Start shallow and grow deep: The development of a Hebrew reading brain

Upasana Nathaniel\textsuperscript{a,b,*}, Yael Weiss\textsuperscript{c}, Bechor Barouch\textsuperscript{a}, Tami Katzir\textsuperscript{d}, Tali Bitan\textsuperscript{a,b,e}

\textsuperscript{a} Psychology Department and Institute for Information Processing and Decision Making, University of Haifa, Israel
\textsuperscript{b} Integrated Brain and Behavior Center (IBBRC), University of Haifa, Israel
\textsuperscript{c} Institute for Learning and Brain Sciences, University of Washington, Seattle, WA, USA
\textsuperscript{d} Department of Learning Disabilities, The E.J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel
\textsuperscript{e} Department of Speech Language Pathology, University of Toronto, Toronto, Canada

\textbf{ABSTRACT}

Brain plasticity implies that readers of different orthographies can have different reading networks. Theoretical models suggest that reading acquisition in transparent orthographies relies on mapping smaller orthographic units to phonology, than reading opaque orthographies; but what are the neural mechanisms underlying this difference? Hebrew has a transparent (pointed) script used for beginners, and a non-transparent script used for skilled readers. The current study examined the developmental changes in brain regions associated with phonological and orthographic processes during reading pointed and un-pointed words. Our results highlight some changes that are universal in reading development, such as a developmental increase in frontal involvement (in bilateral inferior frontal gyrus (IFG) pars opercularis), and increase in left asymmetry (in IFG pars opercularis and superior temporal gyrus, STG) of the reading network. Our results also showed a developmental increase in activation in STG, which stands in contrast to previous studies in other orthographies. We further found an interaction of word length and diacritics in bilateral STG and the visual word form area (VWFA) across both groups. These findings suggest that children slightly adjust their reading depending on orthographic transparency, relying on smaller units when reading a transparent script and on larger units when reading an opaque script. Our results also showed that phonological abilities across groups correlated with activation in the VWFA, regardless of transparency, supporting the continued role of phonology at all levels of orthographic transparency. Our findings are consistent with multiple route reading models, in which both phonological and orthographic processing of multiple size units continue to play a role in children’s reading of transparent and opaque scripts during reading development. The results further demonstrate the importance of taking into account differences between orthographies when constructing neural models of reading acquisition.

\textbf{1. Introduction}

During reading acquisition children learn to map visual and orthographic representations to phonological and semantic ones. However, this process may differ across languages, depending on properties of the orthography such as the consistency with which phonology is represented in the writing system (Frost et al., 1987; Ziegler et al., 2001; Ziegler and Goswami, 2005). The consistency of letter-to-sound representation is high in transparent or shallow orthographies, such as Italian or Greek, providing the necessary articulatory cues for pronunciation. Opaque or deep orthographies on the other hand, such as English or French, have a complex letter-to-sound representation, thus naming may be facilitated by existing lexical representations or stored pronunciations of a word e.g., “plough” (Coltheart, 1978; Landerl et al., 1997). Orthographic transparency was shown to affect the neural reading network in skilled adult readers (Paulesi et al., 2000; Rueckl et al., 2015; Weiss et al., 2016) and to be a key factor in influencing the rate of reading acquisition across different languages (Ellis and Hooper, 2001; Holopainen et al., 2001; Jorm et al., 1984; Katzir et al., 2008; Seymour et al., 2003; Treiman et al., 1990; Vellutino et al., 2004; Ziegler and Goswami, 2005).

In Hebrew, a language with dual versions of script, children’s instruction in school undergoes a shift from reading a transparent (pointed) script in early stages, to reading an un-pointed script later on,
which may affect the interplay of phonological and orthographic processes during reading development. Examining maturational changes that a developing brain undergoes in the early stages of reading acquisition in Hebrew, can provide important insights into how skilled reading emerges across the two versions of script and how orthographic transparency influences the subsequent functional specialisation of neural regions. The goal of the current study was therefore to examine the neurodevelopmental processes associated with reading acquisition in young Hebrew speakers, and examine how regions engaged in phonological and orthographic processing are affected by different levels of orthographic transparency and how they change during development.

1.1. Models of reading acquisition

The cognitive processes underlying reading and reading acquisition have been under debate for several decades. Dual route models, which have been developed for the English orthography, with its abundance of irregularly spelled words, posit a direct and indirect route from orthography to semantics. Based on these models phonological mediation in the indirect route occurs by sequential translation of letters to sounds (Barron, 1986; Coltheart et al., 2001). These models predict that during reading acquisition children shift from reliance on sequential phonological decoding in the indirect route to reliance on the direct route from orthography to semantics, thus reducing the reliance on phonological processes (Coltheart et al., 2001; Perry et al., 2007). In contrast, connectionist and multiple route models posit that phonological processes are always inherent to visual word recognition (Seidenberg et al., 1994), and they do not decay during development (Grainier et al., 2012; Milledge and Blythe, 2019). Rather phonological processing may change from overt, sequential decoding to covert and parallel mapping of orthographic, to phonological and semantic units (Grainier et al., 2012; Milledge and Blythe, 2019).

While reading acquisition literature is dominated by models based on the English orthography (Share, 2021) a number of theories were suggested for differences between orthographies (Frost, 2005; Share, 2008; Ziegler and Goswami, 2005). Both dual route (Frost et al., 1987; Katz and Frost, 1992) and multiple route models (Ziegler and Goswami, 2005) predict that orthographic transparency affects the mapping of orthography to phonology during reading. According to the psycholinguistic grain size theory (Ziegler and Goswami, 2005), children learning to read transparent orthographies rely initially on smaller grain size phonological units, and gradually move to mapping of larger lexical units, while young readers of opaque orthographies show greater reliance on whole-world recognition or mapping of larger orthographic units.

Most studies examining the effect of orthographic transparency on reading acquisition have focused on pairwise comparisons of English as an opaque orthography, and a more transparent orthography. Such cross-linguistic studies have shown that accurate word recognition in transparent orthographies appear close to ceiling just after a year of reading instruction compared to opaque orthographies such as French, Danish, and particularly English (Ellis and Hooper, 2001; Seymour et al., 2003). Such differences in early word recognition and pseudoword reading accuracies have consistently been reported in comparisons between English and Spanish (Goswami et al., 1998), English and Italian (Cosru et al., 1988), and English and Welsh (Spencer and Richard Hanley, 2003). However, it is still not known what are the neural mechanisms underlying these differences in accuracy. While the reading pathways of children acquiring a transparent or opaque orthography may differ from those of children acquiring reading in a dual version script, Hebrew provides the advantage of studying the effect of transparency in a within-language and within-subject design, overcoming the limitations of cross-cultural comparisons.

