# THE RELIABILITY AND VALIDITY OF SARD: A PILOT STUDY IN UAE PRIMARY-GRADE NATIVE ARABIC READERS

A PREPRINT

**Ahmad Nazzal**[1], **Haitham Taha**[1], **Hanada Taha-Thomure**[1,*], **Rabab Saleh**[1]

[1]ZAI Arabic Language Research Center, Zayed University, Dubai, United Arab Emirates

**Corresponding author: Hanada.Thomure@zu.ac.ae*

**ABSTRACT**

Early identification of students at risk for reading difficulties is essential for successful effective intervention, yet validated reliable screening tools for assessment of reading in Arabic are scarce. Therefore, we developed the Smart Arabic Reading Diagnostic (SARD) tool. SARD is a web-based assessment that measures both accuracy and response time across 16 tasks measuring multiple components of reading. This pilot study aims to prove SARD's reliability, convergent and divergent validity. We administered SARD to 354 native Arabic speaking students in Grades 2–6 from United Arab Emirates schools. We computed grade specific means and standard deviations, reliability indices, partial correlations, and component analysis. Cronbach's $\alpha$ for the seven decoding tasks was 0.91, and item–total correlations ranged from 0.76 to 0.84. Principal component analysis showed a dominant factor explaining 67% of variance, and hierarchical clustering separated letter level from word level tasks. We conclude that SARD is a reliable tool for Arabic phonics skills assessment with excellent internal consistency showing convergent and divergent validity. Future studies will explore SARD's test–pretest reliability, criterion validity, and expand population to include different countries.

***Keywords*** Arabic reading · Reading difficulties · Screening tool · SARD battery

## 1 Introduction

Arab countries continuously score below international level in reading skills[Mullis et al., 2023]. Learners in Arabic encounter several unique linguistic and orthographic characteristics that influence their reading acquisition. For exmaple, the alphabetic system of Arabic language comprises 28 letters often grouped by visual similarity and differentiated by the number or placement of diacritical dots. Letter forms vary according to their position (initial, medial, or final) and connectivity within words, making visual recognition context dependent; these connectivity features have been shown to affect word recognition across proficiency levels [Khateb et al., 2013, Taha, 2016, Taha and Khateeb, 2018]. Another key feature is the use of vowelization diacritics (*Tashkeel*) that represent short vowels and other phonological or grammatical markers. In fully vowelized texts—typically used for learners or isolated words—these diacritics facilitate decoding and reduce ambiguity[Abu-Rabia, 1999]. However, the role of vowelization remains debated, with evidence suggesting it may both enhance accuracy and cause visual overload depending on context [Taha and Saiegh-Haddad, 2017a]. Furthermore, Arabic exhibits classical diglossia, with Modern Standard Arabic (MSA) used in formal contexts and diverse regional dialects in everyday speech [Ferguson, 1959, Ayari, 1996]. This linguistic duality impacts phonological awareness, reading, and writing development [Schiff and Saiegh-Haddad, 2018], and has been shown to influence syntactic and grammatical processing in behavioral and electrophysiological studies [El Idrissi Rioui et al., 2021, Khamis-Dakwar et al., 2012].

Taking into account the aforementioned aspects of Arabic language learning, early detection of reading difficulties is crucial for timely intervention; however, to our knowledge, no classroom-feasible screener tools for Arabic with established psychometric reliability and validity currently exist. As such, we developed the Smart Arabic Reading Diagnostic (SARD) tool. SARD is a web-based assessment designed to evaluate multiple components of Arabic reading fluency and comprehension. By integrating measures of accuracy and response time across key linguistic domains,

SARD aims to provide a comprehensive profile of learners' decoding and comprehension skills. The present pilot study examines its internal consistency and structural validity as an initial step toward establishing SARD as a reliable and valid screening instrument for Arabic reading difficulties.

## 2 Methods

### 2.1 General description

SARD is an expert-designed, cloud-based online screening battery for assessing Arabic reading and spelling skills in grades 1 to 6. It is developed by Dr. Haitham Taha and Dr. Hanada Taha-Thomure, who have over 20 years of experience in Arabic language assessment and exam preparation.

### 2.2 Registration process

Schools apply to SARD by completing an online registration form by visiting www.arabisard.com. After approval, the system automatically sends login credentials to school via e-mail. Upon request a parental registration form is provided. The school assigns a coordinator responsible for logging into the platform and uploading students information. Information collected includes name, grade level, date of birth, nationality, native/non-native Arabic speaker. For research and publications, all data are anonymized for the study purposes and researchers have no access to students names.

### 2.3 Test session

SARD is administered in a quiet classroom. Each student is provided with a computer or a tablet with a microphone and speaker. Students set with 5 meters between them to minimize background noise affecting recording of the answers. The coordinator logs in on behalf the student, ensuring the system is ready for the student to begin the assessment. The tool gives a detailed example of how to respond. Instructions are made audible for all tasks. The student starts the assessment independently once the device is ready. Coordinators are available to provide additional support to students, for example 1st and 2nd graders. The maximum duration of the assessment is 90 minutes. The student can finish earlier. Students can pause the test or withdraw from the test at any point without providing any reason or any penalties. No monetary compensation is provided for students or schools for participating. Once done the student logs out of the system, the platform automatically generates individual reports, and the teacher or administrator receives the assessment report results.

### 2.4 Tasks description

The battery includes 14 to 16 tasks, depending on grade level: 14 tasks for grade 1 and 16 tasks for grades 2 through 6.

#### 2.4.1 Reading letters in isolated form task (RL)

The task assesses knowledge of Arabic letter names presented in isolated form. It measures both recognition and retrieval levels. A total of 28 letters are presented in their isolated form on the screen to the participant. Letters are presented in random order. The participant is required to read the letters as quickly and accurately as possible. Responses are recorded via microphone. Total task completion time and accuracy percentage are calculated.

#### 2.4.2 Reading letters in connected form task (RL-POS)

The task assesses knowledge of Arabic letter names presented in connected forms. It measures both recognition and retrieval levels. To ensure full coverage of the Arabic alphabet, certain letters are presented in their connected form either at the initial, medial, or final position of the word. As such, a total of 36 letters stimuli are presented on the screen to the participant. Letters are presented in random order. The participant is required to read the letters as quickly and accurately as possible. Responses are recorded via microphone. Total task completion time and accuracy percentage are calculated.

#### 2.4.3 Reading syllables task (RS)

The task evaluates basic phonological decoding processes required for segmenting and blending phonemes in short and long syllables in Arabic. Arabic's syllables can be short, consisting of a consonant and a short vowel, or long, consisting of a consonant and a long vowel. The task includes a total of 28 items of syllables. The participant is required

to read the syllables presented on the screen aloud. Responses are recorded via microphone. Both total reading time and accuracy are computed at the end of the task.

#### 2.4.4 Reading real words (RW)

Written words in Arabic may appear with full diacritics (tashkeel) or without.

