



Neural representations of phonology in temporal cortex scaffold longitudinal reading gains in 5- to 7-year-old children

Jin Wang^{a,*}, Marc F. Joanisse^b, James R. Booth^a

^a Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

^b Department of Psychology & Brain and Mind Institute, The University of Western Ontario, London, Ontario, Canada

ARTICLE INFO

Keywords:

Phonology
Scaffolding
Refinement
Representation
Access
Grain size
Onset
Rhyme

ABSTRACT

The objective of this study was to investigate whether phonological processes measured through brain activation are crucial for the development of reading skill (i.e. scaffolding hypothesis) and/or whether learning to read words fine-tunes phonology in the brain (i.e. refinement hypothesis). We specifically looked at how different grain sizes in two brain regions implicated in phonological processing played a role in this bidirectional relation. According to the dual-stream model of speech processing and previous empirical studies, the posterior superior temporal gyrus (STG) appears to be a perceptual region associated with phonological representations, whereas the dorsal inferior frontal gyrus (IFG) appears to be an articulatory region that accesses phonological representations in STG during more difficult tasks. 36 children completed a reading test outside the scanner and an auditory phonological task which included both small (i.e. onset) and large (i.e. rhyme) grain size conditions inside the scanner when they were 5.5–6.5 years old (Time 1) and once again approximately 1.5 years later (Time 2). To study the scaffolding hypothesis, a regression analysis was carried out by entering brain activation in either STG or IFG for either small (onset > perceptual) or large (rhyme > perceptual) grain size phonological processing at T1 as the predictors and reading skill at T2 as the dependent measure, with several covariates of no interest included. To study the refinement hypothesis, the regression analysis included reading skill at T1 as the predictor and brain activation in either STG or IFG for either small or large grain size phonological processing at T2 as the dependent measures, with several covariates of no interest included. We found that only posterior STG, regardless of grain size, was predictive of reading gains. Parallel models with only behavioral accuracy were not significant. Taken together, our results suggest that the representational quality of phonology in temporal cortex is crucial for reading development. Moreover, our study provides neural evidence supporting the scaffolding hypothesis, suggesting that brain measures of phonology could be helpful in early identification of reading difficulties.

1. Introduction

Phonological awareness refers to an individual's ability to represent and access the sound structure of spoken words (Treiman and Zukowski, 1991). Phonology has small (e.g. phoneme) and large (e.g. rhyme) units, also known as grain sizes. Studies have shown that phonological awareness is related to reading skill, with small grain awareness of phonemes having a stronger correlation with children's word reading skills than large grain awareness of rhymes (see Melby-Lervåg et al., 2012 for review). Understanding the causal relation between phonological awareness and reading skill is important because it may inform early diagnosis and intervention on children at-risk for reading difficulty.

Behavioral studies have explored the causal relation between

phonological awareness and reading skill. One argument is that phonological awareness is a foundation of later reading, which we refer to as the scaffolding hypothesis. The term “scaffolding” in our study means that phonological awareness contributes to reading acquisition and should predict reading gains. Mechanistically speaking, phonological awareness, the awareness that a spoken word is composed of small sound units, contributes to reading acquisition by facilitating the establishment of a connection between strings of discrete letters and sequences of phonemes. Longitudinal studies have shown that phonological awareness in prereaders (or children with limited reading skills) is predictive of subsequent reading in the first few years of schooling (e.g. Perfetti et al., 1987; Wagner et al., 1997; Roth et al., 2002; Schatschneider et al., 2004; Lervåg et al., 2009). Studies distinguishing between different grain sizes

* Corresponding author.

E-mail address: jin.wang@vanderbilt.edu (J. Wang).

<https://doi.org/10.1016/j.neuroimage.2019.116359>

Received 20 June 2019; Received in revised form 11 November 2019; Accepted 12 November 2019

Available online 14 November 2019

1053-8119/© 2019 Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

have generally found that measures of phonemic awareness are more powerful concurrent and longitudinal predictors of children's reading skills than measures of rhyme awareness, after controlling for pretest reading skills (e.g. [Muter et al., 1998](#); [Hulme, 2002](#); [Muter et al., 2004](#); [Castles and Coltheart, 2004](#)). Training studies have also supported the scaffolding hypothesis. For example, [Lundberg et al. \(1988\)](#) trained phonological awareness in preschoolers at both the phoneme and rhyme level. They found that the training group showed a significant improvement in phonemic awareness and this had a benefit on reading and spelling skills in both grade 1 and 2, as compared to a control group. Additional work shows that phonemic awareness training improves word reading skills in first and second graders, especially for children at risk of dyslexia (e.g. [Torgesen et al., 1992](#); [Bentin and Leshem, 1993](#); [Hatcher et al., 2004](#)). In summary, phonological awareness scaffolds later reading ability, and small grain phonemic awareness plays a more important role in reading acquisition as compared to rhyme awareness.

In contrast to the scaffolding hypothesis, [Ziegler and Goswami \(2005\)](#) proposed that small grain phonemic awareness only appears after reading acquisition, because learning to read helps children distinguish acoustically inseparable sounds by mapping them to letters. We refer to this as the refinement hypothesis; however, longitudinal studies provide inconsistent evidence. For example, [Perfetti et al. \(1987\)](#) tested phonemic awareness and word reading ability in children at four time points during their first grade. They found that reading skill was predictive of later phonemic awareness, supporting the refinement hypothesis. In contrast, [Wagner et al. \(1997\)](#) designed a 5-year longitudinal study on children from kindergarten to 4th grade. They measured their phonological awareness (mainly at phoneme level) and word reading skills every year. They found that word reading skill at the beginning of kindergarten or first grade was not predictive of the individual differences in later phonological awareness in 2nd grade or 4th grade, arguing against the refinement hypothesis. In summary, phonological awareness, especially small grain phonemic awareness, appears to scaffold later word reading skill, but it remains unclear whether early word reading skill refines later phonemic awareness.

As compared to the extensive behavioral studies, little is known about the neural basis of the bidirectional relations between phonological awareness and reading skill. Because performing phonological tasks usually involves multiple cognitive processes, behavioral measures cannot easily tease these components apart. For example, making a rhyming judgment involves activating phonological representations, accessing those representations, and holding those representations in working memory in order to make a decision about whether they rhyme or not. Neuroimaging studies could provide a complementary measurement to probe whether the nature of phonological representations or access to them is related to reading skill by examining the brain regions associated with these different cognitive functions. This question also speaks to the debate on the nature of the phonological deficit in dyslexia; whether children with dyslexia have deficits in the quality of representations or access to those representations ([Peterson and Pennington, 2015](#)), and whether difficulties with phonological awareness are the cause or consequence of their reading difficulties ([Castles and Coltheart, 2004](#)).