1.2. Reading acquisition in Hebrew

The Hebrew script has two versions of orthography, a fully transparent or vowelised version with diacritics (pointed script), and an opaque version (un-pointed script) with partial or no vowel representations. Children at first grade learn to read using the pointed script and gradually transition to reading without diacritics between 2nd and 3rd grades, becoming skilled at reading the un-pointed script by 5th grade (Ravid, 1996; Shany et al., 2012). Diacritics have therefore been found to facilitate word recognition in early stages of reading acquisition in Hebrew (Navon and Shimon, 1981; Ravid, 1996; Shany and Share, 2011). Children learning to read the pointed script quickly master decoding processes (Shattil et al., 2000) as diacritics enhance phonological processing and disambiguate homographs (Shimon, 1999). The benefit of diacritics is especially pronounced in oral reading and in low frequency words (Koriat, 1984). Children throughout elementary school have also demonstrated faster recognition of short compared to long words in pointed Hebrew (Schiff, 2003), consistent with the notion that reading a transparent orthography relies on conversion of smaller phonological units (Ge Luca et al., 2008; Ellis and Hooper, 2001; Havellka et al., 2010). However, the contribution of diacritics to accurate reading of words and text decreases in older, more skilled readers (Bar-Kochva and Breesnitz, 2014; Bar-On et al., 2017; Ravid, 1996; Schiff et al., 2013; Shany et al., 2012; Shimon and Navon, 1982), and can either have facilitatory effects (Navon and Shimon, 1981; Shimon and Navon, 1982), or no effect (Bentin and Frost, 1987; Schiff and Ravid, 2004; Shimon and Sivan, 1994) on word recognition in adulthood. Nevertheless, studies manipulating word length show that even without a clear benefit, adult skilled readers still process words with diacritics through more piecemeal decoding, compared to words with no diacritics (Weiss et al., 2015a, 2015b).

The letters used in the Hebrew orthography mostly represent consonants, however, four letters also represent vowels. The dual function of vowel letters could theoretically interfere with reading (Shimon, 1999), but studies have shown that the presence of vowel letters facilitated reading in second, fourth and sixth graders (Schiff, 2003). Word length studies have revealed that this facilitatory effect of adding a vowel letter was in contrast to the addition of a consonant letter (Schiff, 2003), suggesting that vowel letters provide essential phonological cues for disambiguating potential homographs (Harel, 2005). One notable model discussing these developmental phases is the triplex model of Hebrew reading acquisition (Share and Bar-On, 2018). According to the model, reading acquisition begins with the mapping of letters to sounds in Grade 1 (sub-lexical phase), which builds phonological awareness skills. By Grade 2 children are well versed in reading the pointed script, being less dependent on vowel letters and more dependent on orthographic representations (lexical phase). At the final stage, in upper elementary grades, they transition to reading the un-pointed script (supra-lexical phase), which requires greater reliance on higher-level contexts to solve ambiguities which are very common in the un-pointed script. While much has been learned about the effects of diacritics in reading Hebrew, there has been no direct comparison of the effects of vowel letters and diacritics in beginner and more advanced readers, and the neural basis for reading with these different representations, which the current study aimed to further investigate.

1.3. Neuroimaging studies of reading and reading acquisition

A meta-analysis of studies examining word reading across languages has shown the involvement of three regions in the reading network across all orthographies: the left inferior frontal gyrus (IFG), left superior temporal gyrus (STG), and left occipitotemporal cortex (Bolger et al., 2005). These regions are the focus of interest in the current study as they are relevant to phonological and orthographic processing. The left STG has been associated with access to phonological representations (Leonard and Chang, 2014; Price, 2012) and showed a developmental
The primary goal of the current study was to examine how orthographic transparency affects the development of reading in young Hebrew readers. We used both behavioural (experiment 1) and fMRI measures (experiment 2) to examine the effects of different levels of orthographic transparency on phonological and orthographic processing in children, and how those change during development. This is particularly interesting given the shift in Hebrew reading instruction from the exclusive exposure to the pointed script in early elementary school to the unpointed script in later years. Orthographic transparency was examined by comparing reading pointed and unpointed words, as well as comparing words with and without a single vowel letter. We also manipulated the number of consonants (i.e., 3- vs. 4-consonants), as previous behavioural studies have shown word length effects, i.e., longer responses to long compared to short words, to indicate serial decoding by smaller units (De Luca et al., 2008; Ellis and Hooper, 2001; Hawelka et al., 2010). Our behavioural study with adults showed this effect particularly in the presence of diacritics (Weiss et al., 2015b).

Here we used ROI analyses focusing on three bilateral regions associated with phonological and orthographic aspects of reading: the left IFG pars opercularis, the left STG, and the VWFA. We also included the right hemisphere homologues of these regions to account for the possibility of bilateral cortical involvement in language processing in children (Centanni et al., 2018; Claesens et al., 2007; Everts et al., 2009; Holland et al., 2001; Olulade et al., 2020; Ressel et al., 2008; Szaflarski et al., 2006; Turkeltaub et al., 2003). Below we specify the predictions based on several reading acquisition models, as well as previous neuroimaging findings in adult Hebrew readers.

### 1.4. Current study

This is the first fMRI study to examine reading acquisition in Hebrew. The primary goal of the current study was to examine how orthographic transparency affects the development of reading in young Hebrew readers. We used both behavioural (experiment 1) and fMRI measures (experiment 2) to examine the effects of different levels of orthographic transparency on phonological and orthographic processing in children, and how those change during development. This is particularly interesting given the shift in Hebrew reading instruction from the exclusive exposure to the pointed script in early elementary school to the unpointed script in later years. Orthographic transparency was examined by comparing reading pointed and unpointed words, as well as comparing words with and without a single vowel letter. We also manipulated the number of consonants (i.e., 3- vs. 4-consonants), as previous behavioural studies have shown word length effects, i.e., longer responses to long compared to short words, to indicate serial decoding by smaller units (De Luca et al., 2008; Ellis and Hooper, 2001; Hawelka et al., 2010). Our behavioural study with adults showed this effect particularly in the presence of diacritics (Weiss et al., 2015b).

Here we used ROI analyses focusing on three bilateral regions associated with phonological and orthographic aspects of reading: the left IFG pars opercularis, the left STG, and the VWFA. We also included the right hemisphere homologues of these regions to account for the possibility of bilateral cortical involvement in language processing in children (Centanni et al., 2018; Claesens et al., 2007; Everts et al., 2009; Holland et al., 2001; Olulade et al., 2020; Ressel et al., 2008; Szaflarski et al., 2006; Turkeltaub et al., 2003). Below we specify the predictions based on several reading acquisition models, as well as previous neuroimaging findings in adult Hebrew readers.

### 1. Developmental changes: Based on dual route models (Coltheart et al., 2001; Perry et al., 2007) reading in the younger children group is predicted to rely on serial phonological decoding, manifested in slower responses to longer words (standard word length effect). Older children, in contrast, would show less reliance on serial decoding and an overall decrease in reliance on phonological representations and sublexical phonological segmentation. This would manifest as an age-related decrease in activation in left STG and left IFG pars opercularis for all words. In contrast, while multiple route models (Grainger et al., 2012; Milledge and Brythe, 2019; Ziegler and Goswami, 2005) would have a similar prediction for the young children, they would not predict an age-related decrease in phonological process. Thus, based on these models, older children would not show serial decoding of letters, but they would show no reduction in activation in left STG and left IFG pars opercularis related to sublexical and lexical phonology.

