Keywords: Arabic, Literacy, Orthographic Knowledge, Assessment, Item Response Theory

Orthographic processing has been defined as the ability to acquire, store and use orthographic representations (Stanovich & West, 1986; Apel, 2010, 2011). Theories of reading suggest that this ability to develop and use acquired information about orthography (i.e., any system of symbols or marks that represent spoken language in print; Chomsky, 1970) may be essential to efficient, fluent reading development (Ehri, 1995, 2005; Ehri & McCormick, 1998; Share, 1995; Venezky, 1999). Awareness of orthographic regularities is believed to help students learn the correct spelling of morphologically complex words that vary in phonological form (e.g., "muscle" and "muscular") in a way that enables effective decoding (Apel, 2011; Chomsky, 1970; Graham et al., 2002). Indeed, children who experience reading difficulty tend to exhibit deficits on assessments of orthographic processing (Bergmann & Wimmer, 2008; Rothe et al., 2015).

Much of the current understanding of orthographic processing has been based on results from studies conducted in English or other alphabetic European orthographies (Share, 2008), though there has been a substantial increase in studies examining non-alphabetic orthographies (e.g., radical awareness as an important orthographic processing unit in Chinese reading development; Ho, Ng, & Ng, 2003; Kuo et al., 2015). Findings from one language may not generalize to languages with different orthographic systems, making it essential therefore to expand research across all orthographic systems (Daniels & Share, 2018; Share & Daniels, 2015).

Arabic, a language whose orthography differs substantially from European orthographies, is spoken by 420 million people worldwide (Gordon, 2005), while Arabic orthography is used by millions of non-Arabic speakers (e.g., Pashto, Persian, & Urdu; Mirdehghan, 2010). Furthermore, Arabic is one of the fastest-growing language in the North America, reflecting increasing trends of linguistic diversity across the globe (Brown, 2016). The number of individuals who speak Arabic in the U.S. was estimated to be approximately one million in 2013 and continues to increase (U.S. Census Bureau, 2015). However, there are currently few examples of psychometrically-sound, standardized measures for assessing Arabic literacy development (Authors). Given the need for validated measures to conduct research to examine connections across languages and orthographies, the development and psychometric evaluation of such measures is essential.

The purpose of the present paper is to begin addressing this gap in the literature through the psychometric evaluation of a tool designed to assess one aspect of orthographic processing in Arabic: orthographic knowledge at the lexical level. The results of this work will contribute to ongoing investigation of orthographic knowledge and, more broadly, reading skill in Arabic. Findings may also add to the larger understanding of cross-linguistic differences in orthographic knowledge and provide a foundation for ongoing evaluation of the role of orthographic processing in literacy development. In the following literature review, we define orthographic knowledge in relation to orthographic processing and reading development, discuss measurement of orthographic knowledge at the sublexical and lexical levels, and describe key features of Arabic orthography essential to consider in assessment development.

Defining Orthographic Knowledge

Although orthographic processing is considered to play an integral role in reading and spelling (e.g., Apel, Wolter, & Masterson, 2006), the precise directionality of this relation remains in question (Conrad & Deacon, 2016; Deacon et al., 2019; Querido et al., 2020; Roman, Kirby, Parilla, Wade-Woolley, & Deacon, 2009). Some studies have suggested that orthographic processing uniquely contributes to word reading (Cunningham et al., 2001; Pasquarella et al., 2014; Ricketts, Bishop, Pimperton, & Nation, 2011). Other studies have found no contribution of orthographic processing to gains in word reading (Georgiou, Parrila, and Papadopoulos, 2008), a pattern where word reading gains contribute longitudinally to lexical and sublexical orthographic knowledge (Conrad & Deacon, 2016; Deacon et al., 2012), or even a bidirectional relation where both skills contribute to each other over time (Querido et al., 2020). These discrepancies may stem from the possibility that the construct of orthographic processing is multidimensional, comprised of two distinct underlying aspects: orthographic knowledge and orthographic learning (Deacon et al., 2019). Orthographic learning refers to the dynamic ability to develop orthographic representations (Deacon et al., 2012), whereas orthographic knowledge has been defined as the crystalized store of these recognized representations or patterns (Conrad & Deacon, 2016; Deacon et al., 2019).

Crystalized orthographic knowledge centers on individuals' established awareness of letter patterns and word-specific orthographic representations stored in memory (Apel, 2011; Ehri, 2005, 2014, 2017; Conrad, 2008; Deacon et al., 2019). Chambre, Ehri, and Ness noted that orthographic knowledge involves understanding of phoneme-grapheme correspondence, awareness of morpho-graphemic patterns, and familiarity with general conventional spelling rules (2017). This knowledge can exist either at an implicit level or an explicit level (Ouellette & Sénéchal, 2008a, 2008b; Pacton, Perruchet, Fayol, & Cleeremans, 2001; Wolter & Apel, 2010). Sometimes children learn the orthographic regularities in their lexicon implicitly through exposure to orthographic patterns that recur in different words (Cassar & Treiman, 1997; Ehri, 2017; Treiman & Kessler, 2006; Tunmer & Nicholson, 2011). In contrast, these regularities can also be directly taught and/or recognized through metacognition, such that individuals make conscious decisions about the acceptability of the spelling or letter sequences in one's language (Apel, 2011; Kaefer, 2016).

Orthographic knowledge is integral to many major models of written language acquisition, including the later stages of Ehri's five phase model of reading acquisition (Ehri, 1995, 2005; Ehri & McCormick, 1998), Share's influential self-teaching hypothesis (Shahar-Yames, 2008; Share, 1995; Share, 1999; Tucker, Castles, Laroche, & Deacon, 2016), the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001, 2002), and the connectionist model (Seidenberg, 2005; Seidenberg & McClelland, 1989). More recently, Ehri emphasized that it is through orthographic mapping that connections between the spellings, pronunciations, and meanings of specific words are formed and stored in memory (2005, 2014, & 2017). These connections improve qualitatively as children develop and learn to read and spell.

Measurement of Orthographic Knowledge

In broad terms, orthographic knowledge can be measured at the lexical or sublexical level(s). At the lexical level, the pseudohomophone orthographic choice task put forth by Olson, Forsberg, Wise, and Rack (1994) is a classic task used to measure orthographic knowledge (e.g., Deacon et al., 2013). This task requires participants to choose the correct spelling of a word from a phonologically-matched pair of orthographic representations, with one being correctly spelled and the other a pseudohomophone (e.g., hurt vs. hert). Conversely, tasks designed to measure orthographic knowledge at the sublexical level are based on pseudowords. An example of a typical sublexical assessment of orthographic knowledge in English is the letter sequence judgement task. The task requires participants to indicate whether a given letter sequence would be acceptable (e.g., baff versus bbaf; Cassar & Treiman, 1997; Treiman, 2017).

An advantage of measuring orthographic knowledge at the sublexical level is that sublexical assessments focus on general orthographic knowledge about letter sequences and rules in reading and spelling, rather than orthographic knowledge of specific words (Burt, 2006;

Castles & Nation, 2006; Nation, Angells, & Castles, 2007). Any task based on specific words is susceptible to bias based on individuals' familiarity with the included words. Several studies have even shown that individuals' performance on sublexical orthographic knowledge tasks is distinct but related to performance on lexical tasks (Deacon et al., 2012; Pasquerella et al., 2014). It is likely that lexical and sublexical tasks for assessing orthographic knowledge measure differing underlying abilities.

In the context of Arabic language, the relation between orthographic knowledge at the lexical and sublexical levels is relatively unclear. As discussed in the next section, the rules governing Arabic orthography are intertwined with morphology and semantics. Consequently, evaluation of orthographic knowledge at the sublexical level in Arabic would require the specification of the morphological, semantic, and orthotactic context for any given item. Given the preliminary nature of this work and the meta-cognitive demands of such a task, our focus in the present paper is on examining orthographic knowledge at the lexical level in Arabic.

Arabic Orthography

Arabic is a Semitic language with a primarily consonantal orthography (Daniels, 1992). It has 28 letters, two of which can function as either long vowels or consonants. The three Arabic long vowels /a: /, /u: /, and /i: / also have three corresponding short vowels. Several of the key features of the Arabic writing system are described briefly in the following paragraphs. These include the use of fully vowelized versus partly vowelized writing, homography, ligaturing, diglossia, dense morphology, occasions of phoneme-grapheme mismatch, and context-dependent orthotactic rules (Abu-Rabia, 1996, 1999). Each of these characteristics were directly considered in the development of the current task.

The Arabic long vowels are always explicitly represented in the writing system, whereas the three short vowels (/a/, /u/, /i/) may or may not be represented by diacritics placed on top or underneath the consonants (e.g., ـَ, ـِ, ـُ). Depending on the medium, text may be fully or partly vowelized. Generally, texts written for beginning readers or religious contexts are fully vowelized and include orthographic representations of short vowels using diacritics (Abu-Rabia, 1996, 1999), but texts for adults generally do not include the short vowels, hence partly vowelized. The full vowelization of Arabic script allows for a higher degree of grapheme-phoneme correspondence. Despite the high grapheme-phoneme correspondence possible in fully vowelized texts, Arabic orthography has other features that make it more complex than alphabetic orthographies that are transparent (Share & Daniels, 2015).

Because words can be written without short vowels, Arabic orthography is also characterized by homography. This refers to words that are spelled in the same way, but have different pronunciations. For example, when the three consonant-word /ktb/ appears without short vowels, there are multiple possibilities for reading it: it can be read as /kataba/ “to write”, /kutub/ “books”, or /kutiba/ “was written”. In a partly vowelized text, the reader must infer the correct meaning and pronunciation from the semantic context and the regularity in the morphological information provided by the root and the word pattern (Boudelaa & Marslen-Wilson, 2001; Authors). This ambiguity can be eliminated by the representation of short vowels in the text, because the short vowels explicitly represent additional phonological information, hence restricting the possible pronunciations of the homographic words.