According to the literature review by [Price \(2012\)](#), the phonological processing of speech sounds arises from the functional integration of perceptual processing in the posterior superior temporal gyrus (STG) and articulatory processing in the dorsal inferior frontal gyrus (IFG), which is also called the dorsal stream of speech processing in the dual stream model proposed by [Hickok and Poeppel \(2004, 2007\)](#). According to these models, STG and IFG have distinct roles in phonological processing. STG appears to be a perceptual region associated with phonological representations. Many previous studies, using multivariate pattern analysis or the electrocorticogram (ECoG) technique, have found that STG represents the acoustic and perceptual features of phonology (see review in [Leonard and Chang, 2014](#)) and is able to decode different phoneme categories (e.g. [Boets et al., 2013](#)). In contrast, the dorsal IFG appears to

be unable to represent phonemes (e.g. [Boets et al., 2013](#)). The IFG accesses the phonological representations through connections via the arcuate fasciculus with STG ([Saur et al., 2008](#); [Boets et al., 2013](#)), and is specialized for phonological processing during language production ([Klaus and Hartwigsen, 2019](#)). Increased activation in the dorsal IFG is often found when accessing phonological representations needed for challenging tasks such as segmenting phonemes, processing ambiguous speech, or articulating phonologically dissimilar words (e.g. [Burton et al., 2000](#); [Xie and Myers, 2018](#); [Okada et al., 2018](#)). Overall, the dual-stream model of speech processing and previous empirical studies consistently support the argument that STG is associated with phonological representations whereas the dorsal IFG is an articulatory region associated with accessing phonological representations during more difficult tasks. By examining the left posterior STG and dorsal IFG, we can shed light on whether the nature of the phonological representations and/or access to these representations are related to reading skill.

Neuroimaging studies have suggested that brain activation in STG and IFG during auditory phonological tasks in young children is related to their reading. One line of research compares brain activation during auditory phonological tasks between young children with a family risk of dyslexia and those without a family risk of dyslexia. [Raschle et al. \(2012\)](#) found reduced activation in STG during an auditory onset judgment task in 5-6-year-old pre-readers who have a family risk of dyslexia as compared to those who do not. [Debska et al. \(2016\)](#) found that both IFG and STG were under-activated during an auditory rhyming task in 6-7-year-old children who had a family risk of dyslexia as compared to those who did not. [Vandermosten et al. \(2019\)](#) found that 8-year-old children with a family risk of dyslexia showed atypical phonemic representations in STG, as compared to children without a family risk. These studies suggest that phonological processing, especially phonological representations housed in STG, is related to the reading potential of young children.

Another line of research directly correlates children's brain activation during auditory phonological tasks to their reading skills. [Maurer et al. \(2009\)](#) used the event related potential (ERP) technique and measured 6-year-old kindergarteners' late mismatch negativity (IMMN) to phonemes (one standard/ba/and two deviants/ta/and/da/). They found that the hemispheric lateralization of IMMN in STG when they were in kindergarten predicted their later reading skills. Moreover, the brain measurement for phonological awareness, but not behavioral measures, predicted the children's long-term reading success in 5th grade when children were approximately 11 years old. [Wang et al. \(2018\)](#) used functional magnetic resonance imaging (fMRI) and asked children to do an auditory phonological awareness task including both small (i.e. onset) and large (i.e. rhyme) grain sizes. They found that reading skill was correlated with brain activation for small grain sizes as compared to large grain sizes in left ventral occipitotemporal cortex. This suggests that better reading is associated with the automatic activation of orthography even when just listening to words. However, they only focused on the brain activation in left ventral occipitotemporal cortex, and did not examine phonological processing areas in STG and IFG. Overall, these results suggest that the brain activation during phonological awareness tasks in young children is related reading skill, and that brain measures might be a sensitive tool in predicting future reading skill.

In summary, the previous neuroimaging studies suggest that brain activation in STG and IFG during phonological processing is related to reading ability in young children. However, these previous studies did not examine whether phonological awareness scaffolds reading or whether reading refines phonological awareness. Although the study by [Maurer et al. \(2009\)](#) used a longitudinal design and suggests that phonemic processing as indicated by IMMN scaffolds later reading skill during school years, it did not address whether early reading skill refines later phonological processing in the brain. Moreover, no studies except for [Wang et al. \(2018\)](#) explored how different grain sizes of phonology in the brain relates to reading skills, and they only focused on left ventral occipitotemporal cortex which is implicated in orthographic processing.

Therefore, we still do not know the relation of reading skill to brain regions implicated in phonological processes at different grain sizes.

In the current study, in order to determine the directionality of the relation between phonological processing in the brain and reading skill, we implemented a cross-lagged panel design (Kearney, 2003). We asked children to complete an auditory phonological task in the scanner and a standardized reading test outside of the scanner when they were 5.5–6.5 years old at Time 1 (T1). These children were invited back approximately 1.5 years later when they were 7–8 years old and tested again at Time 2 (T2). In the standardized reading test, participants were asked to read aloud the letters and words. In the phonological task, participants were asked to judge whether the two auditory words presented sequentially shared the same sound. We included a small grain size condition (one phoneme shared in the onset position) and a large grain size condition (two or three phonemes shared in the rhyme position). Because the posterior STG is a perceptual region associated with phonological representations, whereas the dorsal IFG is an articulatory region associated with accessing these representations stored in STG (Hickok and Poeppel, 2004, 2007; Price, 2012; Boets et al., 2013; Ramus, 2014), our analysis focused on the posterior STG and dorsal IFG regions, respectively, as measures of phonological representation versus access. In this way, we could examine how the phonological representations of onset and rhyme in the brain and access to them are related to reading skill.

To examine the scaffolding hypothesis, we tested whether brain activation for phonological processing at T1 predicted reading skill at T2 after controlling for reading skill at T1. Based on previous behavioral studies which found that early phonological awareness in preschoolers predicted later reading skill in school age children (e.g. Roth et al., 2002; Schatschneider et al., 2004; Lervåg et al., 2009), we predicted that brain activation for both small and large grain size phonological processing would predict reading gains. It is possible though that small grain size would play a more important role than large grain size because many behavioral studies have found that phonemic awareness is a better predictor of later reading as compared to rhyme awareness (Muter et al., 1998; Hulme, 2002; Muter et al., 2004; Castles and Coltheart, 2004). However, English is more consistent in mapping from spelling to pronunciation at the rhyme level than at the phoneme level (Treiman et al., 1995), so it is also possible that large grain size should be more predictive of reading gains than small grain size (Ziegler and Goswami, 2005).