1. **Reading real words without diacritics task (RW)** For skilled readers, tashkeel is usually absent. The task evaluates oral reading fluency of real Arabic words without full diacritics. It measures the automaticity of recognizing familiar written words. The present task includes a total of 28 items of un-voweled words (without tashkeel). Participants are required to read the words aloud as they appear on the screen. Responses are recorded via microphone. Measures of total reading time, accuracy are calculated.
2. **Reading real words with diacritics task (RW-Tash)** The task evaluates oral reading fluency of real Arabic words with full diacritics. It measures the automaticity of recognizing familiar written words with diacritics. This task includes a total of 30 items fully vowelized words. Participants read the words aloud as they appear on the screen. Responses are recorded via microphone. Measures of total reading time, accuracy are calculated.

#### 2.4.5 Reading pseudowords task (PW)

The task assesses the efficiency of phonological decoding processes when reading novel (not real) words in Arabic orthography. This task includes a total of 25 words pseudowords constructed according to frequent morphological patterns in Arabic. Pseudowords are presented with full diacritics, with 1 pseudoword displayed on the screen at each trial. The participant is required to read the items aloud. Responses are recorded via microphone. Total reading time and accuracy are calculated.

#### 2.4.6 Reading text task (RT)

The task evaluates oral reading fluency of real words embedded in contextual passages. The task does not assess comprehension but rather fluency. The texts were adapted to grade-level characteristics consistent with reading materials in school books. For example, Grade 1: received 17 words stimuli, Grade 2,3: 71 words stimuli, and Grade 4,5,6: 215 words stimuli. Each text is presented on a computer screen, and participants asked to read aloud clearly. Measures of reading time and accuracy are calculated.

#### 2.4.7 Naming tasks

These tasks measure the efficiency of verbal retrieval speed in response to familiar visual stimuli. In the literature, rapid naming tasks are recognized as reliable predictors of oral reading fluency because they parallel the phonological–verbal retrieval demands involved in reading letters and words. Studies have shown that rapid naming speed is strongly associated with overall reading fluency, particularly with oral word reading fluency. Accordingly, three naming tasks were included in the SARD battery:

1. **Letter naming task (LN)** Five letters are presented in random order, with each letter repeated ten times, yielding 50 stimuli in total. Participants are instructed to name them as quickly and accurately as possible. Verbal retrieval rate for letter forms is recorded.
2. **Number naming task (NN)** Five digits are presented in random order, each repeated ten times, for a total of 50 digit stimuli. Participants name them as quickly and accurately as possible. Verbal retrieval rate for digit forms is measured.
3. **Object naming task (ON)** Five familiar objects are presented in random order, each repeated ten times, resulting in 50 object stimuli. Participants name them quickly and accurately. Verbal retrieval rate for object forms is measured.

#### 2.4.8 Listening comprehension task (LC)

The task evaluates the efficiency of listening comprehension in Standard Arabic. It measures oral language processing efficiency, which contributes to understanding language processing more broadly. Age-appropriate texts were presented, adjusted in length and complexity according to age group level. For example, Grade 1: received 33 words stimuli, Grade 2,3: 89 words stimuli, and Grade 4,5,6: 203 words stimuli. The participant listen to the text via headphones and, immediately after, is required to answer multiple-choice questions assessing both explicit and implicit knowledge. Performance is evaluated by accuracy and response time.

#### 2.4.9 Reading comprehension task (RC)

The task evaluates the efficiency of listening comprehension in Standard Arabic. Reading comprehension is assessed by presenting a written text to the participant. Participants could control the pace of reading and could navigate back and forth through the text. Age-appropriate texts were presented, adjusted in length and complexity according to age group level. For example, Grade 2,3: 76 words – 5 questions, Grade 4,5,6: 140 words – 6 questions. After completing the reading, participants answered multiple-choice questions assessing explicit and inferential understanding. Performance was measured by accuracy and response time.

#### 2.4.10 Phonological awareness task (PA)

Phonological awareness was assessed using a phonological deletion task. The task measures verbal working memory processes. The participant heard a word and was then instructed to omit a given syllable and produce aloud the remaining form. The underlying assumption is that the greater the participant's sensitivity to the phonological structure of words, the more efficiently they can perform such manipulations. The task included a total of 15 items items. Responses are recorded via microhpone and analyzed for both accuracy and reaction time.

#### 2.4.11 Orthographic decision task (OD)

The task evaluates the efficiency of orthographic representations of written Arabic words as it relates to accurate spelling processes and to fluent recognition of written words. The task includes a total of 23 items. The participant is required to select the correct spelling form from among homophonic distractors for a given word. Performance is measured in terms of accuracy and response time.

#### 2.4.12 Verbal memory tasks

These tasks measure verbal short-term and working memory processes:

1. **Immediate verbal memory task (IM)** This task evaluates short-term verbal memory. Participants repeat digit sequences immediately after hearing them. Sequences begin with two digits and gradually increase in length. Accuracy of reproduction is recorded.
2. **Verbal working memory task (PWM)** This task assesses verbal working memory, which plays a central role in information processing and learning. Participants hear a sequence of digits and are required to repeat them in reverse order. Performance is measured by the accuracy of reproducing the reversed sequences.

### 2.5 Data processing and analysis

Data were exported from the SARD platform and processed in `Python`. Column labels were cleaned, and all accuracy and reaction time (RT) fields were converted to numeric values (commas replaced with decimals; dashes treated as missing). Ages were parsed to decimal years, and outliers exceeding three standard deviations from the mean were removed. Long-format tables were created for accuracy and RT, then pivoted back to wide format for multivariate analyses.

The seven decoding tasks (letters, positional letters, syllables, real words, real words with diacritics, pseudowords, and text) were treated as the core scale for reliability and dimensionality analyses. Reaction times were log-transformed to reduce skew. Descriptive statistics (means and standard deviations) for each task were computed by grade. Heteroskedasticity-robust ordinary least squares (OLS) models were fitted for each decoding accuracy and log-RT outcome with grade as a linear predictor to quantify developmental trends and obtain slope estimates.

Internal consistency was assessed using Cronbach's $\alpha$, based on the seven decoding accuracy scores. Convergent validity was evaluated via grade-specific Spearman rank correlations ($\rho$) between each SARD decoding task and the passage comprehension score. Discriminant validity was examined using partial Spearman correlations in which letter-naming accuracy was partialed out.

Correlation matrices for accuracy and RT across tasks were visualized as heatmaps. A scatterplot of mean accuracy versus mean log RT for each task and grade included a locally weighted scatterplot smoothing (LOWESS) curve to illustrate the speed–accuracy trade-off. Dimensionality was examined using principal component analysis (PCA) with $z$-standardized scores, and scree plots and loadings were inspected. Hierarchical agglomerative clustering (Ward's linkage on $1 - |\rho|$) grouped tasks by similarity. All analyses were conducted using the `pandas`, `statsmodels`, and `scikit-learn` Python libraries.