Arabic letters can generally be divided into connecting (i.e., ligatured) letters versus non-connecting letters. Ligaturing between some consecutive letters is an obligatory feature. Therefore, a word containing such letters must be written in a connected “cursive” way; without this ligaturing, its spelling is considered incorrect. In addition, when ligaturing takes place, the connected graphemes change their shape depending on their word position (initial, medial, or final). For example, the letter "h" in Arabic is written in four different ways depending on its position in the word and whether it is connected to previous letter or not. The connectivity (or allography) between letters, in addition to the optional diacritics (short vowels), results in multiple letter sequences being permissible at the sublexical level in Arabic.

Diglossia is another linguistic phenomenon noted in the Arabic writing system. Diglossia is defined as a difference between the standard literary form and the spoken form of a language (Ferguson, 1959). For example, the literary form is used in writing and some formal speech situations, whereas the spoken form is used in daily conversations. Diglossia in Arabic has been observed to impact literacy acquisition (Abu-Rabia, 2000; authors). Diglossia and its effect on reading and writing has been investigated by Saiegh-Haddad (2003, 2004) who reported that the linguistic distance between the standard form and the spoken Arabic exists in all aspects of the language, including phonology, morphology, and semantics. For example, the phoneme /q/ in standard Arabic can be produced as /ʔ/, or /g/, or /k/ in spoken Arabic depending on the spoken dialect. Eviatar and Ibrahim (2000) argued that Palestinian kindergarten and grade one children living in Israel function like bilinguals because the standard literary form is different enough from their spoken form. For a detailed account of diglossia in Arabic, see Maamouri (2009). For a counter argument on diglossia, see Boudelaa & Marslen-Wilson (2013) who presented experimental evidence showing that the standard and the spoken varieties of Arabic are cognitively processed in similar ways despite their structural, functional, and distributional differences.

A description of Arabic orthography would be incomplete without a mention of morphology. Arabic is characterized by its dense morphology, which involves both linear and nonlinear processes. Unlike Indo-European languages, Arabic words are derived by means of combining two abstract units (consonantal root and the mostly vocalic word pattern) in a nonlinear way. This mechanism allows for the derivation of words from the same consonantal root, hence allowing for orthographic transparency, shared semantics, and morphological complexity. The shared orthographic and semantic qualities between the root and its derived words contribute to reading accuracy (Taha & Saiegh-Haddad, 2017; authors), fluency, and comprehension (Layes, Lalonde, & Rebaï, 2017; Authors). We know from research in English that redundancy of letters and letter patterns allow for well-defined orthographic representations which in turn have been shown to be crucial for reading proficiency (Chomsky, 1970; Ehri, 1992; Perfetti, 1992; Rahbari & Sénéchal, 2010; Share, 1999). Arabic also allows for linear morphology when prefixes and suffixes are added linearly to words. The prefixes and suffixes provide information about the grammatical functions of a word such as gender, number (singular, dual, and plural), person, and time. For example, the inflected word /mudarrisata:n/ means "two female teachers". In this word, the "t" is a morpheme denoting the female gender, and the /a:n/ morpheme represents the "dual" marker; both suffixes are added to the noun /mudarris/ "teacher". It should be noted here that in the case of the masculine plural suffix /u:/, it is represented orthographically with two letters "وا", yet it is pronounced only with one sound /u:/.

There are other situations in which orthographic representations do not match their corresponding pronunciations in Arabic. For example, diacritics are used instead of a letter to represent a phoneme in certain grammatical contexts. Specifically, for an indefinite noun in Arabic, double diacritics are added to the vowelized word to represent the /n/ phoneme although the letter "n" is not written (e.g., "an apple" is written as "تفاحةً"). Mismatches between the orthography and phonology occur in several other contexts in Arabic. Consequently, individuals developing literacy in Arabic must know the rules corresponding to each orthographic context.

Additional examples of phoneme-grapheme mismatch include the use of the definite article /ʔal/ before a group of letters known as "sun letters" (see Saiegh-Haddad & Roitfarb, 2014). Sun letters dictate that the preceding /l/ phoneme is not pronounced, despite it being represented orthographically. Saady, Ibrahim, and Eviatar (2015) reported that the rate of missing a target letter was higher in the sun letters context among skilled readers of Arabic. There are also orthotactic rules that place constraints on the sequencing of letters in certain contexts (e.g., the closed /t/ cannot occur in the initial position nor be connected to a letter after it). Moreover, the feature of allography, which was described previously, also influences orthotactic rules in Arabic. This is because letters change their shapes depending on the position of occurrence in a word. Using a letter shape in the wrong position (e.g., medial position instead of final) is considered incorrect spelling (e.g., the letter "s" in the final position is written as "س" as in the word "teacher" (مدرس), but will be deemed inaccurate if written in the middle position allograph "ســ".

The orthographic features of Arabic discussed thus far are components of a complex and multifaceted writing system, which is orthographically deep despite its generally transparent phoneme-grapheme correspondence (Share & Daniels, 2015; Authors). To decode and spell, speakers of Arabic must draw on their knowledge of the conventional rules present in the written language, phoneme-grapheme correspondence and the morphological information inherent in its lexical structure. To develop an assessment of children's orthographic knowledge in Arabic, each of these aspects of the system are important to consider. Further, the underlying dimensionality (or potential divisions of underlying knowledge) of children's orthographic knowledge in Arabic is relatively unknown. Consequently, evaluation of a wide array of orthographic rules pertaining to morphology, diglossia, ligaturing, and orthotactic rules is essential to the development of a valid assessment tool.

The Current Study

The measure developed and evaluated for the current study was an orthographic choice task (Author). Each item consisted of a pair of words: one correctly-spelled target and one incorrectly-spelled foil. Foils were in some cases pseudohomophones, and in other cases were pronounced differently from the target. This task is similar to previously-established tasks intended to assess orthographic knowledge at the lexical level in English and French (Deacon et al., 2013).

The purpose of the current study was to develop a measure consisting of a set of items with strong relations to the latent underlying construct of lexical orthographic knowledge. Correlations between each item and the total test score were examined, followed by confirmatory factor analysis to determine the dimensionality of the measure. Given the nature of the task, and the variations in types of foils used, it was important to test whether there were any sets of items that clustered together. Such clustering would be indicative of additional dimensions beyond the main dimension of orthographic knowledge. After assessing dimensionality, the measure was further refined by removing poorly-fitting items, resulting in a set of well-functioning items which were retained in the final version of the measure. Finally, children's total scores on the refined version of the measure were correlated with an external measure of word reading accuracy in Arabic to provide preliminary information about the tool's construct and potential discriminant validity.

Method

Measure Construction

Orthographic choice task. The primary measure of interest within the present paper was an Arabic Orthographic Choice task (based on Deacon et al., 2013) developed by the first author, and included of set of 60 novel items. The items were constructed to evaluate children's knowledge of the conventional rules of Arabic spelling by targeting the specific characteristics of Arabic orthography noted previously (Authors). For each item, students were presented with two words written in Arabic: a real word and an incorrectly-spelled foil. Children were instructed to choose the word that was spelled correctly.

Of the 60 items, 45 included target-foil pairs that were phonological matches, or homophones. In these cases, the foil can be referred to as a *pseudohomophone*. The remaining 15 items included target-foil pairs that were not phonological matches when pronounced aloud. The target words were selected carefully to be representative of multiple characteristics of Arabic orthography and appropriate for the age group of interest. The item set was constructed by considering the influences of vowelization, instances of homography, ligaturing, diglossia, the connection between morphology, semantics, and orthography, occasions of phoneme-grapheme mismatch, and context-dependent orthotactic rules.

Because the task was intended for use with children at the elementary school age, all target words were fully vowelized. As discussed previously, Arabic can be written with or without the short vowels (Abu-Rabia & Siegel, 2002). Children typically begin learning to read in Arabic using the transparent orthography, but are expected to develop sufficient skills to read using the more opaque orthography as they get older (Abu-Rabia, 2001; Taha, 2016). Only fully-vowelized words were represented in the items to allow for specific examination of children's ability to decode fully vowelized Arabic writing. Full vowelization also eliminated the potential for homography among correctly-spelled items.

To account for the presence of ligaturing and diglossia in Arabic writing, several items were constructed to assess children's knowledge of these features. To evaluate children's awareness of ligaturing, foils were constructed with the incorrect letter form based on ligaturing and the letter's position within the word (e.g., "earthquake" /zil-za:l/ "زل زال " instead of /zilzal/ "زلزال"). To evaluate their awareness of diglossia, foils were constructed to represent a dialectal difference (e.g., "student" /tilmi:z/ "تلميز " instead of /tilmi:ð/ "تلميذ"). It should be noted that diglossia- and coarticulation- in some instances provide opportunities for phonemes to be replaced with phonological neighbors (Asaad & Eviatar, 2013) as in the phonemes /ð/ and /z/, or /d/ and /dˤ/.

Next, items were designed to probe children's understanding of the interconnectedness between morphology, semantics, and orthography in Arabic (Authors). Different parts-of-speech such as adjectives, nouns, pronouns, and verbs were all included as target words. These four commonly-occurring word classes require different vowel patterns. For a child to identify the correct spellings of words from each word class, a thorough familiarity with the orthographic patterns is critical.

The targets also included words that were relatively simple morphologically and those that were more complex. As noted previously, most Arabic content words are comprised of at least two bound morphemes. These include the root and the word pattern. The root provides the general semantic field of the word and the word pattern indicates the grammatical category of the word. Both elements are required for the reader to derive a meaningful word, and distinguish the meaning and function of the word. Moreover, when a prefix and/or suffix is added to these elements forming a simple or complex word, the word is considered morphologically complex (for review on morphological complexity, see Authors). Words without prefixes, suffixes, and infixes are considered morphologically simple (Boudelaa & Marslen-Wilson, 2005; Authors). Both morphologically simple and complex words were included in the target word set.