To examine the refinement hypothesis, we tested whether reading skill at T1 predicted brain activation during phonological processing at T2 after controlling for brain activation at T1. Because previous results are inconsistent (i.e. Perfetti et al., 1987; Wagner et al., 1997), it is unclear whether early reading skill should predict changes in brain activation during phonological processing. Based on the Ziegler and Goswami (2005) argument that reading acquisition facilitates the discovery of phonemes, we expected that, if the refinement effect exists, early reading skill should predict changes in brain activation during phonological processing of small grain sizes compared to large grain sizes.

In our examination of the scaffolding and refinement hypotheses, we focused on the posterior STG and the dorsal IFG regions, because these regions are implicated in representing and accessing phonology. Because activation in the dorsal IFG is also modulated by phonological working memory load (e.g. Fegen et al., 2015) and often found to be involved in effortful phonological tasks that require phonological working memory (e.g. Démonet et al., 1994; Burton et al., 2005; Ramus, 2014), we included phonological working memory as a covariate of no interest in our study in order to tap into how phonological awareness instead of phonological working memory relates to reading skill. Because an ERP study found hemispheric lateralization in STG predicted subsequent reading skill (Maurer et al., 2009) and because fMRI studies on children at risk of dyslexia mainly found differences in the STG (Raschle et al., 2012; Dębska et al., 2016), we expected that the effects of scaffolding or refinement would be more likely to occur in STG rather than IFG.

2. Method

2.1. Participants

Thirty six children (19 girls, mean age = 5.9, range 5.6–6.5 years old at Time 1, mean age = 7.5, range 7.1–8.2 years old at Time 2) were included in this study. Most (31 out of 36) of the 5-6-year-old children at T1 were kindergarteners. Four were pre-kindergarteners (pre-K) and one was just entering first grade. Children at T2 were all first or second graders with one or two years of formal reading education. There were originally 71 participants attending both Time 1 (T1) and Time 2 (T2) fMRI tasks and behavioral standardized testing. Among them, 26 subjects were discarded because of excessive movement (see movement criteria in the data analysis section). Two subjects were discarded due to image quality problem (i.e. ghosting and insufficient brain coverage). Six subjects were discarded because they did not meet our accuracy criteria (see acceptable accuracy criteria in the experimental procedure section). One subject was discarded because the two runs of the auditory phonological task at T2 were more than 6 months apart. Children were recruited from the Austin, Texas metropolitan area. Informed consent was obtained from the parents. The Institutional Review Board at the University of Texas at Austin approved all the following procedures.

Participants completed a series of screening tests and their parents/guardians completed a developmental history questionnaire. The screening tests included the 5-handedness questions in which the children needed to pretend they write, draw, pick, open, and throw something, and the Diagnostic Evaluation of Language Variation (DELV) Part 1 Language Variation Status (Seymour et al., 2003). All the children met the following inclusionary criteria: (1) primarily right-handed, defined as performing at least 3 out of 5 items using their right hand; (2) mainstream English speakers, defined as a mainstream English score of 7, 8, 9, or 11 for 5-, 6-, 7-, and 8-year olds, respectively, on the DELV; (3) no diagnosis of Attention Deficit Hyperactivity Disorder (ADHD), neurological disease, psychiatric disorder, learning or language disorder as reported in the developmental history questionnaire completed by their parents; (4) normal hearing and normal or corrected-to-normal vision as reported in the developmental history questionnaire completed by their parents.

Standardized testing was then administered to assess language ability, nonverbal IQ and phonological working memory. This included the Clinical Evaluation of Language Fundamentals (CELF-5, Wiig et al., 2013) and the Kaufman Brief Intelligence Test, Second Edition (KBIT-2, Kaufman and Kaufman, 2004). All children scored greater than or equal to a standard score of 80 (9th percentile) on KBIT-2 non-verbal scale subtest and on CELF-5 core language score were included in this current study. In addition, phonological working memory was measured by the Comprehensive Test of Phonological Processing (CTOPP-2, Wagner et al., 2013) phonological memory composite score, which included both memory for digits and nonword repetition subtests. We included both IQ standardized scores (as a general cognitive ability index) and the CTOPP-2 phonological memory composite scores in our data analysis as covariates of no interest because our phonological task in the scanner may involve multiple cognitive processes beyond the processes of interest (i.e. phonological representation and access).

2.2. Experimental procedure

2.2.1. Auditory phonological judgment task

Our auditory phonological judgement task was an event-related design (see Table 1). On each trial, children heard two sequentially presented auditory stimuli binaurally through earphones. There were four conditions of the pairs of stimuli: onset, rhyme, non-match, and noise (frequency modulated as the auditory perceptual condition). Participants were asked, “do the two words share the same sounds”. They were instructed to respond to all trials as quickly and accurately as possible with the right index finger for a yes response in the onset, rhyme

Table 1
Examples of the stimuli in the auditory phonological judgment task.

Condition	Response	Brief Explanation	Example
Onset	Yes	The two words start with the same sound	Coat – Cup
Rhyme	Yes	The two words rhyme	Wide – Ride
Non-match	No	The two words have no same sounds	Zip – Cone
Perceptual	Yes	Frequency modulated noise	“Sh – Sh”

conditions, and with the right middle finger for a no response in the non-match condition. For the perceptual condition, participants were asked to press the yes button with the right index finger whenever they heard the stimuli. A blue circle remained on the screen during the auditory stimuli presentation and it turned to yellow to remind the participants to respond. The duration of each word was between 500 and 700 ms (ms) followed by a brief period of silence, with the second word beginning 1000 ms after the onset of the first. The length of the stimuli was counterbalanced across conditions. The blue circle turned to yellow 1000 ms before the trial ended, indicating the need to make a response. The duration of the response interval was 1800 ms. There were 24 trials for each of the four conditions, divided into two runs. The four conditions were pseudo-randomized so there were no more than 5 same responses in a row. To aid in convolving the hemodynamic response, inter-trial intervals were jittered by randomly adding 0, 450 or 900 ms to each trial, in equal proportions for the first run. For the second run, jitters of 0, 375 or 750 ms were similarly added to the trials. Each run lasted about 3 min.

The auditory word conditions were designed according to the following standards. For the onset condition, the word pairs only shared the same initial phoneme (corresponding to one letter of its written form). For the rhyme condition, the word pairs shared the same vowel and final phoneme/cluster (corresponding to 2 to 3 letters at the end of its written form). For the non-match condition, there were no shared phonemes (or letters of its written form). All the words were monosyllabic. Every paired word had no semantic association based on the University of South Florida Free Association Norms (Nelson et al., 1998). There were no significant differences ($p > 0.1$) between conditions in phonotactic frequency (Vitevitch and Luce, 2004), word frequency (Balota et al., 2007), part of speech (Balota et al., 2007), and the phonological and orthographic consistency (Bolger et al., 2008). Neither irregular forms nor inflected forms of words were used. The auditory perceptual condition was made of frequency-modulated noise equated in length and amplitude with the word conditions.