### 2. Effect of Diacritics: Based on both dual route (Perry et al., 2007) and multiple route (Ziegler and Goswami, 2005) models we expect to find differences between reading of pointed and unpointed words, with greater reliance on serial decoding and...
sub-lexical phonological segmentation in reading pointed than un-pointed words; and greater reliance on larger orthographic units in reading un-pointed words. This is expected to be especially true for older children, because young children are expected to rely on serial small-unit decoding for all words. Thus, behaviourally we expect that older children would show an interaction of word length and diacritics, with pointed words showing longer reading time for long words, and the reversed for un-pointed words (short words without diacritics read slower), as seen in adults (Weiss et al., 2015b). In terms of brain activation, we similarly predict greater reliance on phonological segmentation in left IFG pars opercularis and greater activation of phonological representations in left STG, in the reading of pointed words compared to un-pointed words, and greater activation in the VWFA during reading of un-pointed words.

(3) Effects of vowel letters: Vowel letters are expected to facilitate access to phonological representations especially in un-pointed words because they add the missing phonological information. They are also expected to facilitate access to orthographic representations, because the additional letter reduces orthographic competition. Thus, across both age groups we expect to find higher accuracy and shorter RT for words with vowel letters, especially in un-pointed words, as found in adults (Weiss et al., 2015b). We also expect to see decreased activation in STG and VWFA in the reading of words with vowel letters as also seen in adults (Weiss et al., 2015a).

2. Methods

2.1. Experiment #1

2.1.1. Participants

Twenty-eight 2nd grade (ages 7.01 to 8.04, 16 females) and twenty-nine 5th grade students (ages 10.01 to 11.04, 17 females), were recruited from an elementary school in north Israel. Written informed consent was obtained from the parents of all participants and oral consent from the children. The study was approved by the ethics committee of the Faculty of Social Welfare and Health Sciences at the University of Haifa, and by the Ministry of Education. All participants were native Hebrew speakers with no learning disabilities as reported by their teachers and confirmed by our assessments. Their reading level was assessed using the ‘Reading words’ and ‘Reading pseudo-words’ tests from “Alef-Atf, Diagnostic test battery for written language disorders” (Shany et al., 2006), described below. The exclusion criterion was having a score lower than one standard deviation below the mean in both tests. No student was excluded based on this criterion. One 2nd grade participant was excluded from the group analysis because their performance on the experimental task was lower than 3 standard deviations below the group average, in both accuracy and reaction time. This resulted in twenty-seven participants in 2nd grade and twenty-nine participants in 5th grade who were included in the analysis.

2.1.2. Standardized tests

All participants underwent two standardized screening tests, in order to assess their reading and decoding abilities. Screening tests were taken from the “Alef-Atf” battery (Shany et al., 2006): (1) Reading words: participants read aloud 38 nouns with diacritics, which represented different levels of frequency, length, and phonological structure. Different age-appropriate lists were used for the different age groups. The scores indicate the number of accurately read words per minute and the percentage of errors. (2) Reading pseudo-words: participants read aloud 33 pseudo-words with diacritics. 24 of these items represented familiar morpho-phonological structures in Hebrew and nine contained sound structures non-existent in Hebrew. Different age-appropriate lists were used for the different age groups. The obtained scores indicate the number of accurately read pseudowords per minute.

2.1.3. Experimental stimuli

Stimuli was identical to Weiss et al. (Weiss et al., 2015a, 2015b, 2016). 192 concrete Hebrew nouns were used as stimuli, categorized into eight lists, with 24 words in each list: words presented in transparent or non-transparent scripts (with or without diacritics), differed in word length (3- or 4-consonants) and with or without a vowel letter (see Table 1). All words were presented in their typical written form and vowel letters were not removed or inserted into these forms. All words were bi-syllabic, mono-morphemic and were matched for frequency across conditions, both in means and distribution. As there was no available consensus corpus for written Hebrew frequency at the time of data collection in 2012, our frequency ranking was based on subjective rating of ten elementary school teachers on a Likert scale of 1–5, that represents a range of low to high frequency on texts available for second graders. The frequency of the selected words ranged from 2 to 4.8, and the average frequency was equal in all conditions (between 3.4 and 3.6).

2.1.4. Procedure

Stimuli were presented on a computer monitor and participants were required to read them aloud, responses and reaction times were recorded using a voice-activated-key (E-prime, Serial Response Box, PST). The trial began with the presentation of a fixation cross, and the presentation of the word was triggered by the participant. The word appeared on the screen 250 ms after button press and remained there until 1200 ms after the onset of the vocal response, following which it was replaced by a fixation cross. Reaction times were collected starting from the stimulus presentation to the onset of vocalization. Words from the current study were intermixed with 56 words from a different experiment (Haddad et al., 2018) with similar frequencies, resulting in a total of 248 trials. Words with and without diacritics were presented in separate blocks of 124 words each to minimize interference from frequent switching between strategies associated with reading pointed and un-pointed words. Block order was counterbalanced across individuals. Data were collected in two sessions during the second trimester of the school year. In the first session the participants performed the standardized tests individually in a quiet room in the school. All participants passed the inclusion criteria and were invited for a second session where they performed the experimental task.

2.1.5. Statistical analysis

Self-corrected responses and words read by sounding each letter separately were coded as correct responses for the analysis of accuracy, but were omitted from the analysis of reaction time. Reaction time was analysed only for correct responses. 1% of the responses were excluded from the analysis of RT due to technical recording problems. Statistical analysis incorporated separate GLMs with response time and accuracy as dependent variables, and diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with 1 vs. without) as within subject factors, and group as a between subject factor (2nd vs. 5th grade). Results are reported separately for accuracy and reaction time and significant effects are reported with p < .05.

2.2. Experiment #2

2.2.1. Participants

A novel sample of sixteen 2nd and 3rd grade students (ages 7.33 to 9, M = 8.2 ± 0.5 years, 8 females) and nine 5th and 6th grade students (ages 10.5 to 12, M = 11.2 ± 0.54, 3 females), participated in the study (the same sample is also reported in Barouch et al., 2022). Two participants from the younger group were excluded from the analysis due to excessive movement during fMRI scanning (see ‘fMRI data pre-processing’ below) resulting in 14 participants in this group. The study was approved by the Helsinki committee of the Souraski Medical Center. Written informed consent was obtained from the parents of all participants, and oral consent was obtained from the children. All participants were native Hebrew speakers, right-handed, with no neurological
disorders, with normal (or corrected to normal) vision, and with no learning disabilities as reported by teachers and confirmed by our assessments (see below).

2.2.2. Standardized tests

The same screening tests were used as in Experiment #1. In addition to screening tests, we tested participants’ phonological abilities in order to examine its association with brain activation. We therefore computed a composite score by combining performance on two phonological processing tests: Reading Pseudo-words (same as Experiment 1) and a Phono-eme omission test (taken from the “Alef-Taf” battery, Shany et al., 2006) which included 16 mono and bi-syllabic words that were read aloud by the examiner. Participants produced pseudo-words obtained by omitting a designated phoneme positioned at the beginning, middle or end of the word. The score reflects the percentage of errors produced. Age-normed z-scores were created for both tests (for the phoneme omission test, the z-scores were multiplied by -1, as this measure recorded error rates rather than accuracy). We used the average of z-scores from both tests as a composite score of phonological abilities, which was used to correlate with brain activation.