Items were also constructed with consideration of the situations in which graphemes do not map precisely onto specific phonemes in Arabic. In some Arabic words, graphemes represent multiple phonemes or phonemes influenced by coarticulation. To design items to gauge children's knowledge of these orthographic transparency exceptions, some words with imperfect phoneme-grapheme correspondence (i.e., opaque orthography) were included as targets. The foils were written as would be expected based only on the pronunciation of the target word (e.g., "a Dirham" /dɚhamɛn/ was spelled with the "n" letter at the end instead of double diacritic strokes "tanween"). For a child to distinguish correctly between these target-foil pairs, he or she would need to be aware of the specific contexts in which phonemes and graphemes do not map onto each other perfectly.

Finally, several target-foil pairs were created to evaluate participants' knowledge of orthotactic rules, which can change based on the position of letters within words (e.g., closed /ta/ in the initial position). Similar to the previous example, words whose orthography illustrate these context-dependent orthotactic rules were selected as targets. Foils were written to match the target words' phonological pronunciation, but also to include a violation of the context-dependent orthotactic rules.

Word reading accuracy. As part of an external measure designed to assess word reading accuracy, participating children individually completed a set of 87 vowelized word reading items. The words included in the item set were taken from Authors. Words were presented in order of increasing difficulty (Authors). Participants were instructed to read each word out loud with its complete vowelization. Children received one point for reading the word accurately and zero points for an incorrect response. Although this tool has not yet been thoroughly vetted and validated, it was included to provide an external comparison for the orthographic choice task, given the limited body of assessments currently available to assess literacy development in Arabic. Word reading accuracy has been shown to be positively associated with measures of orthographic knowledge (Apel et al., 2006; Berninger, Abbott, Vermeulen, & Fulton, 2006; Cunningham et al., 2001; Deacon, et al., 2013; Roman, et al., 2009). Cronbach's alpha for the word reading measure was 0.98.

Participants

The study sample included 201 Arabic-speaking third grade students attending public schools in the United Arab Emirates. The participants were recruited from elementary public schools in Dubai that were randomly selected from a list of elementary public schools provided by the Ministry of Education in Dubai. With approval from the Research Ethics Board, potential participants were then randomly sampled from a list of students enrolled at those schools provided by the United Arab Emirates Ministry of Education and recruited to participate. The children came from diverse geographical locations and completed testing during the first term of the 2014 school year (mean age = 8.08 years, *SD* = 0.45). All the participants were native speakers of Arabic and attended schools in which the language of instruction is Arabic. The sample was 50% female (*n* = 99).

The first author, who is a native speaker of Arabic, assessed the participants. The children viewed the orthographic choice items in a paper-based format and completed three practice items before beginning the assessment to ensure their understanding of the task. On the practice items, the evaluator provided corrective feedback as needed. Children were shown the two words presented for each item and instructed to circle the word that was spelled correctly. This forced-choice format yielded items that were scored as either correct or incorrect by the test administrator. Children received one point for a correct answer and zero points for an incorrect answer. The participants' total scores were then obtained by summing the points earned across the full measure.

Data Analysis

Several steps were taken to ensure retention of the most valid and reliable items for the orthographic knowledge measure, providing the best balance of statistical fit and theoretical content validity. These included descriptive statistics, evaluating the dimensionality of the measure, measure refinement based on item loadings, and reliability analysis of the final item set. Each step included a review of the findings to assess their consistency with the rules of Arabic orthography. This was done to reduce the likelihood of spurious findings that could negatively impact the measure's content validity.

To obtain a general overview of item functioning, item-level percent correct, its associated standard deviation, and corrected item-total correlations were examined. These descriptive statistics were inspected for evidence of problems in item functioning (e.g., all children responded to a single item correctly or incorrectly). Any item that (a) was answered correctly at a rate above 95% or below 5% and (b) had a negative item-total correlation was flagged. All flagged items were then re-examined by the first author for inconsistencies with their theoretical construction and intended difficulty. Any item with a mismatch between its observed percent correct and theoretical construction was removed from subsequent analysis.

Dimensionality. The underlying factor structure of the items was preliminarily evaluated using a confirmatory factor analysis approach with unweighted least squares means and variance (ULSMV) estimation in Mplus 7.4 (Múthen & Muthen, 2015). ULSMV was selected over weighted least squares means and variances (WLSMV) estimation because of the relatively small size of the sample, which can negatively bias the quality of the WLSMV tests of model fit (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009). Although the measure was written to assess a single underlying construct, the creation of items with different types of targets and foils may have resulted in clustering of children's responses by these characteristics. Unidimensionality could not be assumed without adequate vetting. Failure to account for any underlying similarities among subgroups of items within the full measure would violate the assumption of unidimensionality and consequently invalidate the use of the measure as a single scale (Yang & Kao, 2014). Therefore, five competing models, constructed based on the item characteristics, were compared against a unidimensional (one-factor) model of orthographic knowledge.

The base model, the one-factor model, served as the comparison for each competing model and is shown in Figure S1 (model A). All the competing models were tested against this model to assess change in model fit relative to the parsimony of the model. Because the unidimensional model was the simplest and most parsimonious model tested, any competing models would have to improve the fit significantly to suggest rejection of the unidimensional model. Criteria for model fit statistics were based on Kline's recommendations for indicating a good fit (2005). That is, a root mean square error of approximation (RMSEA) below 0.10, a comparative fit index (CFI) above 0.90, and Tucker Lewis index (TLI) above 0.90. Model comparisons were conducted using the DIFFTEST option for chi-square difference testing in Mplus (Muthén & Muthén, 2015).

One set of competing models was specified based on features of the target words and is shown in Figure S2. First, a model allowing for clustering of children's responses by the word class of the target item (adjective, noun, pronoun, or verb, as depicted in model B) was compared against the unidimensional model. Of the 60 included items, 8 were classified as adjectives, 33 as nouns, 4 as pronouns, and 15 as verbs. Next, a model accounting for the morphological complexity of the target words was tested (simple vs. complex as defined by Boudelaa & Marslen-Wilson, 2005, and Authors; depicted in model C). A total of 46 items were identified as simple, whereas 14 were complex. A model based on the phoneme-grapheme correspondence of the target words was also tested. This model included classifications of the words as fully transparent (i.e., each phoneme corresponded with a single, regular grapheme within the target word) or more opaque (e.g., coarticulation, "sun letters", or tanween). For a coarticulation example, the /d/ consonant is pronounced as an emphatic /dˤ/ because of the adjacent emphatic /dˤ/ "ض" letter as in the word "frog" (ضفدع). In the examples of "sun letters", the /l/ letter in the determiner "the" /al/ is not pronounced when one of the "sun letters" follows /al/. Another example of orthographic opacity is when two different graphemes (regular alef "ا"or "broken alef" "ى") share the same sound /a:/. It is worth noting that in the case of the broken alef "ى", another factor contributes to its opacity is its visual similarity with the letter "y" (ي) with only two dots underneath the letter " ي" (y) making both letters visual neighbors (Asaad & Eviatar, 2013). Slightly over half of the items were classified as transparent (n = 35). The remaining 25 items were classified as opaque (model D).

Two additional models, specified based on the differences between the targets and foils, were compared against the unidimensional model and are illustrated in Figure S3. The first model was a phonological match model, where items that included target-foil pairs with different pronunciation were loaded onto one factor and those that included target-foil pairs with the same pronunciation were loaded onto a second correlated factor (model E). The second additional model was specified based on the general rule of Arabic orthography violated. Items including foils that violated rules of morphological orthography loaded onto one factor, those that violated rules of orthotactic loaded onto another factor, and those that violated rules of phoneme-grapheme correspondence were loaded onto a third factor (model F). As an additional check of the dimensionality of the measure, the eigenvalues were plotted in a scree plot for the model with 59 items.

Measure refinement. Upon identification of the model best representing the underlying factor structure of the measure, items that did not appear to fit into that structure were systematically examined for consistency with their theoretical construction and removed if appropriate. We retained items that both fit well, defined having a standardized loading onto an identified factor of Arabic orthographic knowledge of 0.30 or higher, and that matched the theoretical framework of the scale. The overall measure and individual item functioning were re-evaluated at each step of this process. All items identified as not fitting well through examination of factor loadings were reviewed and discussed carefully prior to removal from the scale. Chi-square difference tests were used to re-assess the fit of the scale after each item or group of items was removed (through constraining item loadings on the underlying factor to zero). A non-significant chi-square test indicated that removing the item from the scale did not significantly worsen the fit of the overall model. Items were removed in in order of their fit within the model, quantified by their factor loading. Groups of items that exhibited the same size loadings were removed simultaneously.

To obtain an estimate of the internal consistency reliability for the final set of items, McDonald's coefficient omega (or omega total; McDonald, 1999) was computed using the item loadings obtained from the final factor analysis. Coefficient omega is considered a robust metric of internal consistency reliability (McNeish, 2017). Coefficient omega allows for item loadings onto the underlying factor to differ in magnitude. Given that the loadings are not expected to be consistent for all items, this metric is a good estimate of internal reliability. Coefficient omega may be interpreted similarly to coefficient alpha (McDonald, 1999), with values below 0.80 being questionable and those above 0.90 being preferred.

Upon identification of the final set of items, two models were examined to evaluate individual item functioning within the context of the final scale. A two-parameter (2PL) model, which allows items to vary in both difficulty and discrimination, was compared against a one-parameter (1PL) model, which allows items to vary in difficulty but holds discrimination constant. Both models were estimated using ULSMV estimation and compared using the DIFFTEST option in Mplus (Muthén & Muthén, 2015).