# 3 Results

## 3.1 Participants

Participants were 354 native Arabic-speaking students (178 girls, 176 boys) in Grades 2 through 6 from the United Arab Emirates. The mean age was 10.75 years (SD = 1.63) (Table 1). Grade 2 had 87 students, grade 3 had 72 students, grade 4 had 55 students, grade 5 had 68 students, and grade 6 had 72 students. All participants had typical vision and hearing and no known learning or developmental disorders. Informed consent was obtained from parents or legal guardians, and verbal assent was obtained from each child in accordance with institutional ethical guidelines.

Table 1: Demographic characteristics of the study sample

| Gender Distribution | Grade Distribution | Age Summary (years) |
|---|---|---|
| Girl: 178 | Grade 2: 87 | Mean: 10.7 |
| Boy: 176 | Grade 3: 72 | Std: 1.7 |
| Total: 354 | Grade 4: 55 | |
| | Grade 5: 68 | |
| | Grade 6: 72 | |

*Note.* Std = standard deviation.

## 3.2 Descriptive statistics and developmental trends

Letter recognition and positional letter identification were already near ceiling in Grade 2 (95% correct) and showed little change thereafter. In contrast, syllable, real-word, diacritised-word, and pseudoword accuracy rose steadily from Grade 2 (67–84%) to Grade 6 (89–94%). Passage reading accuracy increased from 56% in Grade 2 to 87% in Grade 6 (Table 2). Mean log RTs declined across all tasks, with reductions of roughly 0.3–0.4 (30% faster) between Grades 2 and 5; improvement plateaued by Grade 6. Robust OLS models confirmed these patterns: slopes for real-word, pseudoword, and passage accuracy ranged from 2.3–3.7 percentage points per grade ($p < .05$), whereas letter reading showed no significant slope after Grade 3. Log RT slopes were negative and significant across tasks (0.04–0.09 log ms decrease per grade, $p < .001$), consistent with increasing fluency (Table 3).

Table 2: Scores per grade and task

(a) Cognitive, naming, and phonological tasks

| Grade | IM | LN | LC | NN | ON | OD | PA | PWM |
|---|---|---|---|---|---|---|---|---|
| Grade 2 | $4.25 \pm 3.83$ | $6.20 \pm 3.21$ | $5.91 \pm 2.77$ | $6.51 \pm 3.30$ | $7.57 \pm 3.20$ | $5.98 \pm 1.24$ | $5.37 \pm 4.05$ | $2.60 \pm 3.67$ |
| Grade 3 | $2.52 \pm 2.64$ | $6.32 \pm 2.92$ | $4.19 \pm 2.96$ | $6.50 \pm 3.23$ | $7.40 \pm 3.42$ | $6.55 \pm 1.62$ | $6.37 \pm 3.93$ | $2.48 \pm 3.65$ |
| Grade 4 | $1.83 \pm 2.22$ | $6.15 \pm 2.85$ | $3.18 \pm 2.14$ | $6.60 \pm 3.32$ | $6.31 \pm 3.59$ | $6.13 \pm 1.48$ | $3.49 \pm 3.92$ | $0.36 \pm 1.63$ |
| Grade 5 | $2.55 \pm 2.36$ | $7.38 \pm 2.34$ | $3.45 \pm 2.40$ | $7.02 \pm 3.46$ | $7.32 \pm 3.71$ | $6.28 \pm 1.28$ | $4.77 \pm 4.08$ | $0.89 \pm 2.55$ |
| Grade 6 | $2.52 \pm 3.10$ | $7.54 \pm 2.66$ | $3.55 \pm 2.42$ | $7.38 \pm 3.45$ | $7.34 \pm 3.47$ | $6.79 \pm 1.29$ | $4.75 \pm 4.02$ | $0.81 \pm 2.11$ |

(b) Reading and decoding tasks

| Grade | RC | RL | RL-POS | PW | RW | RW-Tash | RS | RT |
|---|---|---|---|---|---|---|---|---|
| Grade 2 | $5.59 \pm 3.53$ | $8.09 \pm 2.76$ | $7.80 \pm 3.19$ | $4.32 \pm 3.24$ | $5.51 \pm 2.75$ | $5.62 \pm 3.28$ | $7.39 \pm 3.45$ | $5.04 \pm 3.40$ |
| Grade 3 | $3.57 \pm 2.84$ | $8.29 \pm 2.95$ | $8.67 \pm 2.22$ | $5.23 \pm 3.19$ | $7.21 \pm 2.86$ | $6.54 \pm 2.89$ | $8.25 \pm 2.95$ | $5.89 \pm 3.04$ |
| Grade 4 | $3.16 \pm 2.21$ | $8.31 \pm 2.93$ | $8.92 \pm 2.61$ | $5.28 \pm 3.27$ | $8.54 \pm 2.69$ | $6.15 \pm 3.13$ | $8.80 \pm 2.76$ | $4.19 \pm 3.55$ |
| Grade 5 | $3.22 \pm 1.77$ | $8.37 \pm 3.12$ | $8.38 \pm 3.22$ | $6.33 \pm 3.27$ | $8.13 \pm 2.48$ | $7.08 \pm 2.85$ | $8.51 \pm 2.51$ | $5.52 \pm 3.11$ |
| Grade 6 | $3.43 \pm 2.12$ | $8.17 \pm 3.34$ | $8.70 \pm 2.73$ | $6.28 \pm 3.62$ | $8.15 \pm 2.60$ | $7.25 \pm 3.45$ | $8.97 \pm 2.55$ | $5.26 \pm 3.64$ |

*Note.* Values are mean $\pm$ SD. IM = Immediate short-term memory; LN = Letter naming; LC = Listening comprehension; NN = Number naming; ON = Object naming; OD = Orthographic decision; PA = Phonological awareness; PWM = Phonological working memory. RC = Reading comprehension; RL = Reading letters; RL-POS = Reading letters in beginning, middle, and end positions; PW = Reading pseudowords; RW = Reading real words; RW-Tash = Reading real words with diacritics; RS = Reading syllables; RT = Reading text.

## 3.3 Reliability

Cronbach's $\alpha$ for the seven decoding accuracy scores was $0.914$. Item–total correlations ranged from $0.76$ to $0.84$, indicating that each task contributed meaningfully to the composite (Table 4).

Table 3: Response time (s) per grade and task

(a) Cognitive, naming, and phonological tasks

| Grade | IM | LN | LC | NN | ON | OD | PA | PWM |
|---|---|---|---|---|---|---|---|---|
| Grade 2 | 11.37 ± 5.84 | 4.74 ± 0.38 | 22.99 ± 16.31 | 4.80 ± 0.50 | 9.52 ± 0.67 | 9.13 ± 4.32 | 9.17 ± 10.57 | 9.93 ± 5.20 |
| Grade 3 | 9.79 ± 4.41 | 4.46 ± 0.59 | 22.68 ± 18.25 | 4.40 ± 0.64 | 9.13 ± 1.04 | 7.00 ± 3.46 | 8.39 ± 10.87 | 9.50 ± 9.27 |
| Grade 4 | 9.97 ± 10.38 | 4.50 ± 0.39 | 15.15 ± 14.60 | 4.47 ± 0.47 | 9.16 ± 0.80 | 7.12 ± 2.46 | 14.84 ± 21.84 | 7.00 ± 3.61 |
| Grade 5 | 10.96 ± 9.92 | 4.76 ± 0.26 | 16.01 ± 11.32 | 4.74 ± 0.17 | 9.65 ± 0.35 | 6.30 ± 3.25 | 8.24 ± 10.00 | 6.47 ± 2.62 |
| Grade 6 | 12.66 ± 21.45 | 4.69 ± 0.23 | 17.75 ± 14.82 | 4.65 ± 0.31 | 9.43 ± 0.77 | 6.29 ± 5.86 | 11.33 ± 19.03 | 9.21 ± 9.47 |