2.2.3. Experimental stimuli

Same as in Experiment #1 (see Table 1).

2.2.4. Procedure

Each participant performed three sessions. The screening tests were conducted in the first session that took place at participants’ home, or in the Language Learning lab at the University of Haifa. The second and third sessions took place at The Functional Brain Imaging Center, in Souraski Medical Center. In the second session participants practiced the experimental task using different words, inside a mock scanner, to help children acclimate to the scanner environment and noise, and practice minimizing their movements. The third session included fMRI scanning of the experimental task and an anatomical scan.

In the functional scans each trial began with 200 ms presentation of a fixation cross followed by the presentation of the stimulus word for 1500 ms, and then a blank screen for 2300 ms. Participants were required to read the word aloud as soon as it appeared on the screen, and their responses and reaction times were recorded by an MRI compatible microphone with noise cancellation (FOMRI™ III system, Optoacoustics Ltd). Stimuli were presented using E-Prime stimulus presentation software (v.2.0, Psychological Software Tools, Inc.).

Words from the current study were intermixed with 56 words from a different experiment (Barouch et al., 2022) with similar frequencies, resulting in a total of 248 trials. Words with and without diacritics were presented in separate runs to minimize interference which may arise from frequent shifting between versions. Two runs of pointed words and two runs of un-pointed words appeared in alternating order, and the order was counterbalanced across individuals. 248 experimental trials were intermixed with 48 baseline trials in which the participants saw a string of asterisks and were required to say the word ‘pass’. Trail interval was jittered with 30% time of null and the sequence of trials was optimized using Optseq (Dale, 1999). A total of 296 trials were acquired in four runs of 5:42 min. A practice list of ten different words was presented to participants immediately prior to the first experimental run.

2.2.5. fMRI data acquisition

Images were acquired using a 3.0 T GE scanner with a standard head coil. The stimuli were projected onto a screen, and viewed through a mirror attached to the inside of the head coil. Functional images were acquired with a susceptibility weighted single-shot EPI (echo planar imaging) with BOLD (blood oxygenation level-dependent), with the following parameters: TE = 35 ms, flip angle = 78°, matrix size = 96 × 96, field of view = 20 cm, slice thickness = 3 mm + 1 mm gap, number of slices = 26 in a sequential ascending order, TR = 2000 ms. One hundred seventy-one images were acquired during each run. In addition, a high resolution, anatomical T1 weighted 3D structural images were acquired (AX SPGR, TR = 9.044 ms, TE = 3.0504 ms, flip angle = 13°, matrix size = 256 × 256, field of view = 25.6 cm, slice thickness = 1 mm) using an identical orientation as the functional images.

2.2.6. fMRI data pre-processing

Scanner images (DICOM) were converted to NiTi format using MRIcon software (https://www.sph.sc.edu/comd/rorden/mricon/; Rorden et al., 2007). Data were analysed using the Statistical Parametric Mapping toolbox for Matlab (SPM12 – Welcome Trust Centre for Neuroimaging, University College London, www.fil.ion.ucl.ac.uk/spm), as well as the ArtRepair toolbox (Mazaika et al., 2009)The images were spatially realigned to the first volume in each run to correct for head movements. Spatially realigned images were then smoothed with a 4-mm isotropic Gaussian kernel and underwent motion adjustment and volume artefact detection and correction (ArtRepair programs: Art Motion Regress, Art Global). We used two ArtRepair parameters: (1) global percent threshold = 1.5 (“percent_thresh”), a measure of the mean signal intensity relative to the mean of the run, and (2) mm/TR = 1.5 (“mv thresh”), a measure of scan-to-scan movement. Based on these parameters, runs which had more than 20% of repaired volumes were discarded from the analysis. Two subjects were subsequently excluded.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Example of stimuli for each experimental condition.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With diacritics</td>
</tr>
<tr>
<td>4-consonants with vowel letter</td>
<td>עיר</td>
</tr>
<tr>
<td>4-consonants without vowel letter</td>
<td>עוב</td>
</tr>
<tr>
<td>3-consonants with vowel letter</td>
<td>לנ</td>
</tr>
<tr>
<td>3-consonants without vowel letter</td>
<td>לות</td>
</tr>
<tr>
<td>Frequency range</td>
<td>1.33 - 4.75</td>
</tr>
</tbody>
</table>
who had all runs exceeding this threshold, and ten other runs were excluded from seven participants, of which five runs were excluded from the pointed condition and five runs were excluded from the un-pointed condition. Sinc interpolation was used for slice time correction to minimize timing errors between slices (Henson et al., 1999). The functional images were then co-registered with the anatomical image and normalized to the standard T1 template volume (MNI). The data were smoothed again with a 5-mm isotropic Gaussian kernel.

### 2.2.7. Statistical analysis

#### 2.2.7.1. Performance in the scanner

Only correct responses were included in the analysis (excluding self-corrected responses from both accuracy and RT analysis). Statistical analysis incorporated separate GLMs with response time and accuracy as dependent variables, and diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with 1 vs. without) as within subject factors, and group as a between subject factor, i.e., younger (2nd-3rd) vs. older children (5th-6th graders). Results are reported separately for accuracy and reaction time and significant effects are reported with $p < .05$.

#### 2.2.7.2. Whole brain group analyses

Statistical analyses at the first level were performed for each participant using GLM analysis for event-related designs. Only correct responses were included in the analysis. We used participants’ RT on each trial as the duration of the event and word frequency was included as a parametric modulator. The model included two levels of diacritics (pointed vs. un-pointed), two levels of word length (3 vs. 4-consonants) and two levels of vowel letters (with 1 vs. without vowel letter), as well as the baseline condition. At the second level, two-sample t-tests were carried out to compare between groups, using first level contrasts of all language conditions. In addition, paired T-tests were used to examine the effects of diacritics (pointed vs. un-pointed) and vowel letters (with 1 vs. without). For descriptive purposes, statistical maps are depicted at uncorrected threshold of $p < .001$, and cluster extent threshold of $k \geq 10$ voxels.

#### 2.2.7.3. ROI analyses

In order to test our specific predictions about developmental changes in phonological and orthographic processing we used region of interest (ROI) analyses in three bilateral regions previously shown in the LH to be involved in: (1) **Phonological processing:** inferior temporal gyrus (Brennan et al., 2013; Desroches et al., 2010; Weiss et al., 2018). (2) **Phonological segmentation:** superior temporal gyrus (IFG) pars opercularis (Burton et al., 2000; Hsieh et al., 2001; Poldrack et al., 2001; Weiss et al., 2015a). (3) **Orthographic processing of written words:** visual word form area (Cohen and Dehaene, 2004; McCandliss et al., 2003). Each of these regions was defined in both the left and right hemispheres to account for the possibility of bilateral cortical involvement in language processing in children. ROIs were defined using the MarsBaR tool (Brett et al., 2002) in SPM. Anatomical masks for IFG pars opercularis and STG were defined using an anatomical mask in the Automated Anatomical Labelling (AAL) atlas in the MarsBaR package. Because STG is a large anatomical structure, the AAL mask for STG was split in approximately equal lengths along the long axis, i.e., posterior portion of the STG from $Y = -54$ to $-24$, and anterior portion of the STG from $Y = -22$ to 6. This approach has been used in previous studies for regions that show a gradual change along the long axis, such as the hippocampus (Collin et al., 2015). The VWFA was defined as a 10 mm sphere centred around the MNI coordinates $x = -42, y = -57, z = -6$ (Cohen and Dehaene, 2004) and its right hemisphere homologue.