External validity. The final set of orthographic knowledge items were then scored according to the final retained model. A Pearson's *r* correlation coefficient was obtained between children's scores on the refined orthographic knowledge task and those on the word reading accuracy measure. Although the external measure is empirically limited, this correlation may provide preliminary information about the relations among literacy constructs in Arabic. Additionally, it serves as an indicator of whether or not the orthographic choice task is measuring its intended underlying construct of orthographic knowledge in Arabic.

Results

Descriptive Findings

Initial examination of the descriptive statistics for the full 60 items (see Table 1) revealed that the percent correct values for the items ranged from 31% correct to 99% correct, with an average of 67% (*SD* = 0.16). One item met criteria for exclusion based on the descriptive results. That item both (a) yielded a percent correct value above 95% and (b) had a negative item-total correlation. Upon theoretical examination of the item, it was determined to be functioning differently than originally anticipated (i.e., no reasonable theoretical rationale supporting the high percent accuracy obtained for the item) and was removed from subsequent analyses. No other items exhibited percent correct values above 95% or below 5% within the participant sample.

Dimensionality

Results from the confirmatory factor analysis revealed that the 59-item unidimensional model (see Model A shown in Figure S1) did not meet criteria for good fit to the data, but may be considered to be in the acceptable range: χ^2 (1652) = 1725.083, RMSEA = 0.015 (90% CI = 0 – 0.023), CFI = 0.875, TLI = 0.870). None of the competing models resulted in a significantly better fit to the data. Model fit statistics for each of the models tested are provided in Table 2, along with the results from the chi-square difference tests used to compare the models against the unidimensional model. Additionally, the scree plot of eigenvalues supported a one-factor solution, with a clear elbow in the plot between one and two factors (see Figure 1).

Measure Refinement

Of the 59 items included in the dimensionality analysis, 27 items loaded onto the unidimensional orthographic knowledge factor at less than 0.30. Following the pre-established steps for item evaluation, these items were systematically re-examined and removed as determined to be appropriate. None of these item removals resulted in a significantly worse fit for the overall scale.

The final scale included 32 items loaded onto a single underlying factor (χ^2 (464) = 502.46, RMSEA = 0.020 (90% CI = 0 – 0.032), CFI = 0.933, TLI = 0.929). The final items included target words representing all four of the word classes originally targeted, targets with simple and targets with more complex morphology, and transparent and more opaque target words. Similarly, the foils were representative of those originally included: both pseudohomophones and non-pseudohomophones were included in the final set of items. The only orthographic feature originally targeted that was not represented in the final item set was the addition of a grapheme to a foil when conventions of Arabic dictate that it should not be written.

Constraining the item loadings (i.e., discrimination) to be equivalent did not result in a significantly worse fit to the data: $\Delta\chi^2(31) = 37.57$, $p = .194$. Consequently, this constraint was retained. Item results from are provided in Table 3. Item characteristic curves, which represent the likelihood of a correct response on a given item conditional on a child's underlying ability, are shown in Figure 2. The test information function curve is provided in Figure 3. Both figures suggest that the developed scale functions optimally for children with lower levels of performance. The sum score on the refined tool was consequently obtained, as indicated by the one-parameter model, and correlated with the external measure of word reading accuracy. Results revealed a significant, positive correlation between children's sum scores on the two measures ($r = 0.793$, $p < .0001$ (95% CI = 0.734 - 0.839).

McDonald's coefficient omega total was computed as a measure of internal reliability for the subset of items retained in the final scale. The model met the assumption of unidimensionality required for this reliability estimate (McNeish, 2017). Given the obtained factor loadings and standard error, the estimate for coefficient omega was 0.987.

Discussion

The purpose of the present study was to examine the psychometrics of a tool designed to assess orthographic knowledge at the lexical level in Arabic. The tool was an orthographic choice task and included 60 items designed for monolingual Arabic-speaking children in Grade 3. The items were written to evaluate a range of skills based on key features of Arabic orthography (e.g., morphological, grammatical, diglossic, and coarticulation features). Measure refinement yielded a set of 32 items which emphasized a range of different features of Arabic orthography and loaded reliably onto a single underlying latent factor.

The final item set appeared to discriminate particularly well among children with lower levels of orthographic knowledge relative to the rest of the sample. This was evidenced by the item characteristic curves, test information curve, and results from the 1-PL model. Children with ability levels below average (*theta* below zero) varied sufficiently in their performance on the items to allow for more precise measurement of their ability. The measure discriminated less effectively among children with very low orthographic knowledge (i.e., two standard deviations below average) and among children with higher ability levels (i.e., 0.75 or more standard deviations above average).

Results suggest that orthographic knowledge at the lexical level may be a unidimensional construct in Arabic. The single-factor model was found to have the best fit, and to explain children's performance on the task better than all the competing multidimensional models. The competing models tested a variety of other possible underlying factors representing orthographic features of Arabic, none of which were found to improve the model fit significantly. Although further work is needed to examine the dimensionality of Arabic orthographic knowledge – and orthographic processing generally – on a broader scale, these preliminary results suggest that lexical orthographic knowledge may be a unitary ability indicated by individuals' ability to distinguish correct versus incorrect spelling in Arabic. The unidimensionality of our findings is in accordance with findings by Commissaire et al. (2014) who found a one factor model of orthographic knowledge to capture both lexical and sub-lexical tasks in both French and English, suggesting a common orthographic knowledge skill across languages and variables.

The significant correlation between the final item set and the external measure of word reading accuracy are consistent with prior work that has identified positive associations, and potential overlap, between orthographic knowledge and word reading measures (Cunningham et al., 2001; Deacon et al., 2013, 2019; Mimeau et al., 2018; Pasquarella et al., 2014). At a minimum, the observed relation suggests that this Arabic orthographic choice task measures a literacy-focused construct closely associated with Arabic word reading. The separability of, and potential directionality between, these constructs requires further examination (e.g., Deacon et al., 2019). At present, these findings suggest that the developed measure has utility in assessing a construct related to reading in Arabic.

Limitations

This study has several limitations that should be noted. Perhaps most importantly, there is some discord regarding the construct validity of the orthographic forced choice task at the lexical level. Some scholars suggest that this task could be measuring the same construct as word reading (Deacon et al., 2019). In other words, this measure may be susceptible to circularity of measurement (Burt, 2006; Castles & Nation, 2006; Deacon et al., 2012), a valid concern given that the external criterion measure included in the present study was a word reading task, and that a high correlation that was found between students' scores on the orthographic choice and word reading accuracy tasks. Deacon and colleagues (2012, 2018, 2019) suggested that this measure could be tracking the outcome of reading rather than an ability that contributes to reading development. They further noted that the potentially interconnected nature of the measures would apply to both the lexical and sublexical levels (Deacon et al., 2012). Given the need for standardized measures and the evidence in support of the use of orthographic choice tasks (Conrad, 2008; Deacon et al., 2013; Olsen et al., 1994), this research can serve as a basis for comparison for future work utilizing alternative tasks. As more Arabic-language assessments of orthographic knowledge and other related skills are developed, the validity of this measure can be assessed by examining its function relative to other validated tools. This will assist practitioners and researchers to better understand both the utility of the developed measure and similar measures.

It is also worth highlighting the fact that all the study participants spoke the same dialect of Arabic and were in third grade. Consequently, generalization of the findings should be made with caution. Results need to be replicated with additional samples of students from other countries and of different ages, in order for the possible influences of educational system, socioeconomic status, and spoken dialect to be understood.

Future Directions

Researchers may consider adapting this task to pilot the assessment of orthographic knowledge at the sublexical level. This may be done by designing a word-like task based on the orthographic patterns represented in this measure. Embedding these orthographic patterns in pseudowords would assist researchers in exploring the feasibility of assessing orthographic knowledge at the sublexical level in Arabic, similar to the studies by Cassar and Treiman (1997), Pacton et al. (2001), Commissaire et al. (2014), and Deacon et al. (2013). This could be particularly informative for assessing children's tacit orthographic knowledge prior to formal reading instruction.

Work is also needed to understand how orthographic knowledge fits into the literacy development of Arabic-speaking individuals. There is some evidence that a child's ability to complete an orthographic choice task may be the result of developing sufficient literacy skills, rather than being a unique ability fundamental to literacy acquisition (Deacon et al., 2012). To assess the role of orthographic knowledge in Arabic literacy development, additional standardized and validated measurement tools are needed. Specifically, scales including items with more open formats and reduced reliance on visual recognition of words (e.g., spelling tasks) may be valuable. With a more comprehensive battery of tools to assess reading and writing development in Arabic, researchers and practitioners may be better equipped to investigate and support the literacy of Arabic-speaking individuals.

Practical Implications

Despite the limitations of present study, the tool examined has the potential to serve as a key resource for educators and researchers, particularly because of the current lack of standardized assessments of Arabic literacy development. This measure may provide insight into distinguishing strengths and weaknesses among Arabic-speaking children who are learning to read (Cassar & Treiman, 1997; Ehri, 2017; Pacton et al., 2001; Treiman & Kessler, 2006). Children's performance on this measure may assist educators in baseline assessment of children's orthographic knowledge and in monitoring progress in word reading over time. Although normative data is not yet available for this tool, educators may use it as a criterion-referenced measure. Because it covers a broad array of orthographic patterns and rules in Arabic, error analysis may provide practitioners with insight into individuals' strengths and weaknesses (Tucker et al., 2016). As literacy is an important part of the responsibilities of speech-language pathologists (American Speech-Language-Hearing Association, 2001), this measure provides an important tool for clinical practitioners.

Notably, the results from this research indicate that the tool can be scored using a simple sum-score system. This is valuable because it can be scored simply by awarding a single point for each correct response and then summing those points upon completion of the task. The reliable nature of the final item set provides preliminary evidence for its use both in research and in educational practice.

A lack of valid assessment tools is detrimental to any researcher. However, developing standardized tools is an iterative process that must begin with some work focused on internal psychometrics and reliability. We consider the work completed in the current study as a step forward in refining and standardizing tools to assess literacy in Arabic. This is a priority for policy makers, researchers, and practitioners, particularly considering that standardized Arabic is the written language used in all the Arab countries.