Participants who scored within an acceptable accuracy range and had no response bias were included in our analysis. Acceptable accuracy was defined as the accuracy of the perceptual condition and the rhyme condition being greater than 50%. We only established a threshold for rhyme but not onset because onset awareness was very difficult for children aged 5–6 years old, with an average accuracy of 62% (as shown in Table 2). We preserved more data and variance by not thresholding based on onset accuracy. By thresholding the rhyme condition, which is easier for young children, we were confident that children understood our task

Table 2
Performance during the auditory phonological task and standardized reading test.

Test		T1 (Mean \pm SD) [range]	T2 (Mean \pm SD) [range]
Phonological task (Accuracies)	Onset	0.62 \pm 0.18 [0.25–0.96]	0.74 \pm 0.13 [0.42–0.96]
	Rhyme	0.78 \pm 0.16 [0.50–1.00]	0.89 \pm 0.10 [0.58–1.00]
	Non-Match	0.76 \pm 0.13 [0.42–0.96]	0.86 \pm 0.11 [0.54–1.00]
	Perceptual	0.92 \pm 0.09 [0.63–1.00]	0.94 \pm 0.08 [0.71–1.00]
	Raw scores of Reading Skill	28 \pm 11 [12–57]	47 \pm 8 [29–63]

in the scanner. However, the limitation of thresholding one but not the other condition is that we have a risk of truncating the true distribution of rhyme awareness. The lack of response bias was indicated by the accuracy difference between the rhyme condition and the non-match condition being lower than 40%. The accuracies for each condition during our auditory phonological task inside the scanner at both T1 and T2 are shown in Table 2.

In order to make sure the participants understood the task well and to acclimate them to the scanner environment, they were required to complete the same task with different stimuli in the mock scanner and a short practice just before the fMRI scanning session.

2.2.2. Reading ability

The raw scores of the Woodcock-Johnson III Test of Achievement Letter-Word Identification subtest (Woodcock et al., 2001), in which children were required read the visually presented letters and words out loud, served as our measure of reading ability. The mean and standard deviation of reading skills at both T1 and T2 are also shown in Table 2.

2.3. Data acquisition

Participants lay in the scanner with a response button box placed in their right hand. To keep participants focused on the task so that they would respond in time, visual stimuli were projected onto a screen, viewed via a mirror attached to the inside of the head coil. Participants wore earphones to hear the auditory stimuli and two ear pads were used to attenuate the scanner noise. The two phonological task runs were counterbalanced across participants.

Images were acquired using 3.0T Skyra Siemens scanner with a 64-channel head coil. The blood oxygen level dependent (BOLD) signal was measured using a susceptibility weighted single-shot echo planar imaging (EPI) method. Functional images were acquired with multiband EPI (TE = 30 ms, flip angle = 80, matrix size = 128 \times 128, FOV = 256 mm², slice thickness = 2 mm without gap, number of slices = 56, TR = 1250 ms, Multi-band accel. factor = 4, voxel size = 2 \times 2x2mm). A high resolution T1 weighted MPRAGE scan was acquired with the following scan parameters: TR = 1900 ms, TE = 2.34 ms, matrix size = 256 \times 256, field of view = 256 mm², slice thickness = 1 mm, number of slices = 192.

2.4. Data analysis

fMRI data was analyzed using Statistical Parametric Mapping (SPM12, <http://www.fil.ion.ucl.ac.uk/spm>). First, all functional images were realigned to their mean functional image across runs. The anatomical image was then segmented and warped to a pediatric tissue probability map template to get the transformation field. An anatomical brain mask was created by combining the segmentation products (i.e. grey, white and cerebrospinal fluid), and then applied to its original anatomical image to produce a skull-stripped anatomical image. After that, the mean functional image and all functional images were co-registered to the skull-stripped anatomical image. Then, all the functional images were normalized to a pediatric template by applying the transformation field to them. We created this pediatric tissue probability map template using CerebroMatic (Wilke et al., 2017), a tool that makes SPM12 compatible pediatric templates with user-defined age, gender, and magnetic field. We input the following information into CerebroMatic: the unified segmentation parameters described in Wilke et al. (2017), which were estimated from 1919 participants (downloaded from <https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/>) and user defined age as 5.5–8 years old with one-month interval, gender as two females and two males at each age interval and magnetic field strength as 3T, resulting in a sample of 124 to create the appropriate pediatric template for our research. After normalization, smoothing was applied to all the functional images with 6 mm isotropic Gaussian kernel.

To reduce movement effects on brain signal, Art-Repair (<http://cibers.stanford.edu/tools/human-brain-project/artrepair-software.html>) was used to identify outlier volumes, defined as those with volume-to-volume head movement exceeding 1.5 mm in any direction, head movement greater than 5 mm in any direction from the mean functional image across runs, or deviations of more than 4% from the mean global signal intensity. The outlier volumes were repaired by interpolation between the nearest non-outlier volumes. Subjects included in our study had no more than 10% of the volumes repaired in each run and no more than 6 consecutive volumes repaired in each run. Six motion parameters estimated in the realignment step were entered in the first level modeling as regressors and the repaired volumes were deweighted (Mazaika et al., 2009).

Statistical analyses at the first level were calculated using an event-related design with the four conditions (i.e. onset, rhyme, non-match and perceptual) as conditions of interest. A high pass filter with a cut-off of 128s and an SPM default mask threshold of 0.5 were applied. All experimental trials were included in the analysis. Word and noise pairs were treated as individual events for analysis and modeled using a canonical hemodynamic response function (HRF). Contrast maps were generated for onset > perceptual and rhyme > perceptual for each participant at first level analysis. Then we used one sample t-tests at the group level analysis to show brain regions that were activated during onset and rhyme processing (Table 3 and Fig. 1). Two anatomical masks were used to isolate our regions of interest. The posterior left STG was defined as the posterior half of STG with $y < -24$ (Hickok and Poeppel, 2000), while the dorsal left IFG was defined as the opercular part of the

left IFG (Boets et al., 2013; Ramus, 2014) by using the anatomical automatic labeling (AAL) atlas template from WFU PickAtlas toolbox (http://www.nitrc.org/projects/wfu_pickatlas). Because the AAL atlas was based on the adult brain, we warped the T1 structure of the AAL atlas to our pediatric T1 template using AFNI's 3dNwarp non-linear coregistration and then applied this transformation to the AAL atlas. In this way, the anatomical atlas masks were aligned with our pediatric T1 template.