(b) (Reading and decoding tasks

| Grade | RC | RL | RL-POS | PW | RW | RW-Tash | RS | RT |
|---|---|---|---|---|---|---|---|---|
| Grade 2 | 158.92 ± 318.27 | 71.58 ± 2.65 | 75.94 ± 19.47 | 71.06 ± 7.76 | 76.93 ± 11.69 | 80.42 ± 14.59 | 64.03 ± 9.74 | 55.53 ± 41.00 |
| Grade 3 | 99.67 ± 116.11 | 72.15 ± 0.97 | 92.20 ± 0.93 | 77.57 ± 1.43 | 86.76 ± 0.70 | 92.55 ± 1.92 | 72.23 ± 1.11 | 44.64 ± 24.02 |
| Grade 4 | 79.51 ± 97.28 | 71.78 ± 1.24 | 91.94 ± 1.11 | 77.33 ± 1.41 | 86.33 ± 1.64 | 92.42 ± 1.36 | 71.92 ± 1.23 | 53.43 ± 36.84 |
| Grade 5 | 128.04 ± 148.06 | 71.84 ± 1.41 | 92.02 ± 0.80 | 77.48 ± 0.95 | 86.50 ± 0.91 | 92.50 ± 0.87 | 72.07 ± 0.72 | 66.70 ± 37.64 |
| Grade 6 | 104.20 ± 110.65 | 71.71 ± 1.25 | 91.90 ± 0.80 | 77.30 ± 2.45 | 86.14 ± 1.73 | 92.31 ± 1.15 | 71.94 ± 0.83 | 51.68 ± 31.00 |

*Note.* Values are mean ± SD in seconds. IM = Immediate short-term memory; LN = Letter naming; LC = Listening comprehension; NN = Number naming; ON = Object naming; OD = Orthographic decision; PA = Phonological awareness; PWM = Phonological working memory; RC = Reading comprehension; RL = Reading letters; RL-POS = Reading letters in beginning, middle, and end positions; PW = Reading pseudowords; RW = Reading real words; RW-Tash = Reading real words with diacritics; RS = Reading syllables; RT = Reading text.

Table 4: Cronbach's $\alpha$ and item–total correlations for reading tasks

| Task | Value |
|---|---|
| **Total reliability** | |
| Cronbach's $\alpha$ | 0.914 |
| **Item–total correlations** | |
| Reading letters | 0.622 |
| Reading letters in beginning, middle, and end positions | 0.666 |
| Reading syllables | 0.788 |
| Reading real words | 0.793 |
| Reading real words with diacritics | 0.840 |
| Reading pseudowords | 0.768 |
| Reading text | 0.672 |

### 3.4 Convergent and discriminant validity

Convergent and discriminant validity assess whether related constructs cohere as expected within a measurement model. Convergent validity is demonstrated when tasks designed to measure related abilities show moderate to strong correlations, indicating that they tap a common underlying construct. Discriminant validity is established when theoretically distinct abilities—such as low-level symbol recognition versus higher-level decoding or comprehension—show weak or absent correlations.

Spearman correlations between SARD decoding tasks and passage comprehension varied across grades. In Grade 2, significant correlations emerged for real words ($\rho = .31$, $p < .01$), diacritised words ($\rho = .23$, $p < .05$), pseudowords ($\rho = .24$, $p < .05$), and text reading ($\rho = .26$, $p < .05$). By Grade 6, pseudoword decoding remained significant ($\rho = .29$, $p = .02$), and passage reading showed the strongest association ($\rho = .46$, $p < .001$). Letter-level correlations were generally weak or non-significant, consistent with ceiling effects.

Partial correlations controlling for letter-naming accuracy reduced letter-level associations to near zero (e.g., Grade 4 positional letters: $\rho_{\text{partial}} = 0.09$, $p = 0.55$), whereas word-level effects remained moderate (e.g., Grade 6 diacritised words:$\rho_{\text{partial}} = 0.30$, $p = 0.02$; pseudowords: $\rho_{\text{partial}} = 0.26$, $p = 0.05$). These results support *convergent validity* for word and pseudoword decoding—both of which align with comprehension performance—and *discriminant validity* for letter naming, which reflects a distinct, more basic visual–phonological mapping skill. Together, these findings indicate that the SARD decoding battery captures the intended construct of reading fluency rather than general processing speed or symbol identification (see Table 5).

Table 5: Convergent validity between decoding tasks and passage comprehension across grades

| Grade | Task | $\rho$ | $p$ |
|---|---|---|---|
| Grade 2 | Reading letters | 0.238 | 0.036 |
| | Reading letters in beginning, middle, and end positions | 0.133 | 0.241 |
| | Reading syllables | 0.003 | 0.981 |
| | Reading real words | 0.310 | 0.005 |
| | Reading real words with diacritics | 0.228 | 0.043 |
| | Reading pseudowords | 0.238 | 0.035 |
| | Reading text | 0.264 | 0.019 |
| Grade 3 | Reading letters | -0.113 | 0.360 |
| | Reading letters in beginning, middle, and end positions | 0.249 | 0.042 |
| | Reading syllables | 0.109 | 0.381 |
| | Reading real words | 0.232 | 0.057 |
| | Reading real words with diacritics | 0.231 | 0.058 |
| | Reading pseudowords | 0.230 | 0.061 |
| | Reading text | 0.155 | 0.209 |
| Grade 4 | Reading letters | -0.038 | 0.798 |
| | Reading letters in beginning, middle, and end positions | 0.010 | 0.948 |
| | Reading syllables | 0.261 | 0.083 |
| | Reading real words | 0.084 | 0.575 |
| | Reading real words with diacritics | 0.042 | 0.781 |
| | Reading pseudowords | 0.056 | 0.710 |
| | Reading text | 0.115 | 0.448 |
| Grade 5 | Reading letters | -0.062 | 0.652 |
| | Reading letters in beginning, middle, and end positions | 0.007 | 0.960 |
| | Reading syllables | -0.098 | 0.458 |
| | Reading real words | 0.051 | 0.696 |
| | Reading real words with diacritics | 0.127 | 0.333 |
| | Reading pseudowords | 0.135 | 0.308 |
| | Reading text | 0.181 | 0.171 |
| Grade 6 | Reading letters | 0.061 | 0.637 |
| | Reading letters in beginning, middle, and end positions | 0.241 | 0.062 |
| | Reading syllables | 0.143 | 0.273 |
| | Reading real words | 0.135 | 0.294 |
| | Reading real words with diacritics | 0.369 | 0.003 |
| | Reading pseudowords | 0.295 | 0.020 |
| | Reading text | 0.464 | 0.000 |

### 3.5 Correlation structure

Inter-task correlation analysis revealed coherent and interpretable patterns across the SARD battery. Accuracy scores were strongly and positively correlated across decoding tasks, indicating that performance on one task tended to predict performance on others. Two main clusters emerged: one comprising letter-level tasks (letter naming and positional letter reading) and another including syllable, real-word, diacritised-word, pseudoword, and text reading. This pattern suggests that while all tasks draw on common decoding skills, letter recognition forms a distinct but related foundational process within the broader reading construct.