Conclusion

The purpose of this study was to provide foundational information about the assessment of orthographic knowledge in Arabic. The findings indicate that this orthographic choice task provides reliable information about a single underlying ability for Arabic-speaking children in Grade 3. Further research is needed to evaluate the reliability and validity of this tool across different age groups, as well as its validity in relation to others reading skills such as reading accuracy and fluency. The current study is an early step on a long road toward understanding literacy development in Arabic.

References

Abu-Rabia, S. (1996). The role of vowels and context in reading of highly skilled native Arabic readers. *Journal of Psycholinguistic Research, 25,* 629-641. https://doi.org/10.1007/BF01712413

Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research,* 28, 93-101. https://doi.org/10.1023/A:1023291620997

Abu-Rabia, S. (2000). Effects of exposure to literary Arabic on reading comprehension in diglossic situation. *Reading and Writing: An Interdisciplinary Journal, 13*, 147-157. https://doi.org/10.1023/A:1008133701024

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing: An Interdisciplinary Journal,* 14, 39-59. https://doi.org/10.1023/A:1008147606320

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English-speaking Canadian children. *Journal of Psycholinguistic Research, 31*, 661-678. https://doi.org/10.1023/A:1021221206119

American Speech-Language-Hearing Association. (2001). *Roles and responsibilities of speech-language pathologists with respect to reading and writing in children and adolescents* [Position Statement]. Retrieved from http://www.asha.org/policy/PS2001-00104/#FN1

Apel, K. (2010). Kindergarten children's initial spoken and written word learning in a storybook context. *Scientific Studies of Reading, 14*, 440-463. https://doi.org/10.1080/10888431003623496

Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in Schools*, *42*(4), 592–603. https://doi.org/10.1044/0161-1461(2011/10-0085)

Apel, K., Wolter, J. A., & Masterson, J. J. (2006). Effects of phonotactic and orthotactic probabilities during fast mapping on 5-year-olds' learning to spell. *Developmental Neuropsychology, 29*, 21-42. https://doi.org/10.1207/s15326942dn2901_3

Asaad, H. & Eviatar, Z. (2013). The effects of orthographic complexity and diglossia on letter naming in Arabic: A developmental study. *Writing Systems Research, 5,*156-168. http://dx.doi.org/10.1080/17586801.2013.862163

Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography: Evidence from phonological and orthographic lexical decisions. *Cognitive Neuropsychology*, *25*(5), 653–676. https://doi.org/10.1080/02643290802221404

Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of Psycholinguistic Research, 39*, 141-163. https://doi.org/10.1007/s10936-009-9130-6

Berninger, V. W., Abbott, R. D., Vermeulen, K., & Fulton, C. M. (2006). Paths to reading comprehension in at-risk second grade readers. *Journal of Learning Disabilities, 39*, 334-351. https://doi.org/10.1177/00222194060390040701

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition, 81*, 65-92. https://doi.org/10.1016/S0010-0277(01)00119-6

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes, 20,* 207-260. https://doi.org/10.1080/01690960444000106

Boudelaa, S., & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes, 28,* 1453-1473. https://doi.org/10.1080/01690965.2012.719629

Brown, A. (2016). The challenges of translating the U.S. census questionnaire into Arabic. Washington, DC: Pew Research Center. Retrieved from http://pewrsr.ch/1XVvdmc

Burt, J. S. (2006). What is orthographic processing skill and how does it relate to word identification in reading? *Journal of Research in Reading, 29*, 400-417. https://doi.org/10.1111/j.1467-9817.2006.00315.x

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89,* 631-644. https://doi.org/10.1037/0022-0663.89.4.631

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151-179). New York, NY: Psychology Press. https://doi.org/10.4324/9780203841211

Chambré, S. J., Ehri, L. C., & Ness, M. (2017). Orthographic facilitation of first graders' vocabulary learning: Does directing attention to print enhance the effect? *Reading and Writing, 30*, 1137-1156. https://doi.org/10.1007/s11145-016-9715-z

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review*, 40, 287-309. https://doi.org/10.17763/haer.40.2.y7u0242x76w05624

Commissaire, E., Pasquarella, A., Chen, X., & Deacon, S. H., (2014). *Written Language & Literacy, 17,* 16-39. http://dx.doi.org/10.1075/wll.17.1.02com

Conrad, N. J. (2008). From reading to spelling and spelling to reading: Transfer goes both ways. *Journal of Educational Psychology, 100*, 869-878. https://doi.org/10.1037/a0012544

Conrad, N. J., & Deacon, S. H. (2016). Children's orthographic knowledge and their word reading skill: Testing bidirectional relations. *Scientific Studies of Reading*, *20*(4), 339–347. https://doi.org/10.1080/10888438.2016.1183128

Cunningham, A. E., Perry, K. E. & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing: An Interdisciplinary Journal, 14***,** 549-568. https://doi.org/10.1023/A:1011100226798

Daniels, P. T. (1992). The syllabic origin of writing and the segmental origin of the alphabet. In P. A. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 83-110). Philadelphia, PA: John Benjamins Publishing Company.

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading,* 22, 101-116. https://doi.org/10.1080/10888438.2017.1379082

Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg? Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition*, *122*(1), 110–117. https://doi.org/10.1016/j.cognition.2011.09.003

Deacon, S. H., Commissaire, E., Chen, X., & Pasquarella, A. (2013). Learning about print: The development of orthographic processing and its relationship to word reading in first grade children in French immersion. *Reading and Writing: An Interdisciplinary Journal, 26,* 1087-1109. https://doi.org/10.1007/s11145-012-9407-2

Deacon, S. H., Pasquarella, A., Marinus, E., Tims, T., & Castles, A. (2019). Orthographic processing and children's word reading. *Applied Psycholinguistics*, *40*(2), 509–534. https://doi.org/10.1017/S0142716418000681

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*, 116-125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167-188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18,* 5-21. https://doi.org/10.1080/10888438.2013.819356

Ehri, L. C. (2017). Orthographic mapping and literacy development revisited. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 127-146). Philadelphia, PA: John Benjamins Publishing Company.

Eviatar, Z., & Ibrahim, R. (2000). Bilingual is as bilingual does: Metalinguistic abilities of Arabic-speaking children. *Applied Psycholinguistics, 21*, 451-471. https://doi.org/10.1017/S0142716400004021

Ferguson, C. A. (1959). Diglossia. *Word, 15,* 325-340. https://doi.org/10.1080/00437956.1959.11659702

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16,* 625-641. https://doi.org/10.1080/10705510903203573

Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, *94*(4), 669–686. https://doi.org/10.1037/0022-0663.94.4.669

Gordon, R. G. (2005). *Ethnologue: Languages of the world* (15th Ed.). Dallas, TX: SIL International.

Ho, C.S.-H., Ng, T.-T., & Ng, W.-K. (2003). A "radical" approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research,* 35, 849–878. https://doi:10.1207/s15548430jlr3503_3

Kaefer, T. (2016). Integrating orthographic and phonological knowledge in early readers:

Implicit and explicit knowledge. *Child Development Research*, *2016*. https://doi.org/10.1155/2016/6036129

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Kuo L-J, Kim T-J, Yang, X., Li, H., Liu Y., Wang, H., Hyun Park, J. & Li, Y. (2015). Acquisition of Chinese characters: The effects of character properties and individual differences among second language learners. *Frontiers in Psychology, 6,* 986, 1-10. https://doi: 10.3389/fpsyg.2015.00986

Layes, S., Lalonde, R., & Rebaï, M. (2015). Reading speed and phonological awareness deficits among Arabic-speaking children with dyslexia. *Dyslexia, 21*, 80-95. https://doi.org/10.1002/dys.1491

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Maamouri, M. (2009). *Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region*. Washington, DC: World Bank, Mediterranean Development Forum.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives, 11*, 71-101. https://doi.org/10.1080/15366367.2013.831680

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Erlbaum.

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*, 412-433. https://doi.org/10.1037/met0000144

Mimeau, C., Ricketts, J., & Deacon, S. H. (2018). The role of orthographic and semantic learning in word reading and reading comprehension. *Scientific Studies of Reading, 22,* 384-400. https://doi.org/10.1080/10888438.2018.1464575

Mirdehghan, M. (2010). Persian, Urdu, and Pashto: A comparative orthographic analysis. *Writing Systems Research, 2,* 9-23. https://doi.org/10.1093/wsr/wsq005

Muthén, L. K., & Muthén, B. O. (1998-2015*). Mplus user's guide* (7th Ed). Los Angeles, CA: Muthén & Muthén.

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*, 71-84. https://doi.org/10.1016/j.jecp.2006.06.004

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243-277). Baltimore: Paul H. Brookes Publishing Co.