Statistical significance for the group level analysis within the whole brain mask (174,428 voxels) was defined using Monte Carlo simulations using AFNI's 3dClustSim program (see <http://afni.nimh.nih.gov/>). 3dClustSim carries out a user-specified number of Monte Carlo simulations of random noise activations at a particular voxel-wise alpha level within a masked brain volume. Following the suggestions made by Eklund et al. (2016) article regarding the inflated statistical significance achieved using some packages, we used 3dFWHMx to calculate the smoothness of the data for every single participant, using a spatial autocorrelation function (ACF = 0.48, 4.59, 14.88), and then averaged those smoothness values across all participants. This average smoothness value was then entered into 3dClustSim to calculate the cluster size needed for significance for a given ROI. The threshold for the size of a significant cluster was 109 voxels within the whole brain mask at a voxel-wise threshold at $p < 0.001$ uncorrected and cluster-wise threshold at $p < 0.05$.

To examine the scaffolding hypothesis, the top 100 voxels showing maximal activation (regardless of significance) for the contrast of onset > perceptual at T1 were selected based on the t values of that contrast for every participant within the anatomical mask of the posterior left STG as the individualized regions of interest (ROIs). The overlap among participants' individualized ROI for onset > perceptual at T1 were plotted (see Fig. 2(A), upper left panel). Beta values associated with all conditions were then extracted from these individualized top 100-voxel ROIs by using Marsbar toolbox (<http://marsbar.sourceforge.net/tutorial/index.html>). Brain activations for onset > perceptual at T1 were calculated as the betas for the onset condition minus betas for the perceptual condition. We chose this brain activation extraction method because Tong et al. (2016) has found that selecting the top activated voxels within an ROI, as compared to other methods, is more powerful in finding group differences and Fedorenko et al. (2010) showed that defining ROIs functionally in each individual shows greater specificity. This approach has also been used in other recently published studies (e.g. Suárez-Pellicioni et al., 2018; Younger et al., 2019). After extracting the brain activation from the top-activated voxels, a hierarchical regression analysis was run in SPSS, with non-verbal IQ, phonological working memory, and reading skill at T1 entered into the model as covariates of no interest and brain activation of onset > perceptual at T1 entered as the covariate of interest. The dependent measure was reading skill at T2 (see Table 4). In this way, we examined whether the representational quality of phonemic awareness scaffolds later reading. The same analysis was done using the brain contrast of rhyme > perceptual to examine whether the representational quality of rhyme awareness scaffolds later reading. The overlap among participants' individualized ROI for rhyme > perceptual at T1 were also plotted (see Fig. 1(A) upper right panel).

To examine the refinement hypothesis, the top 100 voxels showing maximal activation (regardless of significance) for the contrast of onset > perceptual at T2 were selected based on the t values of that contrast for every participant within the anatomical mask of posterior left STG as the individualized regions of interest (ROIs). The overlap among participants' individualized ROI for onset > perceptual at T2 were plotted (see Fig. 2(A) lower left panel). Then beta values associated with all conditions were extracted from these individualized 100-voxel ROIs by using Marsbar toolbox (<http://marsbar.sourceforge.net/tutorial/index.html>). Brain activations for onset > perceptual at T1 and T2 were calculated as the betas for the onset condition minus betas for the perceptual condition. After that, a hierarchical regression analysis was run in SPSS, with non-verbal IQ, phonological working memory, brain activation of onset > perceptual at T1 entered into the model as

Table 3
Group level brain activation for onset and rhyme processing at both T1 and T2.

Contrast	Brain region	Brodman Area	Coordinate	Voxel	T	
Onset > Perceptual at T1	Left Superior Temporal Gyrus	22	-58 -14 8	2996	11.80	
	Right Superior Temporal Gyrus	22	66 -10 4	1706	11.06	
	Left Inferior Temporal Gyrus	20	-46 -50 -14	250	5.56	
	Left Inferior Frontal Gyrus	45	-52 38 4	828	5.42	
	Left Superior Temporal Pole	34	-24 8 -22	146	5.39	
	Left Cingulate Gyrus	32	-10 24 36	150	4.56	
	Left Precentral/Inferior Frontal Gyrus	44	-54 10 30	114	4.33	
	Rhyme > Perceptual at T1	Left Middle/ Superior Temporal Gyrus	22	-60 -12 -2	2138	11.20
		Right Superior Temporal Gyrus	22	66 -8 4	1530	9.33
		Left Inferior Frontal Gyrus	47	-44 26 2	213	5.03
Left Inferior Temporal Gyrus		37	-42 -36 -14	139	4.73	
Onset > Perceptual at T2		Left Superior Temporal Gyrus	22	-62 -10 2	1746	12.51
	Right Superior Temporal Gyrus	22	66 -4 -2	1388	12.25	
	Left Inferior Temporal Gyrus	20	-44 -46 -16	146	4.87	
	Rhyme > Perceptual at T2	Left Superior Temporal Gyrus	22	-60 -10 2	1638	13.40
		Right Superior Temporal Gyrus	22	66 -2 -2	1230	11.26
Left Inferior Frontal Gyrus		47	-36 32 -14	120	6.28	
Left Inferior Temporal Gyrus		37	-42 -40 -16	185	4.91	

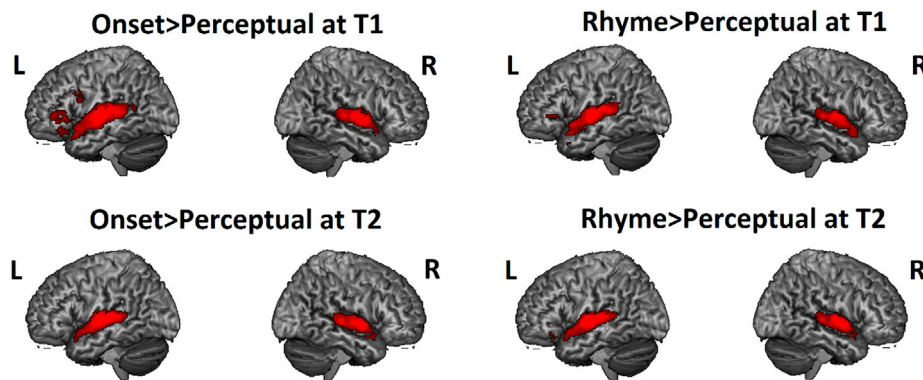


Fig. 1. Group level brain activation during onset and rhyme processing at T1 and T2. Group maps thresholded at voxel-wise $p < 0.001$ uncorrected and cluster-wise $p < 0.05$ within the whole brain mask. Clusters with size greater than 109 voxels are shown. L = left hemisphere; R = right hemisphere.

covariates of no interest and reading skill at T1 entered as the covariate of interest. The dependent measure was brain activation of onset > perceptual at T2 (see Table 5). In this way, we examined whether early reading skills refine the later representational quality of phonemic awareness. The same analysis was done using the contrast of rhyme > perceptual to examine whether early reading skill refines the later representational quality of rhyme awareness. The overlaps among participants' individualized ROI for rhyme > perceptual at T2 were also plotted (see Fig. 2(A) lower right panel).