Reaction time (RT) correlations showed a similar structure but with predominantly negative coefficients. Because shorter RTs represent faster responses, negative correlations mean that children who were quicker on one task also tended to be quicker on others—indicating a shared fluency or processing efficiency factor. Likewise, negative correlations between RT and accuracy show that faster readers were often more accurate, reflecting the co-development of speed and proficiency. Despite differences in sign, the overall clustering pattern remained the same as for accuracy, distinguishing letter-level from word-level decoding processes. (Figure 1)

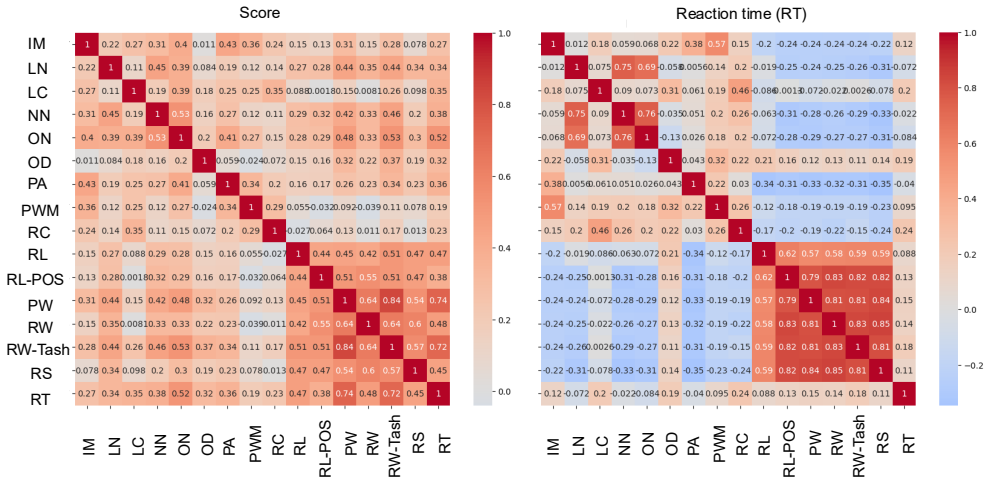


Figure 1: (a) Heatmap of inter-task correlations for accuracy scores across all cognitive, phonological, and decoding tasks. Warmer colors indicate stronger positive correlations, while cooler colors indicate weaker or negative relationships. A clear clustering pattern is observed, distinguishing letter-level tasks (letters and positional letters) from higher-level decoding tasks (syllables, real words, diacritised words, pseudowords, and text). (b) Heatmap of inter-task correlations for mean reaction times (RT). Correlations are generally weaker than for accuracy, reflecting task-specific processing speed patterns. Letter-level tasks again cluster together, while word- and text-level decoding tasks form a separate group, suggesting shared fluency and automaticity demands within each domain. IM = Immediate short-term memory; LN = Letter naming; LC = Listening comprehension; NN = Number naming; ON = Object naming; OD = Orthographic decision; PA = Phonological awareness; PWM = Phonological working memory; RC = Reading comprehension; RL = Reading letters; RL-POS = Reading letters in beginning, middle, and end positions; PW = Reading pseudowords; RW = Reading real words; RW-Tash = Reading real words with diacritics; RS = Reading syllables; RT = Reading text.

### 3.6 Speed–accuracy

Across Grades 2–6, mean log reaction times (RTs) and accuracy scores showed wide variation, reflecting developmental differences in reading efficiency and cognitive control. The LOWESS curve revealed a clear U-shaped trend between speed and accuracy: at lower grades, faster responses were linked to higher accuracy, consistent with gains in basic decoding skill; however, as tasks became more complex and readers engaged in more strategic or integrative processing, accuracy continued to improve even with longer RTs. This indicates a developmental progression from automatic letter and word recognition toward more deliberate, comprehension-oriented reading. Higher-grade students clustered in the upper-right region of the plot, showing slower but more accurate performance, whereas younger students displayed greater variability, with some showing fast but error-prone responses. Overall, the pattern reflects the expected trade-off between speed and accuracy as children transition from surface decoding to fluent, controlled reading. (Figure 2)

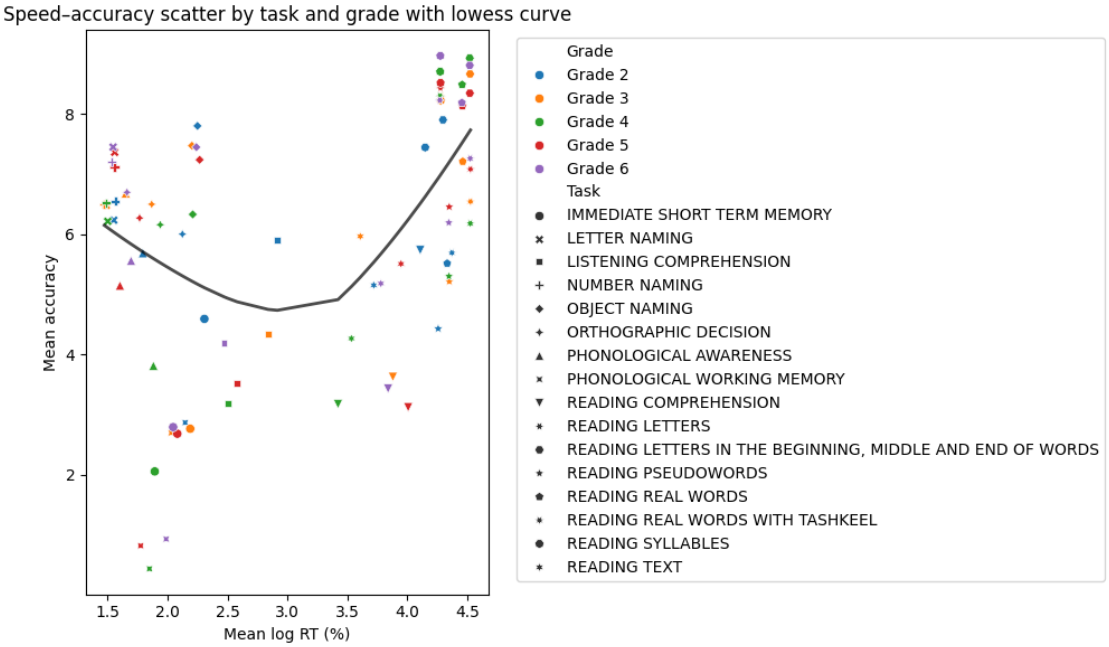


Figure 2: Speed–accuracy relationship in Arabic reading screening tasks across Grades 2–6. Each point represents the mean accuracy and mean log reaction time (RT) for a specific task–grade combination. Colors indicate grade levels and marker shapes represent different tasks. The locally weighted scatterplot smoothing (LOWESS) curve shows a U-shaped pattern, with faster responses initially associated with higher accuracy, followed by a plateau and a rise in RTs at high accuracy levels. This pattern reflects the developmental shift from early, effortful decoding—where rapid responding may reduce accuracy—to more efficient, controlled reading, where slower but more accurate responses emerge as fluency consolidates.