The top 100 most activated voxels in each of the eight basic reading conditions (2 levels of diacritics x 2 levels of vowel letters x 2 levels of word length) $\times$ the asterisks baseline condition, were selected based on t-values of that contrast within each ROI anatomical mask, separately for each participant. Beta values associated with each condition from the individualised 100-voxel ROIs were then extracted using the MarsBaR toolbox for SPM (Brett et al., 2002). This enabled us to select voxels that were most responsive and sensitive to the experimental manipulation and were therefore more accurate in detecting neural effects. This brain activation extraction method has previously been shown to be more powerful in finding group differences compared to other methods (Tong et al., 2016).

Statistical analyses were carried out using IBM SPSS Statistics Software (v. 19). Separate repeated measures GLM analyses were conducted for each of the three ROIs (i.e., STG, IFG pars opercularis, VWFA), with % signal change from the individualised top 100 activated voxels in each ROI as the dependent variable and hemisphere, diacritics, vowel letters and word length as within-subject variables, and group as between-subject variable. For the STG analysis the anterior and posterior regions were included as another within-subject factor. For interactions between group and/or diacritics with one of the other experimental conditions (vowel letters or length) further analyses were carried out separately for each group, or separately for pointed and un-pointed words respectively. Because the manipulation of length was intended to distinguish between decoding of small units vs. identification of larger units in the transparent and non-transparent scripts, the effect of length was further examined only when there was an interaction with diacritics.

In order to examine the effect of phonological awareness on phonological and orthographic processing during reading we computed a composite score of phonological abilities by combining measures of two phonological processing tests taken from the “Alfel-Taf” battery (Shany et al., 2006): Phoneme Omission and Reading pseudo-words. We assessed whether phonological ability is associated with phonological and orthographic processing differentially in transparent and non-transparent words, and whether this changes across age groups. We conducted individual GLMs within each region with the same within subject and between subject factors used in the above ROI analysis while including the phonological composite score as a covariate in the GLM. Across all analyses, significant effects are reported at the level of $p < .05$.

### 3. Results

#### 3.1. Experiment #1

##### 3.1.1. Screening tests

All participants performed within one standard deviation from the mean on our screening measures (rate of reading words and pseudo-words), as computed based on the age-appropriate norms of the standardized tests (Shany et al., 2006). The scores are presented in Table 2.

##### 3.1.2. Experimental task accuracy

We ran a repeated-measures GLM analysis for accuracy as the dependent variable and the following within subject factors: diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with 1 vs. without), and group as a between subject factor (2nd vs. 5th grade). See Supplementary Table S1 for average performance in all conditions. The analysis showed a significant effect for diacritics.

### Table 2

<table>
<thead>
<tr>
<th>Reading words</th>
<th>2nd graders (n = 27)</th>
<th>5th graders (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>number per minute raw score</td>
<td>45.2 (10.4)</td>
<td>59.2 (14.3)</td>
</tr>
<tr>
<td>number per minute z-score</td>
<td>0.98 (0.7)</td>
<td>0.17 (0.7)</td>
</tr>
<tr>
<td>Reading pseudo-words</td>
<td>24.9 (5.0)</td>
<td>33.0 (5.8)</td>
</tr>
<tr>
<td>number per minute raw score</td>
<td>0.79 (0.6)</td>
<td>0.96 (0.6)</td>
</tr>
<tr>
<td>number per minute z-score</td>
<td>0.96 (0.6)</td>
<td>0.96 (0.6)</td>
</tr>
</tbody>
</table>
group: $F(1, 54) = 11.47, p = .001$, with higher accuracies in 5th graders (Fig. 1a) and a significant effect of diacritics: $F(1, 54) = 45.07, p < .001$, showing better performance for pointed words across groups (Fig. 1a).

A two-way interaction of diacritics and group was significant: $F(1, 54) = 8.48, p = .005$, and there was a three-way interaction between diacritics, length and group: $F(1, 54) = 8.91, p = .004$. These interactions were followed by separate analyses within each group, which revealed a significant effect of diacritics in both groups, grade 2: $F(1, 26) = 27.83, p < .001$ and grade 5: $F(1, 28) = 17.18, p < .001$, suggesting that diacritics improved performance in both groups, but more so in 2nd graders. Only 5th graders showed an interaction of diacritics and length: $F(1, 28) = 6.20, p = .019$, and follow-up analyses revealed that the effect of length was significant only for words without diacritics, length: $F(1, 28) = 9.58, p = .004$, with long words being more accurate than short words. There were no effects in 2nd graders (Fig. 1a).

The main analysis also showed a significant main effect of vowel letters: $F(1, 54) = 13.88, p < .001$. However, a significant three-way interaction of vowel letters, diacritics and length: $F(1, 54) = 10.56, p = .002$, that was followed by separate analyses for with and without diacritics, revealed an opposite and significant effects of vowel letters in each condition. In un-pointed words the presence of a vowel letter reduced reaction time (Fig. 2b).

### 3.2. Experiment #2

#### 3.2.1. Screening tests

Here we report mean raw scores for all participants included in the final group analysis, as well as mean z-scores computed based on age-appropriate norms of the standardized tests (Shany et al., 2006) see Table 3. All participants performed within two standard deviations of the mean of their age group norms. Age-normed z-scores on pseudo-word reading and phoneme omission showed a significant correlation across groups ($r = .393, p = .032$), z-scores from these two measures were combined into a phonological composite score, as described above, which was then correlated with brain activation (described later).

#### 3.2.2. Performance accuracy in the scanner

A repeated-measures GLM analysis for accuracy as the dependent...
variable was conducted with the following within subject factors: diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with 1 vs. without), and group as a between subject factor (younger children vs. older children). See Supplementary Table S2 for average performance in all conditions. The analysis showed a significant effect of length: \( F(1, 21) = 10.64, p = .004 \), with higher accuracies in older children (Fig. 3a). There was a significant main effect of diacritics: \( F(1, 21) = 9.33, p = .006 \). However, a two-way interaction of diacritics and group: \( F(1, 21) = 7.51, p = .012 \) that was followed by separate analyses within each age group, revealed a significant effect of diacritics only in younger children: \( F(1, 13) = 14.27, p = .002 \), suggesting more accurate performance with diacritics than without diacritics, and no effect in older children (Fig. 3a).

The main analysis also showed a significant main effect of vowel letters: \( F(1, 21) = 5.95, p = .024 \), with higher accuracy for words with vowel letters than without (Fig. 3b).

Table 3
Participants’ average performance (and standard deviation) on the screening tests. Standardized scores are based on the norms in Shany et al. (2006).