The same procedure of analyses was conducted using the mask of the opercular part of left IFG (instead of the posterior left STG) to examine the scaffolding and refinement hypotheses between reading skill and phonological access at different grain sizes. The overlap of individualized ROI within the opercular part of IFG were also plotted (see Fig. 2(B)).

3. Results

3.1. Scaffolding hypothesis

3.1.1. Representational quality of phonological awareness in STG scaffolds later reading regardless of grain size

The regression analysis showed that brain activation in STG for both phoneme (onset > perceptual) and rhyme (rhyme > perceptual) significantly predicted reading skill at T2 after controlling for the reading skill at T1 and covariates of no interest such as nonverbal IQ and phonological working memory at T1 (see Table 4).

In order to compare which of the two predictors (phoneme or rhyme) explained more variance in reading gains over time, we examined if the brain activation for one predictor still significantly predicted reading skill at T2, while controlling for the other predictor. We found that the brain activation for onset > perceptual no longer significantly predicted reading skill at T2 [$\Delta R^2 = 0.019$, $p = .169$] after controlling for brain activation for rhyme > perceptual. We also did not find that the brain activation for rhyme > perceptual predicted reading skill at T2 [$\Delta R^2 = 0.001$, $p = .725$] after controlling for brain activation for onset > perceptual. These results show that there are no significant differences between the two predictors (i.e. phonemic and rhyme processing in STG) in predicting reading gains over time.

3.1.2. Phonological access in IFG does not scaffold later reading regardless of grain size

The regression analysis showed that the brain activation in IFG for phonemes (onset > perceptual) did not significantly predict reading skill at T2 after the effects of the covariates of no interest were accounted for (see Table 4). The regression analysis also showed that the brain activation in IFG for rhymes (rhyme > perceptual) did not significantly predict reading skill at T2, after the effects of the covariates of no interests were accounted for (see Table 4).

3.2. Refinement hypothesis

3.2.1. Reading skill does not refine later phonological representation in STG regardless of grain size

The regression analysis showed that reading skill did not significantly predict phonemic representation (onset > perceptual) in STG at T2 after the effects of the covariates of no interest were accounted for (see Table 5). The regression analysis also showed that reading skill did not significantly predict rhyme representation (rhyme > perceptual) in STG at T2 after the effects of the covariates of no interest were accounted for (see Table 5).

3.2.2. Reading skill does not refine later phonological access in IFG regardless of grain size

The regression analysis showed that reading skill did not significantly predict phonemic access (onset > perceptual) in IFG at T2 after the effects of the covariates of no interest were accounted for (see Table 5). The regression analysis also showed that reading skill did not significantly predict rhyme access (rhyme > perceptual) in IFG at T2 after the effects of the covariates of no interest were accounted for (see Table 5).

3.3. Behavioral results

Table 2 shows the behavioral performance of participants during the auditory phonological task inside the scanner and reading skill outside the scanner. In parallel with the analysis of brain data, the same regression analysis, using accuracy of the phonological task inside the scanner, was also conducted to examine the scaffolding and refinement hypotheses. We found that the accuracies for both the onset and rhyme conditions at T1 did not significantly predict reading skill at T2, after controlling for the reading skill at T1, nonverbal IQ and phonological working memory [onset: $\Delta R^2 = 0.019$, $p = .234$; rhyme: $\Delta R^2 = 0$, $p = .948$]. Moreover, we found that reading skill at T1 did not significantly predict accuracies of the onset or rhyme conditions at T2 after controlling for accuracies at T1 for the onset or rhyme conditions as well as nonverbal IQ and phonological working memory [onset: $\Delta R^2 = 0.083$, $p = .101$; rhyme: $\Delta R^2 = 0.013$, $p = .519$]. Overall, no evidence was found either for the scaffolding or for the refinement hypotheses in behavioral analysis.

We also analyzed our data by including both behavioral performance and brain activation during the phonological task in one hierarchical regression model. The results stayed the same as when we calculated the models separately for behavioral and brain data. The brain activation for both onset and rhyme in STG at T1 still significantly predicted reading skill at T2 after controlling for T1 reading skill and the behavioral performance in our phonological task at T1 [onset: $\Delta R^2 = 0.059$, $p = .027$; rhyme: $\Delta R^2 = 0.055$, $p = .038$], showing that brain measurement of phonological processing predicts unique variance in reading gains, above