### 3.7 PCA and clustering

Principal component analysis (PCA) of the seven decoding scores revealed a dominant first component (PC1) accounting for 66.7% of the total variance, with all tasks loading positively (0.33–0.41), reflecting a general reading proficiency factor. The second component (13.5% of the variance) differentiated letter-level tasks (positive loadings) from word- and text-level tasks (negative loadings). Hierarchical clustering using Ward's method on $1 - |\rho|$ yielded a dendrogram with a clear bifurcation between letter-level and word-level tasks (Figure 3).

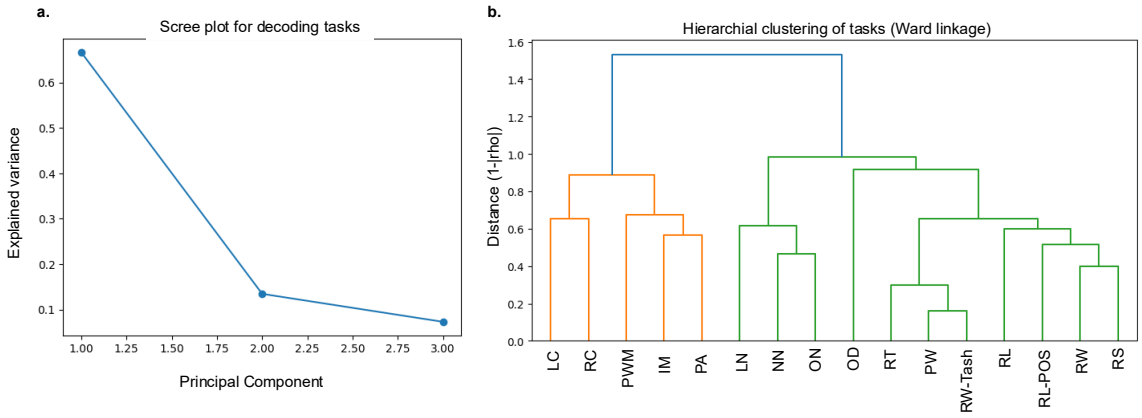


Figure 3: (a) Scree plot showing the proportion of variance explained by each principal component (PC) derived from the seven decoding accuracy scores. The first component (PC1) accounts for 66.7% of the total variance, with all tasks loading positively (0.33–0.41), reflecting a general decoding proficiency factor. (b) Hierarchical clustering of decoding tasks using Ward's linkage on $1 - |\rho|$. The resulting dendrogram reveals two main clusters, distinguishing letter-level tasks (letters and positional letters) from word- and text-level decoding tasks, consistent with the PCA structure. IM = Immediate short-term memory; LN = Letter naming; LC = Listening comprehension; NN = Number naming; ON = Object naming; OD = Orthographic decision; PA = Phonological awareness; PWM = Phonological working memory; RC = Reading comprehension; RL = Reading letters; RL-POS = Reading letters in beginning, middle, and end positions; PW = Reading pseudowords; RW = Reading real words; RW-Tash = Reading real words with diacritics; RS = Reading syllables; RT = Reading text.

## 4 Discussion

This pilot study provides the first evaluation of SARD battery, a web-based tool designed to assess Arabic phonics skills in schools, and investigates its reliability in terms of internal consistency and its multidimensional validity. The tool showed excellent internal consistency. In particular, the decoding tasks exhibited excellent internal consistency ($\alpha = .91$) and strong item–total correlations, supporting the interpretation of the battery as a coherent measure of decoding skill.

### 4.1 Internal consistency and structure

Dimensionality analyses further confirmed that SARD primarily measures a single latent dimension: the first principal component explained approximately two thirds of the variance and loaded positively on all tasks, while the second component differentiated letter-level from word-level tasks. Hierarchical clustering produced a similar bifurcation, indicating that the battery captures both foundational symbol recognition and more integrative lexical processing. Together, these analyses demonstrate that SARD provides internally consistent measures of Arabic decoding, captures expected developmental improvements in both accuracy and speed, and correlates appropriately with reading comprehension. However, as no comparison was made with an established diagnostic reading assessment in Arabic, these findings should be considered preliminary.

### 4.2 Developmental Trends

The present findings provide clear evidence for differentiated developmental trajectories across levels of reading processing between Grades 2 and 6. While letter recognition and positional letter identification were already near ceiling by Grade 2, more complex orthographic and phonological decoding tasks continued to improve steadily through Grade 6. This pattern underscores that the fundamental visual–orthographic mapping skills underlying letter recognition are largely consolidated early in literacy acquisition, whereas word- and text-level decoding and fluency continue to develop well into later primary school years.

### 4.3 Growth in Word- and Text-Level Accuracy

The gradual increase in accuracy for syllables, real words, diacritised words, and pseudowords (rising from approximately 70% to over 90%) suggests ongoing refinement in orthographic–phonological integration and lexical access. The significant positive slopes (2.3–3.7 percentage points per grade) indicate that these improvements are both robust and sustained across grades. These results are consistent with stage-based models of reading development (e.g., Ehri 2005, Share 1995), which posit that automaticity in word recognition and efficient decoding emerge only after children accumulate sufficient exposure to print and strengthen sublexical–lexical mappings. The finding that pseudoword accuracy improved at a comparable rate to real-word accuracy is particularly noteworthy, as it reflects enhanced phonological decoding skill rather than mere lexical memorization. Similarly, the gains in reading diacritised words—especially relevant in orthographies where vowel markers carry key phonological distinctions—indicate more precise grapheme–phoneme conversion and increased sensitivity to orthographic detail.

### 4.4 Fluency Gains and Response Time Reductions

Across all tasks, response times declined by approximately 0.3–0.4 log units between Grades 2 and 5, corresponding to roughly 30% faster performance. These reductions, coupled with the accuracy improvements, reflect growing automaticity and fluency in both lower- and higher-level reading processes. The plateau observed by Grade 6 suggests that, by this stage, children's performance on many sublexical and lexical tasks approaches asymptotic levels of efficiency. Similar nonlinear growth patterns have been reported in prior longitudinal studies (e.g., Kim and Wagner 2015, Seymour et al. 2003), where early rapid gains are followed by slower refinements once decoding becomes automatized.