Ouellette, G. P., & Sénéchal, M. (2008a). A window into early literacy: Exploring the cognitive and linguistic underpinnings of invented spelling. *Scientific Studies of Reading, 12*, 195-219. https://doi.org/10.1080/10888430801917324

Ouellette, G., & Sénéchal, M. (2008b). Pathways to literacy: A study of invented spelling and its role in learning to read. *Child Development*, 79*,* 899-913. https://doi.org/10.1111/j.1467-8624.2008.01166.x

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401-426. https://doi.org/10.1037/0096-3445.130.3.401

Pasquarella, A., Deacon, S. H., Chen, X., Commissaire, E., & Au-Yeung, K. (2014) Acquiring orthographic processing through word reading: Evidence from children learning to read French and English. *International Journal of Disability, Development and Education, 61*, 240-257. https://doi.org/10.1080/1034912X.2014.932579

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp.145-174). Hillsdale, NJ: Erlbaum.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11,* 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *Decade of behavior. On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-214). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Querido, L., Fernandes, S., Verhaeghe, A., & Marques, C. (2020). Lexical and sublexical orthographic knowledge: relationships in an orthography of intermediate depth. *Reading and Writing*, *0123456789*. https://doi.org/10.1007/s11145-020-10052-2

Rahbari, N., & Sénéchal, M. (2010). Learning to read and spell in Persian: A cross-sectional study from Grades 1 to 4. *Developmental Psychology, 46,* 1514-1527. https://doi.org/10.1037/a0020377

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in grades 4, 6, and 8. *Journal of Experimental Child Psychology, 102*, 96-113. https://doi.org/10.1016/j.jecp.2008.01.004

Rothe, J., Cornell, S., Ise, E., & Schulte-Körne, G. (2015). A comparison of orthographic processing in children with and without reading and spelling disorder in a regular orthography. *Reading and Writing*, *28*(9), 1307–1332. https://doi.org/10.1007/s11145-015-9572-1

Saady, A., Ibrahim, R., & Eviatar, Z. (2015). Language-specific and language-general factors in text reading in Arabic: Evidence from the missing letter effect. *Psicológica, 36*, 123-142. https://link.gale.com/apps/doc/A453289833/IFME?u=tall85761&sid=IFME&xid=ac8acd4f

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics, 24,* 431-451. https://doi.org/10.1017/S0142716403000225

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics, 25,* 495-512. https://doi.org/10.1017/S0142716404001249

Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad & M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 3-28). Dordrecht, the Netherlands: Springer.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14,* 238-242. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568. https://doi.org/10.1037/0033-295X.96.4.523

Shahar-Yames, D., & Share, D. L. (2008). Spelling as a self-teaching mechanism in orthographic learning. *Journal of Research in Reading, 31*, 22-39. https://doi.org/10.1111/j.1467-

9817.2007.00359.x

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*, 151-218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of self-teaching hypothesis. *Journal of Experimental Child Psychology, 72,* 95-129. https://doi.org/10.1006/jecp.1998.2481

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin, 134,* 584-615. https://doi.org/10.1037/0033-2909.134.4.584

Share, D. L., & Daniels, P. T. (2015). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). *Writing Systems Research*, *1*, 17-31. https://doi.org/10.1080/17586801.2015.1016395

Stanovich, K. E., & West, R. F. (1986). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433.

Taha, H. (2016). Deep and shallow in Arabic orthography: New evidence from reading performance of elementary school native Arab readers. *Writing Systems Research, 8,* 133-142. https://doi.org/10.1080/17586801.2015.1114910

Taha, H., & Saiegh-Haddad, E. (2017). Morphology and spelling in Arabic: Development and interface. *Journal of Psycholinguistic Research, 46*, 27-38. https://doi.org/10.1007/s10936-016-9425-3

Treiman, R. (2017). Learning to spell words: Findings, theories, and issues. *Scientific Studies of Reading, 21,* 265-276. https://doi.org/10.1080/10888438.2017.1296449

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology,* 98, 642-652. https://doi.org/10.1037/0022-0663.98.3.642

Tucker, R., Castles, A., Laroche, A., & Deacon, S. H. (2016). The nature of orthographic learning in self-teaching: Testing the extent of transfer. *Journal of Experimental Child Psychology, 145*, 79-94. https://doi.org/10.1016/j.jecp.2015.12.007

Tunmer, W. E., & Nicholson, T. (2011). The development and teaching of word recognition skill. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research: Volume IV* (pp. 405-431). New York, NY: Routledge.

U.S. Census Bureau. (2015). *2009-2013 American Community Survey*. Washington, DC: U.S. Department of Commerce. Retrieved from https://www.census.gov/programs-surveys/acs/

Venezky, R. (1999). *The American Way of Spelling.* New York: Guilford Press.

Wolter, J. A., & Apel, K. (2010). Initial acquisition of mental graphemic representations in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 53*,* 179-195. https://doi.org/10.1044/1092-4388(2009/07-0130)

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry, 26*, 171-177. https://doi.org/10.3969/j.issn.1002-0829.2014.03.010

Table 1

Descriptive Statistics for All Items

Item Number	Percent Accuracy	*SD*	Corrected Item-Total Correlation	**Item Number**	Percent Accuracy	*SD*	Corrected Item-Total Correlation
1	0.89	0.31	0.18	**31**	0.46	0.5	0.17
2	0.58	0.50	0.22	**32**	0.63	0.48	0.06
3	0.80	0.40	0.20	**33**	0.46	0.5	-0.02
4	0.56	0.50	0.16	**34**	0.59	0.49	0.15
5	0.81	0.39	0.42	**35**	0.58	0.49	0.11
6	0.69	0.46	0.24	**36**	0.36	0.48	-0.16
7	0.82	0.38	0.37	**37**	0.64	0.48	0.17
8	0.45	0.50	0.00	**38**	0.69	0.46	0.13
9	0.92	0.27	0.32	**39**	0.64	0.48	0.31
10	0.86	0.34	0.30	**40**	0.39	0.49	0.11
11	0.75	0.43	0.23	**41**	0.99	0.1	-0.01
12	0.31	0.46	-0.01	**42**	0.69	0.46	0.16
13	0.82	0.38	0.34	**43**	0.65	0.48	0.24
14	0.88	0.33	0.30	**44**	0.73	0.44	0.31
15	0.72	0.45	0.32	**45**	0.66	0.47	0.26
16	0.89	0.31	0.28	**46**	0.54	0.5	0.20
17	0.37	0.49	0.13	**47**	0.38	0.49	0.01
18	0.81	0.39	0.26	**48**	0.72	0.45	0.21
19	0.70	0.46	0.23	**49**	0.76	0.43	0.20
20	0.76	0.43	0.31	**50**	0.72	0.45	0.35
21	0.54	0.50	0.27	**51**	0.70	0.46	0.24
22	0.91	0.29	0.35	**52**	0.59	0.49	0.30
23	0.88	0.32	0.36	**53**	0.37	0.48	-0.25
24	0.69	0.46	0.23	**54**	0.73	0.45	0.10
25	0.72	0.45	0.19	**55**	0.49	0.5	0.12
26	0.88	0.32	0.23	**56**	0.68	0.47	0.23
27	0.70	0.46	0.32	**57**	0.61	0.49	0.09
28	0.85	0.35	0.39	**58**	0.63	0.48	0.20
29	0.72	0.45	0.17	**59**	0.64	0.48	0.10
30	0.77	0.42	0.40	**60**	0.69	0.46	0.34

Table 2

Fit Indices for Models of Arabic Orthographic Processing

Model		χ^2	*df*	$\Delta\chi^2$	Δdf	ΔSig.	RMSEA	LB	UB	CFI	TLI
A	Unidimensional	1725.08	1652	--	--	--	0.015	<0.001	0.023	0.87	0.87
B	Target-Based										
	Word Classes	1718.97	1646	5.71	6	.456	0.015	<0.001	0.023	0.87	0.87
	Morphological Complexity	1724.21	1651	0.44	1	.508	0.015	<0.001	0.023	0.87	0.87
	Phoneme-Grapheme	1723.74	1651	2.03	1	.154	0.015	<0.001	0.023	0.87	0.87
C	Foil-Based										
	Phonological Match	1723.92	1651	1.43	1	.232	0.015	<0.001	0.023	0.87	0.87
	Rule Violation	1722.64	1649	0.72	3	.721	0.015	<0.001	0.023	0.87	0.87

Note. $\Delta\chi^2$ is reported for the model comparisons against the 1-factor model. Numbers written in parentheses represent the number of factors included in the model.

Table 3

Item characteristics obtained from final one-parameter model estimated with ULSMV

Item	Item Difficulties	*SE*	Percent Correct
Item 9	-2.933	0.322	0.92
Item 22	-2.799	0.301	0.91
Item 1	-2.617	0.279	0.89
Item 16	-2.561	0.276	0.89
Item 23	-2.507	0.264	0.88
Item 26	-2.507	0.267	0.88
Item 14	-2.454	0.262	0.88
Item 10	-2.257	0.25	0.86
Item 28	-2.211	0.26	0.85
Item 7	-1.954	0.231	0.82
Item 13	-1.954	0.237	0.82
Item 18	-1.836	0.225	0.81
Item 5	-1.798	0.233	0.81
Item 30	-1.547	0.209	0.77
Item 20	-1.479	0.21	0.76
Item 49	-1.455	0.216	0.76
Item 11	-1.446	0.212	0.75
Item 44	-1.291	0.206	0.73
Item 48	-1.228	0.203	0.72
Item 50	-1.228	0.208	0.72
Item 15	-1.192	0.202	0.72
Item 25	-1.192	0.208	0.72
Item 24	-1.072	0.196	0.69
Item 27	-1.072	0.197	0.70
Item 60	-1.044	0.198	0.69
Item 6	-1.042	0.198	0.69
Item 56	-0.985	0.197	0.68
Item 45	-0.868	0.193	0.66
Item 39	-0.727	0.191	0.64
Item 52	-0.479	0.187	0.59
Item 2	-0.405	0.186	0.58
Item 21	-0.247	0.184	0.54

Note. Items are ordered from least difficult to most difficult. Item discriminations were fixed at 0.547 (*SE* = 0.032). Item difficulties can be interpreted as the (z-scored) underlying ability level needed for a child to have a probability of .50 of answering the item correctly. The underlying latent trait mean was set to zero, with a standard deviation of 1.