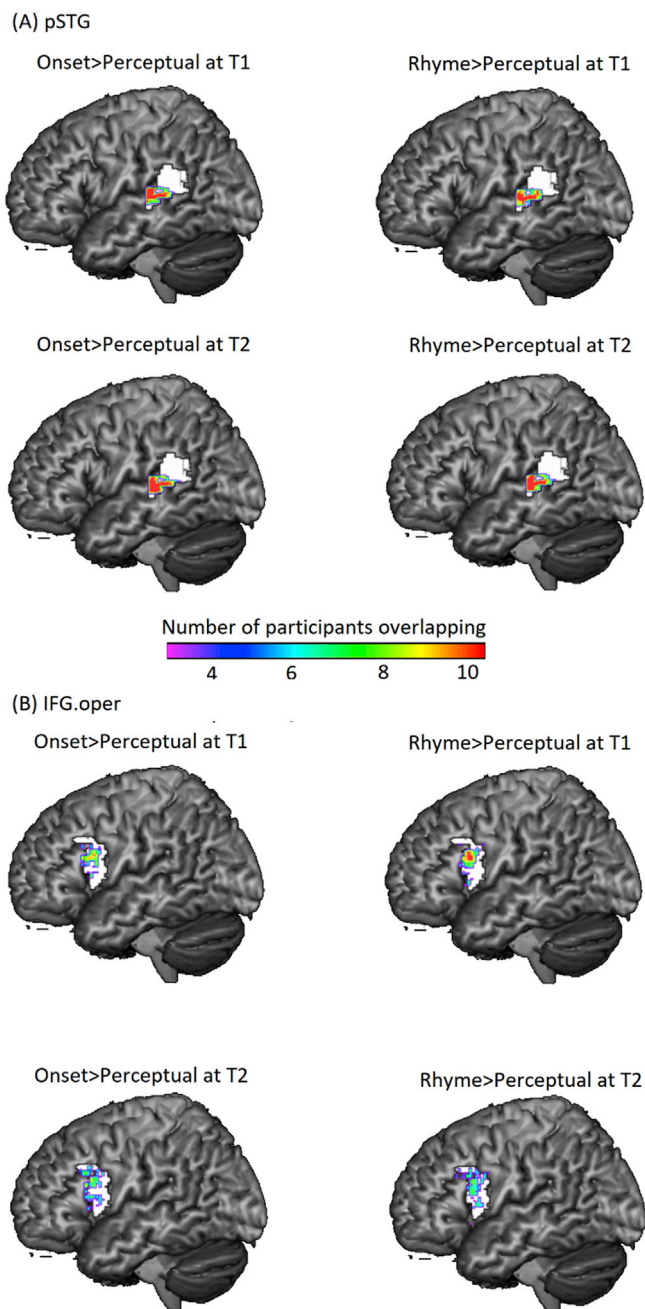


Fig. 2. Regions of interest in temporal and frontal cortex. (A) Overlap of individualized ROI in the posterior superior temporal gyrus (pSTG) (B) Overlap of individualized ROI in the opercular part of inferior frontal gyrus (IFG.oper). The two anatomical regions (i.e. pSTG and IFG.oper) are outlined in white. For each brain region, the upper left panel shows the number of participants showing overlap for onset > perceptual at T1. The upper right panel shows the number of participants showing overlap for rhyme > perceptual at T1. These ROIs defined at T1 were used in the examination of scaffolding hypothesis. The lower left panel shows the number of participants showing overlap for onset > perceptual at T2. The lower right panel shows the number of participants showing overlap for rhyme > perceptual at T2. These ROIs defined at T2 were used in the examination of refinement hypothesis.

and beyond behavioral measurement.

4. Discussion

The objective of the current study was to investigate the relationship between longitudinal changes in the neural basis of phonological

processing and in reading skill. Phonological processing was measured at the small grain size of phonemes (i.e. onset) and also at the large grain size of rhymes. We implemented a cross-lagged panel design by measuring brain and behavior when children were 5.5–6.5 years old and again when they were 7–8 years old. We investigated the relation of reading skill to both the posterior STG that has been implicated in representing phonology, and the dorsal IFG that has been implicated in access to phonology. We found that only the STG, regardless of grain size, predicted reading gains, suggesting that the quality of phonological representations but not the access to these representations scaffolds the development of reading in young children. We did not find evidence that reading predicted changes in either STG or IFG, and therefore, there was no evidence that reading refines the nature of phonological processing.

The scaffolding hypothesis was supported by the finding that that brain activation during phonological processing at T1 predicted reading gains. Our finding is consistent with the previous ERP study by Maurer et al. (2009), which found that the hemispheric specialization of the phoneme IMM in kindergarten predicted later reading skill; however, this study did not account for reading skill differences in kindergarten, so the effect could be autoregressive. Our finding is also consistent with longitudinal behavioral studies which showed that early phonological awareness in kindergarten predicted reading skills in the first few years of reading instruction (e.g. Roth et al., 2002; Schatschneider et al., 2004; Lervåg et al., 2009) and with training studies which showed that instruction in phonological awareness can improve reading in young children (Lundberg et al., 1988; Torgesen et al., 1992; Bentin and Leshem, 1993; Hatcher et al., 2004).

The prediction of reading gains by brain activation was specific to the STG, suggesting the representational quality of phonology is most important for reading acquisition. Our result is in alignment with previous fMRI studies on young children with a family risk of dyslexia (Raschle et al., 2012; Dębska et al., 2016). Both of these studies showed that at-risk children had alterations in STG during phonological processing, as compared to children without a family risk of dyslexia. Our finding is also consistent with the previous ERP study by Maurer et al. (2009), which found that the brain activation in STG during an auditory phoneme task in kindergarten predicted later reading skill in school years. Although the Maurer et al. (2009) study used a longitudinal design, they did not measure the reading skill in kindergarten to control for autoregressive effects. Our study, by using a cross-lagged panel design, more strongly suggests that early representational quality of phonology in STG leads to reading gains.

We interpreted increased activity in STG as better quality phonological representations because previous developmental studies have consistently shown that as children grow older and become better at phonological awareness, they show increased activation in STG during auditory tasks (e.g. Booth et al., 2004; Cao et al., 2011). Dyslexic children, who often have phonological deficits, tend to show less activity in STG as compared to typically developing children (see literature review in McCandliss and Noble, 2003). Some studies have also found increased activation of STG after an intensive phonological-based training (e.g. Simos et al., 2002). Together, these previous studies on children suggest that increased activation in STG is associated with better phonological awareness. Even though our measurement of activation might not be a direct index of the quality of phonological representations, increased activation in STG, according to previous literature, can be indicative of a better quality of phonological representations. Therefore, our finding that greater activation in STG was related with faster reading gains suggests that the representational quality of phonology is important for later reading acquisition.

In contrast to the finding in STG, our results in 5-to-7-year-old children do not support the idea that access to phonological representations through dorsal IFG at T1 is an important predictor of future reading skill. This is consistent with studies that have found differences in temporal, but not frontal, cortex in young children aged approximately 5–6 years old with a family risk of dyslexia (Raschle et al., 2012) and in adult

Table 4
The result of the hierarchical regression analyses examining the scaffolding hypothesis.

Step	Predictor	Dependent measure											
		Reading skill at T2			Reading skill at T1			Rhyme > Perceptual in STG at T2			Rhyme > Perceptual in IFG at T2		
		β	R ²	ΔR^2	β	R ²	ΔR^2	β	R ²	ΔR^2	β	R ²	ΔR^2
Model 1	Nonverbal IQ	-.055			-.055			-.055			-.055		
	Phonological working memory	.083			.083			.083			.083		
	Reading skill at T1	.805 ***	.653		.805 ***	.653		.805 ***	.653		.805 ***	.653	
Model 2	Nonverbal IQ	-.079			-.046			-.053			-.061		
	Phonological working memory	.091			.075			.039			.082		
	Reading skill at T1	.850 ***			.844 ***			.844 ***			.839 ***		
	Onset > Perceptual in STG at T1	.261 *	.719	.066									
	Rhyme > Perceptual in STG at T1				.225*	.702	.049						
	Onset > Perceptual in IFG at T1							.208	.694	.041			
	Rhyme > Perceptual in IFG at T1										.140	.672	.019

Note: *p < .05, **p < .01, ***p < .001.

Table 5
The result of the hierarchical regression analyses examining the refinement hypothesis.