<table>
<thead>
<tr>
<th></th>
<th>Younger children (n = 14)</th>
<th>Older children (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading words</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number per minute raw score</td>
<td>33.75 (10.75)</td>
<td>48.0 (8.10)</td>
</tr>
<tr>
<td>number per minute z-score</td>
<td>-0.05 (0.67)</td>
<td>-0.53 (0.40)</td>
</tr>
<tr>
<td><strong>Reading pseudo-words</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>number per minute raw score</td>
<td>19.85 (4.17)</td>
<td>21.88 (5.23)</td>
</tr>
<tr>
<td>number per minute z-score</td>
<td>0.14 (0.49)</td>
<td>-0.01 (0.41)</td>
</tr>
<tr>
<td><strong>Phoneme omission</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% errors raw score</td>
<td>27.36 (20.87)</td>
<td>12.97 (13.45)</td>
</tr>
<tr>
<td>% errors z-score</td>
<td>-0.58 (0.78)</td>
<td>-0.40 (0.57)</td>
</tr>
</tbody>
</table>

3.2.3. Reaction times in the scanner

A repeated-measures GLM was also conducted with reaction time as the dependent variable and the following within subject factors: diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with vs. without), and group as a between subject factor (younger children vs. older children). See Supplementary Table S2 for average reaction time in all conditions. The analysis showed significantly faster responses in older children: \( F(1, 21) = 8.25, p = .009 \) (Fig. 4). There was also a significant main effect of length: \( F(1, 21) = 6.24, p = .021 \) and a two-way interaction of group and length: \( F(1, 21) = 4.41, p = .048 \). This was followed by separate analyses within each group, which revealed a significant effect of length only in younger children: \( F(1, 13) = 8.70, p = .011 \), suggesting slower reaction times for words with 4-consonants, and no effects in older children. The analysis in younger children also showed a significant effect of diacritics: \( F(1, 13) = 5.06, p = .042 \), with slower responses for words without diacritics than with diacritics.

3.2.4. Whole-brain analyses

Whole-brain analyses were performed to examine the main effects of group (across all conditions), diacritics (pointed vs. un-pointed) and vowel letters (with 1 vs. without). No activation was found after correcting for multiple comparisons (\( p < .05 \), FWE corrected), therefore, the results are presented at the uncorrected level (\( p < .001 \), uncorrected, \( k \geq 10 \)) for descriptive purpose. A two-sample \( t \)-test comparing between groups across all conditions revealed significantly greater activation for older vs. younger children in right supramarginal gyrus and right middle temporal gyrus (Table 4a; Fig. 5a), and no significant effects for younger vs. older children. A paired-sample \( t \)-test testing the effect of diacritics across age groups and across all other conditions did not reveal any significant activations. Lastly, paired sample \( t \)-test of vowel letters across groups and all other conditions revealed significantly greater activation for words with vowel letters compared to without in bilateral fusiform.
gyrus, right inferior occipital gyrus, right superior temporal gyrus and right middle frontal gyrus (Table 4b, Fig. 5b). No significant active clusters were found in the comparison for words without vs. with vowel letters. We further examined our hypotheses using ROI analyses.

3.2.5. ROI analyses

We conducted separate GLM analyses for the three bilateral ROIs, i.e., IFG pars opercularis, STG and VWFA, with % signal change for the individualised top 100 voxels as the dependent variable. The following within subject factors were used: region (anterior vs. posterior – only in STG), hemisphere (left vs. right), diacritics (pointed vs. un-pointed), length (3-consonants vs. 4-consonants) and vowel letters (with 1 vs. without), and group as a between subject factor (younger vs. older children). All significant main effects and interactions from these analyses are summarized in Table 5. The following sections describe these effects and their follow-up analyses by our key experimental factors, i.e., (i) group, and its interaction with hemisphere, (ii) vowel letters, (iii) diacritics, and its interactions with length and hemisphere.

(i) Effects of age and its interaction with hemispheric lateralization

In order to identify general developmental shifts in reading strategy, and developmental changes in hemispheric asymmetry we looked for regions showing a main effect of age group, or an interaction of group by hemisphere. A significant main effect of group was seen in bilateral IFG
effect of diacritics, we examined regions showing an interaction with diacritics.

The manipulation of word length was included in the study as an indication of reliance on piecemeal decoding of small orthographic units (greater activation for long vs. short words), vs. reliance on larger orthographic units (greater activation for short vs. long words). We therefore also looked at regions showing an interaction of diacritics and length. Two regions, STG and VWFA showed a two-way interaction of diacritics and length across groups. Follow-up analysis of STG, split by diacritics, revealed that for words presented with diacritics, greater activation was found for long compared to short words: \( F(1,21) = 10.21, p = .005 \) (see Fig. 7, upper panels), and there was no effect of length for words without diacritics.

Follow-up analysis in the VWFA, split by diacritics, showed that for words presented without diacritics, there was a non-significant trend that short words elicited greater activation in comparison to long words: \( F(1, 21) = 3.15, p = .091 \) (Fig. 7, lower panels), and there was no effect of length for words presented with diacritics.

There was also a significant three-way interaction of hemisphere, diacritics and group in the VWFA (see Table 5). Follow-up analysis split by group, revealed that only in the younger group there was a marginal effect for greater activation in the right VWFA for words presented without diacritics compared to with diacritics: \( F(1,13) = 4.59, p = .053 \), and no difference in the left VWFA. Finally, we also found a three-way interaction of region, diacritics and vowel letters in STG (see Table 5), however, follow-up analysis split by region did not reveal significant results in either subregion of STG.

3.2.6. Correlations with phonological ability

In order to examine the effect of phonological ability on phonological and orthographic processing during reading we included the composite score of phonological ability in the above GLM analyses as a covariate. We also included a measure of word reading ability (i.e., Reading words test) as an additional covariate in the GLM to control for general word recognition ability. We only report regions that showed a main effect of phonological ability or an interaction of phonological abilities with diacritics or age. Only in the VWFA there was a main effect of phonological abilities: \( F(1, 19) = 4.93, p = .039 \), and no effect of word reading ability: \( F(1, 19) = 1.54, p = .231 \), and no interaction with diacritics or group. We found a significant positive correlation between the phonological composite score and activation in the VWFA across diacritics (see Fig. 8), after controlling for reading ability \( (r = 0.482, p = .035) \).

---

Table 4  
Regions showing activation in the whole-brain analysis: (a) Activation across all conditions in older vs. younger children. (b) Activation in words with vowel letters vs. without vowel letters, across groups. Significant at threshold \( p < .001 \) uncorrected, with cluster extent \( k \geq 10 \).

<table>
<thead>
<tr>
<th>Area</th>
<th>BA</th>
<th>H</th>
<th>Z</th>
<th>Voxels</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
</table>
| (a) All conditions: Older vs. Younger children
| Supramarginal Gyrus   | 40 | R   | 3.76 | 47     | 50  | -50 | 34   |
| Middle Temporal Gyrus | 22 | R   | 3.38 | 24     | 66  | -38 | -6   |
| (b) Vowel letters: with 1 vs. without
| Fusiform Gyrus        | 37 | R   | 4.25 | 249    | 30  | -58 | -12  |
| Precuneus             | 31 | L   | 3.93 | 217    | -12 | -70 | 20   |
| Cingulate gyrus       | 32 | R   | 3.78 | 160    | 10  | -26 | 28   |
| Thalamus              | 41 | L   | 4.81 | 137    | 18  | -34 | 2    |
| Fusiform Gyrus        | 37 | R   | 4.14 | 100    | -28 | -60 | -14  |
| Inferior occipital gyrus | 17 | R   | 3.86 | 99     | 40  | -74 | -10  |
| Superior Temporal Gyrus | 22 | R   | 3.65 | 94     | 50  | -46 | 12   |
| Middle Frontal Gyrus  | 10 | R   | 3.41 | 33     | 2   | 4   | 48   |

(iii) Effects of Diacritics

We also looked at how phonological and orthographic processes are affected by orthographic transparency. While we did not find a main

---

Table 5  
Significant main effects and interactions in each ROI.