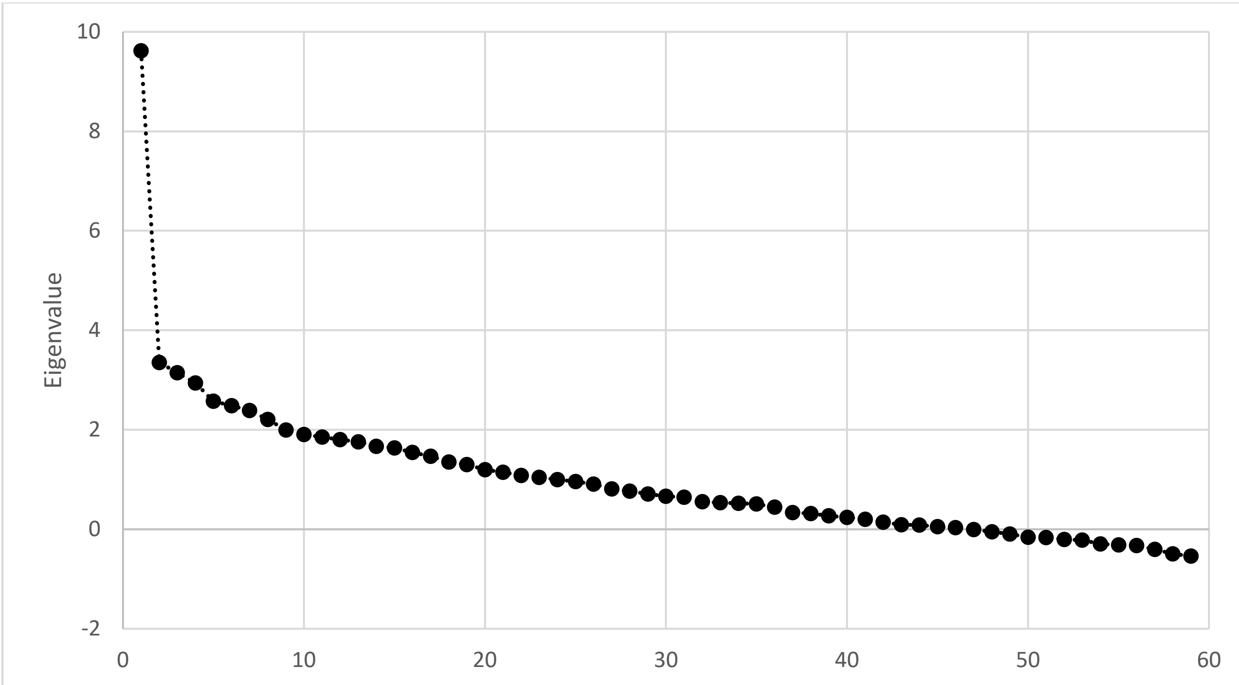


Figure 1. Scree plot showing eigenvalues for model with 59 items.

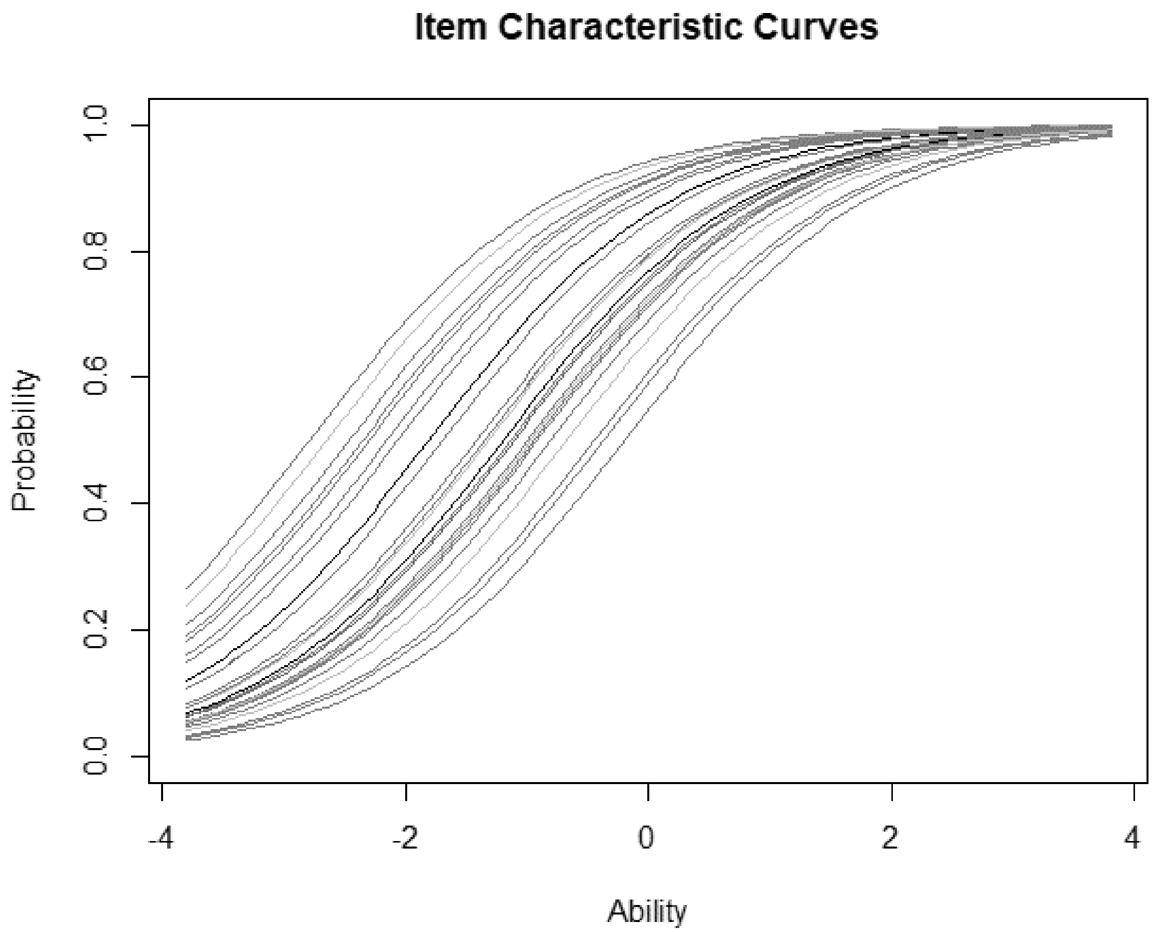


Figure 2. Item characteristic curves for final set of items.

Note. Each line represents a different item's item characteristic curve, which is a function of children's underlying ability (average ability = 0) and the probability of responding correctly to that item.

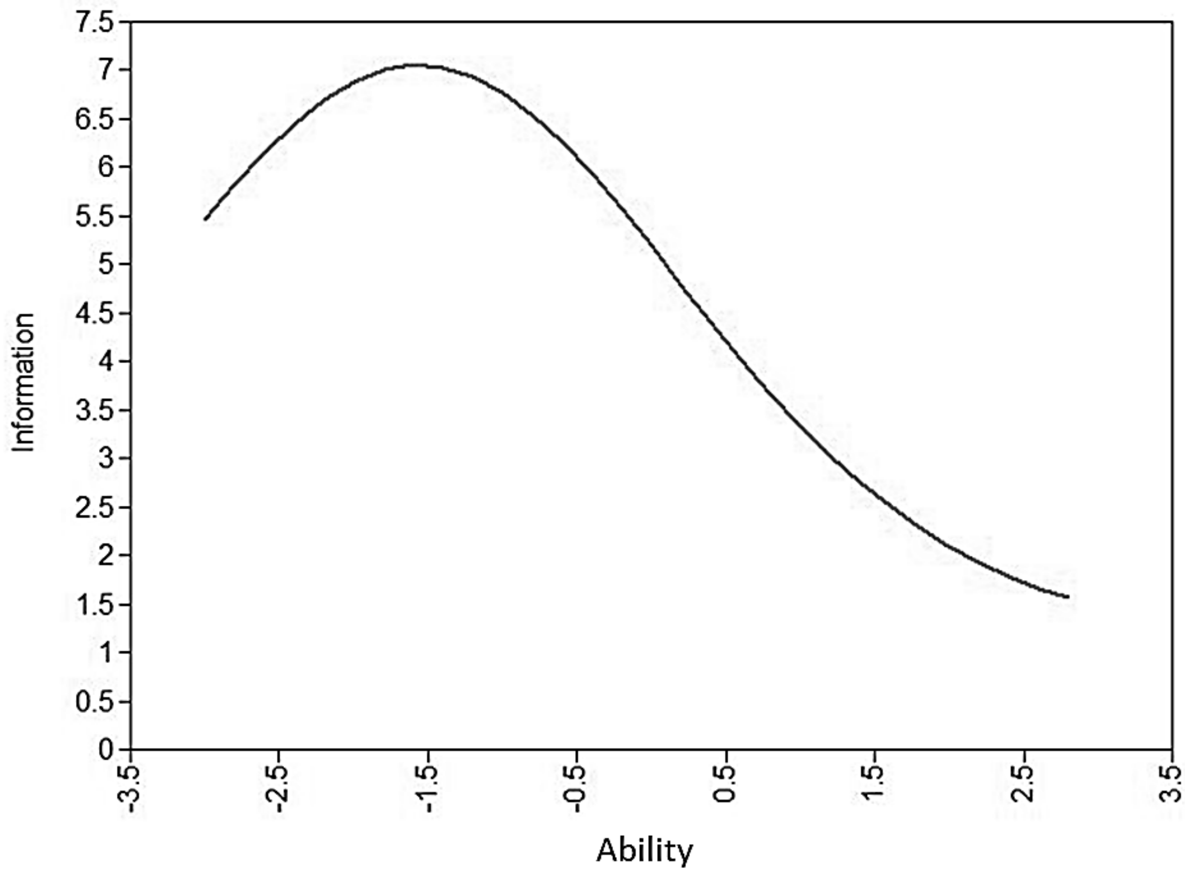
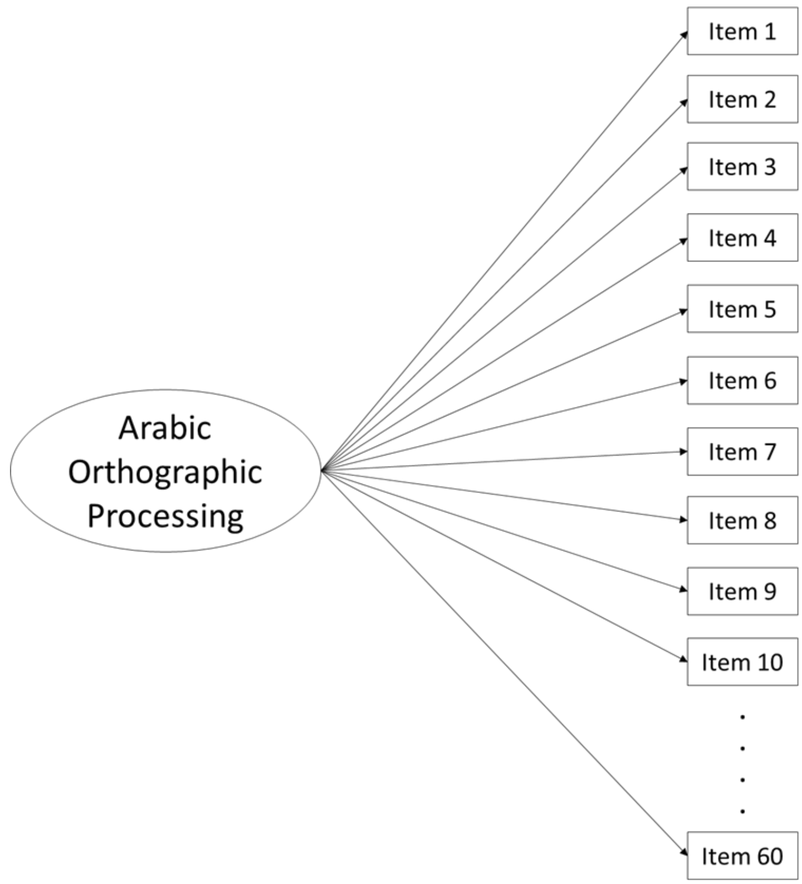