### 4.5 Reading Comprehension and Connected Text Processing

Passage-level reading showed the most substantial absolute improvement, increasing from 56% accuracy in Grade 2 to 87% by Grade 6. This growth, combined with reductions in response time, suggests parallel development in decoding, fluency, and comprehension skills. As comprehension relies not only on word recognition but also on vocabulary knowledge, working memory, and inferential reasoning, the observed progress likely reflects broader cognitive–linguistic maturation. The modest variability in response time at the passage level also points to individual differences in comprehension monitoring and text integration skills that persist even at later grades.

### 4.6 Stability in Basic Skills

By contrast, tasks assessing letter identification and letter-position recognition reached ceiling early and showed no significant change after Grade 3. This plateau confirms that alphabetic knowledge stabilizes relatively early in literacy development, serving as a foundation upon which higher-order reading and language skills are built. The lack of further improvement does not indicate stagnation but rather successful consolidation of foundational orthographic representations.

### 4.7 Correlation Structure and Reading Subsystems

The correlation analyses revealed a coherent internal structure within the SARD battery, supporting its construct validity and alignment with established models of reading development. Strong positive correlations among decoding accuracy measures indicate that performance on one reading task reliably predicts performance on others, consistent with the notion of a shared core decoding competence [Perfetti et al., 1992, Share, 1995]. However, the emergence of two distinct clusters—one centered on letter-level tasks (letter naming and positional letter reading) and another encompassing syllable, real-word, diacritised-word, pseudoword, and text reading—suggests a hierarchical organization of reading subskills.This bifurcation aligns with prior work demonstrating that letter knowledge and letter-position encoding represent foundational processes that are mastered early and serve as precursors for word- and text-level reading [Ehri,

2005, Seymour et al., 2003]. In the context of Arabic, this distinction is particularly salient: letter and positional recognition require sensitivity to complex visual–orthographic patterns and contextual letter forms [Abu-Rabia, 1997, Taha and Saiegh-Haddad, 2017b]. As decoding tasks progress from isolated letters to words and connected text, performance reflects increasing integration of phonological, orthographic, and morphological knowledge, consistent with multi-level models of literacy acquisition [Ziegler and Goswami, 2005].The strong within-cluster correlations among syllable, pseudoword, and real-word tasks also suggest that these measures tap overlapping phonological decoding mechanisms, whereas diacritised-word and text reading add layers of lexical and semantic processing. This mirrors evidence from cross-linguistic studies indicating that pseudoword reading shares the strongest association with phonological awareness, while real-word and text reading reflect both phonological and lexical fluency [Frost, 2012, Kim and Wagner, 2015].

### 4.8 Reaction Time Correlations and Fluency Factors

Reaction time (RT) correlations revealed a parallel structure to that observed for accuracy, though with negative coefficients due to the inverse relationship between RT and processing efficiency. Children who responded quickly on one task tended to be fast across others, indicating a shared fluency or processing-speed factor underlying decoding [Wolf and Katzir-Cohen, 2001]. This shared variance supports models that conceptualize fluency as a domain-general skill emerging from automatization of lower-level decoding operations [LaBerge and Samuels, 1974]. Moreover, negative correlations between RT and accuracy confirm that faster readers were typically more accurate—reflecting co-development of speed and proficiency during literacy acquisition [Fuchs et al., 2001]. However, the clustering pattern also showed that letter-level fluency was separable from word- and text-level fluency, suggesting distinct processing demands. Letter naming speed, for instance, is often considered an index of rapid automatized naming (RAN), which predicts later reading fluency and efficiency [Norton and Wolf, 2012]. The current findings therefore reinforce the theoretical distinction between automatic visual identification of symbols and more integrative decoding at the lexical level.

### 4.9 Developmental Dynamics of the Speed–Accuracy Relationship

The U-shaped pattern observed in the speed–accuracy analysis provides further insight into developmental shifts in reading strategy. Among younger readers, faster responses were associated with higher accuracy, indicating that improved decoding efficiency accompanies gains in automaticity at early stages. As tasks became more complex, however, accuracy continued to improve even with longer RTs, reflecting a transition from surface-level decoding to more deliberate, meaning-oriented processing.This developmental shift echoes findings from previous studies showing that as readers mature, they engage more controlled strategies involving morphological and semantic analysis, particularly in morphologically rich languages like Arabic [Ibrahim et al., 2002]. The clustering of higher-grade students in the "slower but more accurate" region of the speed–accuracy space suggests that mature readers balance fluency with comprehension monitoring and orthographic precision—consistent with interactive models of reading that emphasize coordination of decoding and comprehension [Kintsch and Rawson, 2005].

### 4.10 Limitations

The study have several limitations. The cross sectional design precludes assessment of test–retest reliability and long term predictive validity. The sample, although representative, it was limited to native speakers in schools in UAE; regional and dialectal differences across the Arab world may require separate norming studies. Although we examined associations with passage comprehension, we did not compute sensitivity and specificity against an external diagnostic benchmark. Moreover, convergent validity was assessed using internal reading comprehension task.

### 4.11 Future Directions

Future research should establish criterion validity to confirm SARD's screening accuracy, sensitivity, specificity, and receiver operating characteristic (ROC) performance. Although the present findings demonstrated strong internal consistency and interpretable inter-task correlations—indicating that SARD reliably captures the latent structure of Arabic reading subskills—criterion validity is essential to determine whether these psychometric strengths translate into diagnostic precision. Establishing sensitivity and specificity relative to independent benchmarks (e.g., standardized literacy assessments or teacher-identified reading difficulties) would clarify SARD's utility as an early-identification tool for decoding and fluency deficits.

Given that several SARD tasks showed ceiling effects at the letter level but continuous growth at the syllable-, word-, and text-levels, criterion validation should focus particularly on those mid- to higher-order tasks (pseudoword, real-word,

diacritised-word, and passage reading). These tasks demonstrated robust developmental gradients and strong cross-task correlations—properties that make them promising discriminators between typical and at-risk readers. ROC analyses could quantify the optimal cut-off points for each subtest, indicating how well SARD distinguishes children with emerging fluency problems from those following typical developmental trajectories.

Furthermore, longitudinal validation would allow examination of SARD's predictive validity: whether Grade 2–3 performance reliably forecasts reading comprehension and fluency outcomes in later grades. Incorporating measures such as the area under the ROC curve (AUC), positive predictive value (PPV), and negative predictive value (NPV) would provide a comprehensive profile of SARD's classification performance across developmental stages.

Finally, given the distinctive features of Arabic orthography—morphological richness, diglossia, and optional diacritics—future work should compare ROC and sensitivity–specificity parameters across dialectal varieties and orthographic presentations. Doing so would establish the generalizability of SARD's diagnostic thresholds and its robustness to linguistic variation. Integrating machine-learning approaches (e.g., logistic regression or random-forest models) with large-scale normative data could further refine SARD's predictive accuracy, enabling adaptive, data-driven screening of Arabic literacy skills in educational and clinical settings.

## 5 Conclusion

We conclude that SARD is a reliable and valid tool for screening Arabic reading difficulties in primary schools. SARD demonstrated excellent internal consistency, captured expected developmental improvements in decoding accuracy and fluency, and showed convergent and divergent validity. Principal component and cluster analyses further confirmed that SARD primarily measures a single latent reading factor while distinguishing letter level from word level skills. Its digital format enables rapid administration and scoring, making it well suited for classroom use and progress monitoring.