<table>
<thead>
<tr>
<th>Effects</th>
<th>df</th>
<th>( F ), ( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferior frontal gyrus pars opercularis</td>
<td>(1, 21)</td>
<td>6.31, .021</td>
</tr>
<tr>
<td>Hemisphere</td>
<td>(1, 21)</td>
<td>19.87, &lt;.001</td>
</tr>
<tr>
<td>Hemispheres x group</td>
<td>(1, 21)</td>
<td>5.60, .028</td>
</tr>
<tr>
<td>Superior temporal gyrus</td>
<td>(1, 21)</td>
<td>4.45, .048</td>
</tr>
<tr>
<td>Hemisphere x group</td>
<td>(1, 21)</td>
<td>6.24, .021</td>
</tr>
<tr>
<td>Hemispheres x length</td>
<td>(1, 21)</td>
<td>5.55, .029</td>
</tr>
<tr>
<td>Diacritics x length</td>
<td>(1, 21)</td>
<td>11.95, .002</td>
</tr>
<tr>
<td>Region x diacritics x vowels</td>
<td>(1, 21)</td>
<td>4.57, .045</td>
</tr>
<tr>
<td>Length x vowels</td>
<td>(1, 21)</td>
<td>5.37, .031</td>
</tr>
<tr>
<td>Region x hemisphere x diacritics x length x vowels x group</td>
<td>(1, 21)</td>
<td>18.08, &lt;.001</td>
</tr>
<tr>
<td>Hemispheres x length</td>
<td>(1, 21)</td>
<td>21.61, &lt;.001</td>
</tr>
<tr>
<td>Hemispheres x diacritics x group</td>
<td>(1, 21)</td>
<td>5.04, .036</td>
</tr>
<tr>
<td>Diacritics x length</td>
<td>(1, 21)</td>
<td>5.54, .029</td>
</tr>
<tr>
<td>Vowels x length</td>
<td>(1, 21)</td>
<td>14.98, .001</td>
</tr>
</tbody>
</table>

---

Fig. 5. Whole brain analysis: (a) Older children compared to younger children, across all conditions. (b) With vowel letters compared to without vowel letters across groups. Significance threshold \( p < .001 \) uncorrected, cluster extent \( k \geq 10 \).

pars opercularis and STG (see Table 5), with more activation for older children compared to younger children (see Fig. 6). There was also a significant two-way interaction between hemisphere and group in both of these regions. Follow-up analyses split by group revealed significant effect of hemisphere in the older children group in both IFG pars opercularis: \( F(1, 8) = 15.77, p = .004 \), and STG: \( F(1, 8) = 7.02, p = .029 \), with more activation in the left hemisphere compared to the right, and no significant differences between hemispheres in the younger group (see Fig. 6).

(ii) Effects of Diacritics

We also looked at how phonological and orthographic processes are affected by orthographic transparency. While we did not find a main
4. Discussion

The current study examined the effect of orthographic transparency on phonological and orthographic processing during reading acquisition in young Hebrew speakers. In a behavioural (experiment #1) and an fMRI (experiment #2) study, we manipulated the levels of orthographic transparency using diacritics and vowel letters, and their interaction with word length. We examined their effect on word reading and on the neural activity in regions associated with phonological and orthographic processes in younger (2nd & 3rd graders) and older (5th & 6th graders) Hebrew reading children.

Fig. 6. *General developmental changes*: Main effect of group, and interaction between hemisphere and group in inferior frontal gyrus pars opercularis (IFG_Oper), and superior temporal gyri (STG). LH: left hemisphere, RH: right hemisphere. Error bars indicate standard errors. Significant effects ($p < .05$) are marked by asterisks, a larger asterisk indicates a significant difference between groups.

Fig. 7. *Interaction of word length and diacritics*: shown in bilateral superior temporal gyri (STG) and bilateral visual word form area (VWFA), across groups. Error bars indicate standard errors. Significant effects ($p < .05$) are marked by asterisks.
To summarise our behavioural results, the benefit of diacritics was observed in both experiments, mostly for younger children. Younger children had higher accuracy when reading words in the presence of diacritics in both experiments, and also faster reaction time for reading pointed words in experiment #2. In contrast, older children read words with and without diacritics at a similar speed in both experiments, and only benefited from diacritics in accuracy in experiment #1. Younger children also showed an effect of word length on reaction times in both experiments, being slower for long words compared to short ones. On the other hand, only 5th graders in experiment #1 showed a reversed length effect for un-pointed words, reading long words more accurately than short words. Finally, in experiment #1, across both age groups, the presence of vowel letters facilitated performance for words without diacritics (in accuracy and RT), but reduced accuracy for words with diacritics.

4.1. Behavioural results

4.1.1. Effects of diacritics and their interaction with word length and age

Our behavioural results from both experiments showed greater facilitation effect of diacritics in the younger group in comparison to the older children, who, having more reading experience of un-pointed words and a richer lexicon, read words with and without diacritics with similar speed. Younger children also showed a main effect of word length in RT, i.e., slower responses for long compared to short words, regardless of diacritics. These findings are consistent with our predictions and with findings from young readers in other orthographies (Samuels et al., 1978), suggesting that younger children rely on serial piecemeal decoding, i.e., overtly mapping individual letters to sounds, therefore reading longer words slower than short words, regardless of transparency. Contrary to our prediction, older children did not show length effects for pointed words like skilled adults (Weiss et al., 2015b), suggesting they were able to identify the orthographic patterns of whole words even when they are pointed, and hence did not have to rely on serial decoding of individual letters. This could be because older children are exposed to the pointed script more often than adults and it has been less time since they learned it, hence they may be more familiar with it.

Interestingly, older children did show a reversed word length in un-pointed words, reading longer words more accurately than short words, consistent with our prediction and with performance in adults (Weiss et al., 2015b). This advantage for longer words may result from their smaller orthographic neighbourhood, i.e., fewer orthographically similar words (Coltheart et al., 1977), reducing orthographic competition and making long words easier to identify as whole-word units. The finding of this effect for un-pointed words in older children further supports our conclusion that older children, like adults (Frost, 2005; Katz and Frost, 1992; Weiss et al., 2015b) identify un-pointed words as larger orthographic units.