Figure 3. Test information function curve for final set of items.



Model A

Figure S1. Unidimensional model of Arabic orthographic processing.

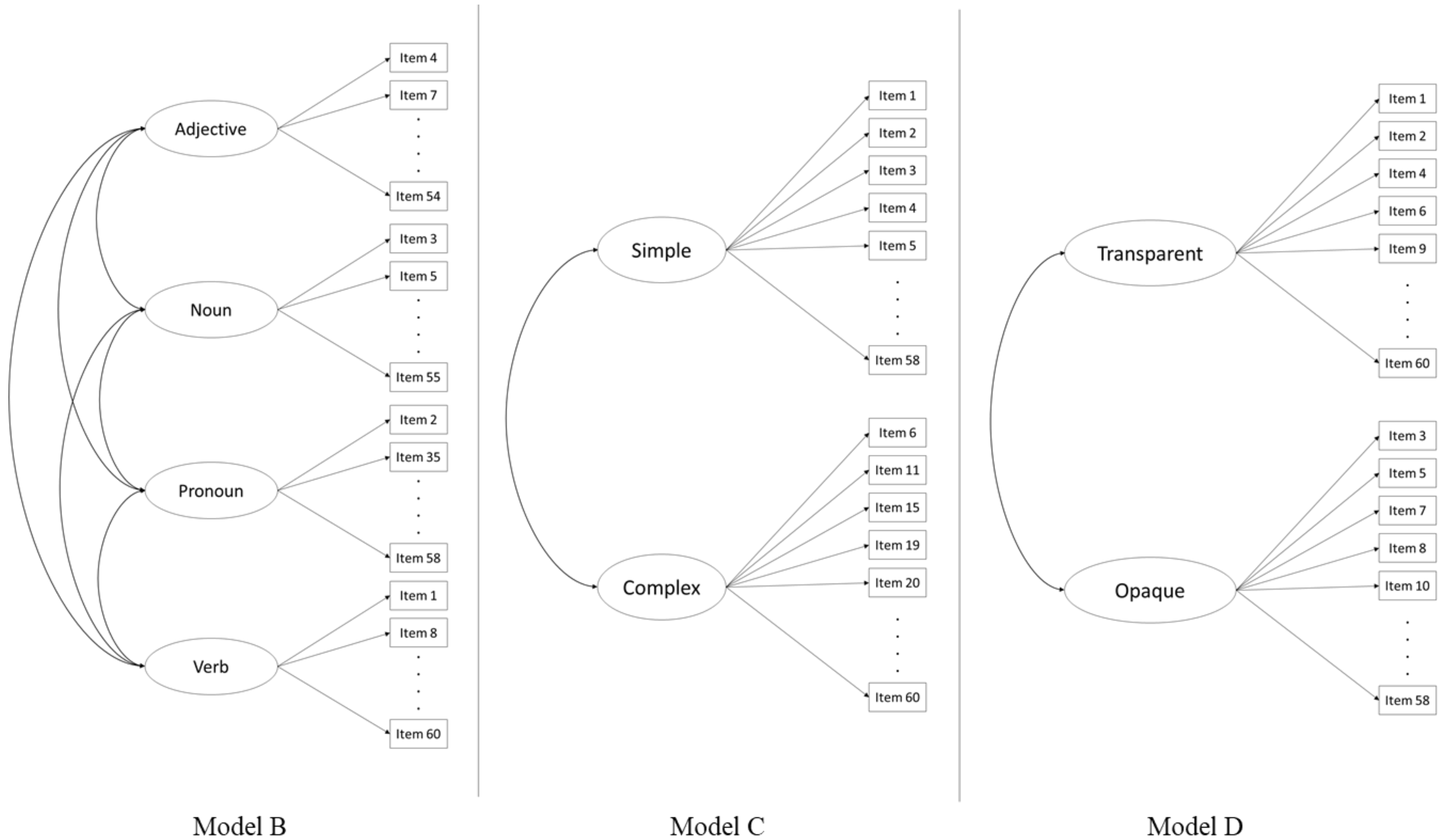


Figure S2. Competing models of Arabic orthographic processing based on target word characteristics.

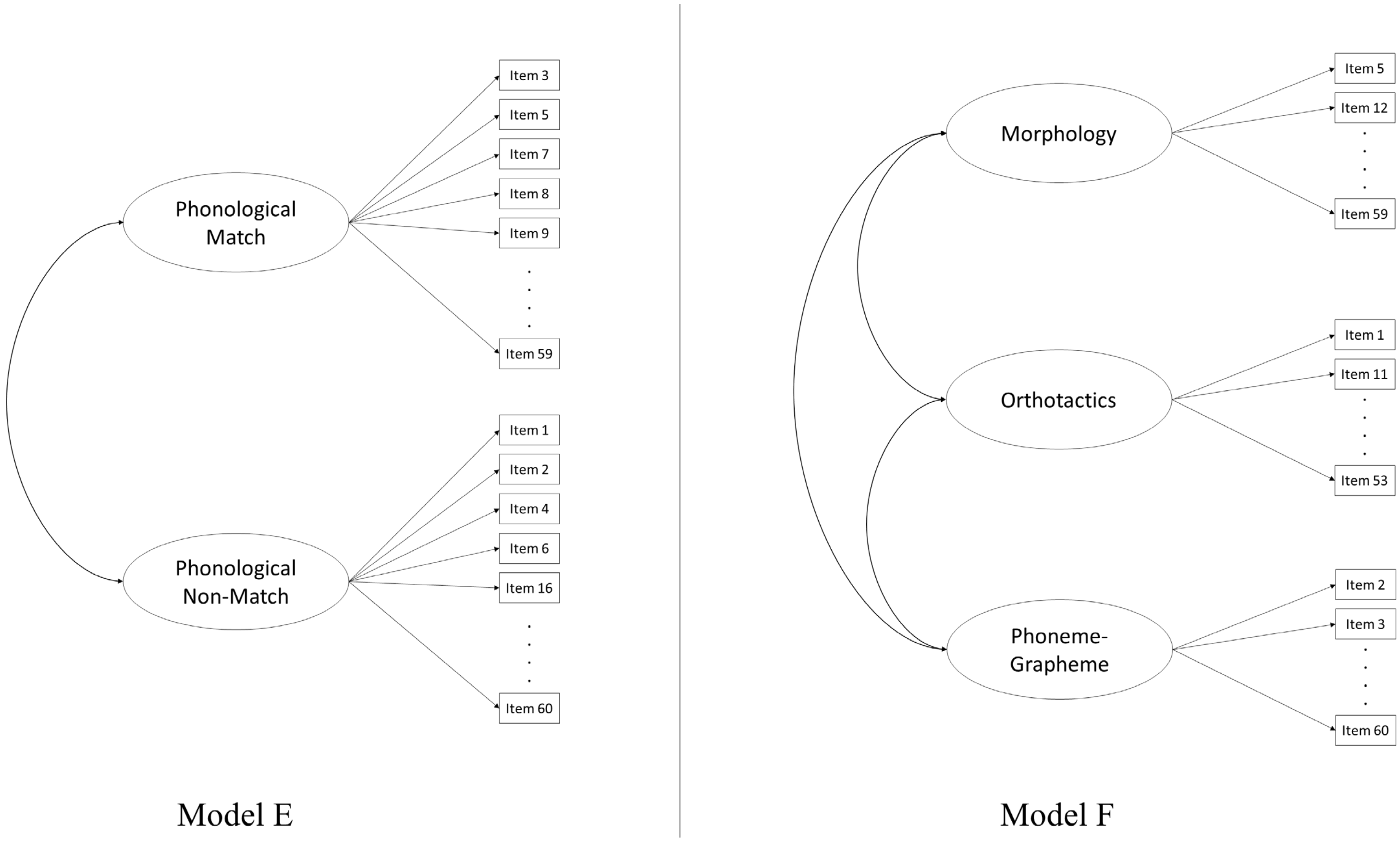


Figure S3. Competing models of Arabic orthographic processing based on foil characteristic