## References

Ina V. S. Mullis, Matthias von Davier, Pierre Foy, Beth Fishbein, Kathryn A. Reynolds, and Emma Wry. *PIRLS 2021 International Results in Reading*. TIMSS & PIRLS International Study Center, Boston College, 2023. doi:10.6017/lse.tpisc.tr2103.kb5342. URL https://doi.org/10.6017/lse.tpisc.tr2103.kb5342.

Asaid Khateb, Haitham Y. Taha, Inbal Elias, and Raphiq Ibrahim. The effect of the internal orthographic connectivity of written Arabic words on the process of visual recognition: A comparison between skilled and dyslexic readers. *Writing Systems Research*, 5(2):214–233, 2013. doi:10.1080/17586801.2013.834244. URL https://doi.org/10.1080/17586801.2013.834244.

Haitham Taha. The development of reading and spelling in Arabic orthography: Two parallel processes? *Reading Psychology*, 37(8):1149–1161, 2016. doi:10.1080/02702711.2016.1193580. URL https://doi.org/10.1080/02702711.2016.1193580.

Haitham Taha and Hala Khateeb. Statistical learning and orthographic preferences among native Arab kindergarten and first graders. *Writing Systems Research*, 10(1):15–25, 2018. doi:10.1080/17586801.2018.1473313. URL https://doi.org/10.1080/17586801.2018.1473313.

Salim Abu-Rabia. The effect of Arabic vowels on the reading comprehension of second- and sixth-grade native Arab children. *Journal of Psycholinguistic Research*, 28(1):93–101, 1999. doi:10.1023/A:1023291620997. URL https://doi.org/10.1023/A:1023291620997.

Haitham Taha and Elinor Saiegh-Haddad. Morphology and spelling in arabic: Development and interface. *Journal of Psycholinguistic Research*, 46(1):27–38, 2017a. doi:10.1007/s10936-016-9425-3. URL https://doi.org/10.1007/s10936-016-9425-3.

Charles A. Ferguson. Diglossia. *WORD*, 15(2):325–340, 1959. doi:10.1080/00437956.1959.11659702. URL https://doi.org/10.1080/00437956.1959.11659702.

Salah Ayari. Diglossia and illiteracy in the Arab world. *Language, Culture and Curriculum*, 9:243–253, 1996. doi:10.1080/07908319609525233. URL https://doi.org/10.1080/07908319609525233.

Rachel Schiff and Elinor Saiegh-Haddad. Development and relationships between phonological awareness, morphological awareness, and word reading in spoken and standard Arabic. *Frontiers in Psychology*, 9:356, 2018. doi:10.3389/fpsyg.2018.00356. URL https://doi.org/10.3389/fpsyg.2018.00356.

Souad El Idrissi Rioui, Mohamed Saad Rigar, and Abderrazak Grine. The impact of mandatory IFRS adoption on earnings management: Evidence from Morocco—a multinomial logit approach. In *Journal of Physics: Conference*

*Series*, volume 1743, page 012013, 2021. doi:10.1088/1742-6596/1743/1/012013. URL https://doi.org/10.1088/1742-6596/1743/1/012013.

Reem Khamis-Dakwar, Karen Froud, and Peter Gordon. Acquiring diglossia: Mutual influences of formal and colloquial Arabic on children's grammaticality judgments. *Journal of Child Language*, 39(1):61–89, 2012. doi:10.1017/S0305000910000784. URL https://doi.org/10.1017/S0305000910000784.

Linnea C. Ehri. Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2):167–188, 2005. doi:10.1207/s1532799xssr0902_4.

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

Young-Suk Grace Kim and Richard K. Wagner. Text (oral) reading fluency as a construct in reading development. *Scientific Studies of Reading*, 19(3):224–242, 2015. doi:10.1080/10888438.2015.1007375. URL https://doi.org/10.1080/10888438.2015.1007375.

Philip H. K. Seymour, Mikko Aro, and Jane M. Erskine. Foundation literacy acquisition in european orthographies. *British Journal of Psychology*, 94:143–174, 2003. doi:10.1348/000712603321661859. URL https://doi.org/10.1348/000712603321661859.

Charles A. Perfetti, Sulan Zhang, and Iris Berent. Reading in english and chinese: Evidence for a universal phonological principle. *Advances in Psychology*, 94:227–248, 1992. doi:10.1016/S0166-4115(08)62798-3. URL https://doi.org/10.1016/S0166-4115(08)62798-3.

Salim Abu-Rabia. Reading in arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native arabic readers. *Reading and Writing*, 9(1):65–78, 1997. doi:10.1023/A:1007962408827. URL https://doi.org/10.1023/A:1007962408827.

Haitham Taha and Elinor Saiegh-Haddad. Morphology and spelling in arabic: Development and interface. *Journal of Psycholinguistic Research*, 46(1):27–38, 2017b. doi:10.1007/s10936-016-9425-3. URL https://doi.org/10.1007/s10936-016-9425-3.

Johannes C. Ziegler and Usha Goswami. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131(1):3–29, 2005. doi:10.1037/0033-2909.131.1.3. URL https://doi.org/10.1037/0033-2909.131.1.3.

Ram Frost. Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5):263–279, 2012. doi:10.1017/S0140525X11001841. URL https://doi.org/10.1017/S0140525X11001841.

Maryanne Wolf and Tami Katzir-Cohen. Reading fluency and its intervention. *Scientific Studies of Reading*, 5(3):211–239, 2001. doi:10.1207/S1532799XSSR0503_2. URL https://doi.org/10.1207/S1532799XSSR0503_2.

David LaBerge and S. Jay Samuels. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6(2):293–323, 1974. doi:10.1016/0010-0285(74)90015-2. URL https://doi.org/10.1016/0010-0285(74)90015-2.

Lynn S. Fuchs, Douglas Fuchs, Michelle K. Hosp, and Jim Jenkins. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3):239–256, 2001. doi:10.1207/S1532799XSSR0503_3. URL https://doi.org/10.1207/S1532799XSSR0503_3.

Emily S. Norton and Maryanne Wolf. Rapid automatized naming (ran) and reading fluency: implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, 63:427–452, 2012. doi:10.1146/annurev-psych-120710-100431. URL https://doi.org/10.1146/annurev-psych-120710-100431.

Raphiq Ibrahim, Zohar Eviatar, and Judith Aharon-Peretz. The characteristics of arabic orthography slow the development of visual word recognition. *Neuropsychology*, 16(3):322–326, 2002. doi:10.1037/0894-4105.16.3.322. URL https://doi.org/10.1037/0894-4105.16.3.322.

Walter Kintsch and Katherine A. Rawson. Comprehension. In Margaret J. Snowling and Charles Hulme, editors, *The Science of Reading: A Handbook*, pages 211–226. Blackwell Publishing, 2005. doi:10.1002/9780470757642.ch12. URL https://doi.org/10.1002/9780470757642.ch12.