4.1.2. Processing of vowel letters

As predicted, our behavioural results showed a facilitatory effect of vowel letters across both age groups. This effect was found across both levels of transparency in experiment #2, and specifically for un-pointed words in experiment #1. This effect is consistent with previous studies in skilled Hebrew readers, suggesting that vowel letters provide phonological information that is missing in un-pointed words (Frost, 1995; Schiff and Ravid, 2004; Weiss et al., 2015a). Additionally, vowel letters may also improve word recognition by facilitating access to the orthographic representation, in a similar manner to the effect of an additional consonant reducing orthographic neighbourhood competition (Weiss et al., 2015b). The opposite (negative) effect of vowel letters on the accuracy of words with diacritics (in experiment #1), is consistent with a previous study showing a similar effect in dyslexic adult readers (Weiss et al., 2015b). In pointed words, where the phonological information is fully specified, the inherent ambiguity of vowel letters may hinder word recognition, particularly in groups with less stable orthographic representations.

4.2. Neural activation

Our ROI analyses focused on three bilateral regions, two of which are relevant for phonological aspects of reading acquisition, including IFG pars opercularis (Burton et al., 2000; Hsieh et al., 2001; Poldrack et al., 2001) and STG (Brennan et al., 2013; Desroches et al., 2010), and the third, the VWFA, involved in orthographic processing (Cohen and Dehaene, 2004; McCandliss et al., 2003). These regions revealed developmental changes, as well as effects of hemispheric lateralization, diacritics and word length, discussed separately in the following sections.

4.2.1. Developmental changes in activation and lateralisation

The ROI analysis revealed an overall increase in activation for older children compared to younger children in IFG pars opercularis. This finding is consistent with reading acquisition studies across languages and orthographies which showed a developmental increase in activation in left IFG more generally (Bitan et al., 2007; Brown et al., 2005; Cherodath and Singh, 2015; Chyl et al., 2018; Holland et al., 2001;
There were no significant simple effects but only a trend for a reversed effect of word length in activation in anterior STG. This suggests that children, both young and old, rely more on larger orthographic units when reading words without diacritics.

### 4.2.2. Effects of diacritics and their interaction with word length

We found an interaction between diacritics and word length (across the two age groups), which resulted from opposite effects of length in words with and without diacritics in both bilateral STG and bilateral VWFA. Namely, only pointed words showed a standard word length effect (i.e., greater activation for long than short words), and this was significant in STG. While there were no parallel behavioural findings in the current study, these results are in line with previous behavioural findings in skilled adult Hebrew readers, who showed slower responses for long compared to short words only when presented with diacritics (Weiss et al., 2015b). These results suggest that when reading pointed words children engage in processing smaller units, resulting in greater activation for longer than shorter words, more than when reading un-pointed words. Given the possible association of STG with phonological processing, this may indicate processing of smaller phonological units. The absence of similar effects in reaction time suggests that pro-

### 4.2.3. Correlations with standardized tests

Lastly, our correlational analysis showed that individual phonological processing abilities correlated with activation in bilateral VWFA, during reading of words with and without diacritics across groups, i.e., better phonological abilities were associated with higher activation even when controlling for word reading ability. Given the possible involvement of the VWFA in orthographic processing (Cohen and Dehaene, 2004; Dehaene and Cohen, 2011; Glezer et al., 2009; Hirshorn et al., 2016; Stevens et al., 2017), this finding supports the interpretation of the reversed length effect as reflecting greater orthographic competition when faced by short words, due to their large orthographic neighbourhoods. Thus, this interaction of word length and diacritics in the VWFA suggest that children, both young and old, rely more on larger orthographic units when reading words without diacritics.

### 4.3. Limitations

One major limitation of the current study is the small number of participants in the fMRI study, particularly in the older group. Cultural factors may have contributed to families’ reluctance to participate in an imaging study. The great difficulty in recruiting children for the study also resulted in a large age-range within each group of children. This problem is compounded by the higher levels of head motion in this age range, which resulted in exclusion of 18% of the data. These factors have reduced the statistical power of the study and may have contributed to the absence of larger differences in brain activation between the two age groups. The small sample size in the fMRI experiment may also explain some of the differences in behavioural results between experiments #1 and #2.

### 4.4. Summary

This is the first fMRI study to examine the developmental processes associated with reading acquisition in young Hebrew speakers. While some of our results are unique to the properties of the Hebrew dual orthography, they provide important insights into the effects of orthographic transparency, and the nature of developmental changes during reading acquisition more generally.

Our study shows age related difference between the younger and older children groups in both behavioural and neural measures. The
and Goswami, 2005) that suggest that with age, processing of several sub-lexical phonological units do not decrease with age, but may change route (Grainger et al., 2012) and with the grain size hypothesis (Ziegler and Goswami, 2005) that suggest that with age, processing of several size units can occur in parallel. They further show that the role of sub-lexical phonological units do not decrease with age, but may change and become more implicit (Grainger et al., 2012; Milledge and Blythe, 2019). The developmental increase in frontal activation and in left lateralization during reading in Hebrew, which is similar to findings in other orthographies (Bitan et al., 2007; Turkeltaub et al., 2003), indicate that these maturational trajectories may be independent of the specific orthography the children are reading. On the other hand, the developmental increase seen in bilateral STG, is in contrast to previous studies that have shown a developmental decrease in STG in English and Chinese speakers (Bitan et al., 2007; Cao et al., 2010), suggesting that some developmental changes may depend on the specific orthography.

Our findings comparing words with and without diacritics and their interaction with word length show that across both age groups, children rely on both orthographic and phonological processes to read pointed and un-pointed words. However, children adjusted their reliance on different aspects of the neural reading network depending on the transparency of the script. For pointed words they process smaller phonological units (as evident by the interaction of word length and diacritics in STG). For reading un-pointed words children in both age groups tend to rely on larger orthographic units (evidenced by the interaction of word length and diacritics in VWFA). These differences in the neural mechanisms involved in reading the two versions of the orthography are in line with the grain size hypothesis (Ziegler and Goswami, 2005), suggesting that in transparent orthographies (early) reading involves greater reliance on decoding of small orthographic and phonological units compared to (early) reading in an opaque orthography. The findings of the current study show that this is true even when the same individual reads in two scripts that differ in transparency.

Finally, our results also show that phonological abilities are associated with access to orthographic processing in the VWFA. This finding is in line with Share’s self-teaching hypothesis (Share, 1995; Share and Bar-On, 2018), suggesting that learning to decode in the early stages of reading helps children acquire and develop their orthographic knowledge. The finding of this correlation across both pointed and un-pointed words, and across both age-groups is a further support for multiple route and connectionist models which argue for a continued role of phonology at all stages of reading acquisition (Milledge and Blythe, 2019; Grainger et al., 2012) and for all levels of orthographic transparency.

Author contributions

Upasana Nathaniel: Methodology, Formal analysis, Investigation, Visualization, Writing. Yael Weiss: Conceptualization, Methodology, Investigation. Bechor Barouch: Methodology, Formal analysis. Tami Katzir: Conceptualization, Methodology, Funding acquisition, Writing. Tali Bitan: Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Funding acquisition, Writing.

Funding

The study was funded by the Israel Foundation Trustees (IFT), grant 34/2011 to Tali Bitan, and by the Israel Science Foundation (ISF) - Jerusalem, Israel grant 1142/11 to Tali Bitan and Tami Katzir.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2022.108376.

References

CERCOR/BHP186.