Step	Predictor	Dependent measures											
		Onset > Perceptual STG at T2			Rhyme > Perceptual STG at T2			Onset > Perceptual IFG at T2			Rhyme > Perceptual IFG at T2		
		β	R ²	ΔR^2	β	R ²	ΔR^2	β	R ²	ΔR^2	β	R ²	ΔR^2
Model 1	Nonverbal IQ	-.170			-.101			-.015			-.028		
	Phonological working memory	-.003			-.089			-.208			-.240		
	Onset > Perceptual in STG at T1	.635***	.434										
	Rhyme > Perceptual in STG at T1				.795***	.656							
	Onset > Perceptual in IFG at T1							.633***	.416				
	Rhyme > Perceptual in IFG at T1										.461**	.267	
Model 2	Nonverbal IQ	-.189			-.140			.050			-.003		
	Phonological working memory	-.011			-.107			-.180			-.230		
	Onset > Perceptual in STG at T1	.642***											
	Rhyme > Perceptual in STG at T1				.812***								
	Onset > Perceptual in IFG at T1							.604***					
	Rhyme > Perceptual in IFG at T1										.445**		
	Reading skill at T1	.061	.437	.003	.131	.671	.015	-.203	.450	.034	-.080	.272	.005

Note: *p < .05, **p < .01, ***p < .001.

illiterates (Dehaene et al., 2010). Our result is also consistent with Vandermosten et al. (2019) which only examined the STG as region of interest and showed atypical phonological representations in 8-year-old children with a family risk of dyslexia, as compared to children without family risks. However, Dębska et al. (2016) found that children aged 6–7 years old with a family risk of dyslexia showed under-activation in both frontal and temporal cortex as compared to controls. Corina et al. (2001) and Kovelman et al. (2012) found reduced activation during auditory rhyming tasks in the frontal but not temporal cortex in 10–13-year-old and 7-13-year-old dyslexic children as compared to typically developing children. During an auditory phoneme discrimination task, Boets et al. (2013) found that even though dyslexic adults had intact phonological representation in the STG, they had difficulty accessing them through the dorsal IFG. One possibility is that phonological representations housed in STG are more important in reflecting individual differences of reading in young children while the access of phonology in IFG becomes a more salient predictor of individual differences of reading skill when children grow older. Future longitudinal studies with older children are needed to examine whether there is a developmental shift from the quality of phonological representations to access to those representations being more predictive of reading gains.

The activation of dorsal IFG is not only affected by the efficiency of accessing phonological information, but also influenced by working memory load (e.g. Fegen et al., 2015). In the current study, we included the phonological working memory as a control variable because our phonological awareness task included a working memory component. The two auditory words in our task were presented sequentially, which required children to maintain the first word in working memory in order

to judge whether the two words shared the same sound or not. We attempted to rule out this working memory explanation in order to more purely tap into the quality of phonological representations and access. Previous studies have suggested that phonological working memory and phonological awareness are two distinct components, which have unique contributions to reading (e.g. Gathercole and Baddeley, 1993). Phonological awareness is a prerequisite for distinguishing the component sounds in spoken words, whereas phonological memory is used to store sounds of spoken words so that they can be transferred to long-term memory. In terms of their relation to reading development, phonological awareness contributes to reading skill at an earlier age, whereas working memory only plays a role later (e.g. Nithart et al., 2011). We conducted an additional analysis by taking out phonological working memory from the regression model, and we found that the activation in IFG for onset processing was significantly predictive of reading gains. All the rest of findings remained the same. The fact that we only found activation in IFG for onset but not rhyme predicted reading is likely because it is more difficult to segment the onset and maintain it in working memory. This additional finding, even though it was not the main question of our study, suggests that phonological working memory also scaffolds reading gains in 5 to 7-year-old children.

One of the goals of our study was to determine whether phonological processing of small grain size (i.e. onset) or large grain size (i.e. rhyme) was predictive of reading gains. None of the previous neuroimaging studies on young children explored how different grain sizes play a role in reading skill. Previous studies used only small grain size tasks (i.e. onset judgement: Raschle et al., 2012; phoneme listening: Maurer et al., 2009) or large grain size tasks (i.e. rhyming: Dębska et al., 2016). We

showed that brain activation during phonological processing of both small and large grain size predicted reading gains, and there was no difference between the two. This finding is not consistent with previous behavioral studies on young children, which showed that phonemic awareness was more predictive of later reading skill as compared to rhyme awareness (Melby-Lervag et al., 2012; Muter et al., 1998; Hulme, 2002; Muter et al., 2004; Castles and Coltheart, 2004). However, some argue that reading skills develop from small grain size letter-phoneme mapping to larger grain size mapping, so it is possible that children at 7-8-year-old children in our study are relying more on rhymes to read words (Frith, 1985). This may be particularly important for English word reading because English is more consistent in orthography to phonology mapping at the rhyme level than at the phoneme level (Treiman et al., 1995). It is also possible that brain measures may be better at predicting reading gains than behavioral measures. In fact, we failed to find any scaffolding or refinement effects when we analyzed accuracy of the in-scanner task, probably because behavioral performance is the final product of several processing steps. For example, in our phonological task, the participants need to activate, maintain and compare the phonological representations. A single behavioral measure is a summation of all those cognitive steps. By looking at brain activation in STG associated with phonological representations, perhaps we have a more sensitive measure than behavioral performance. Our result is consistent with previous studies which showed that only brain but not behavioral measures were more reliable predictors of reading growth (e.g. Maurer et al., 2009; Hoeft et al., 2011). It is also possible that the lack of a scaffolding effect at the behavioral level in our study might be because our sample size was smaller than previous behavioral studies (e.g. Perfetti et al., 1987; Wagner et al., 1997).

As shown by the map of top activated voxels in Fig. 2, the small and large grain sizes of phonological processing activated a similar region of the planum temporale across subjects. This result suggests that different phonological grain sizes share a neural mechanism in 5- to 7- year-old children. Previous fMRI studies, using cross-modal tasks on children aged 8–13 years old (McNorgan et al., 2013a, 2013b) suggest that, as part of the phonological loop, the planum temporale is more sensitive to large-grain size representations. On the other hand, as word representations unfold over time, posterior superior temporal sulcus should be more sensitive to smaller grain-size representations. They argue, from a computational perspective, literacy experience drives the representations in posterior superior temporal cortex to increasingly multimodal in nature at the small grain size due to the experience of mapping graphemes to phonemes. We may have failed to find this distinction in phonological representations because young children do not have enough reading experience to develop these mappings. Instead, because planum temporale is often indicated as a region for analyzing physical features of speech sounds (e.g. Davis and Johnsrude, 2003; Liebenthal et al., 2005), our participants were likely relying on this region to make a phonological judgement. Future studies should examine whether there is a divergence for different grain sizes of phonology in temporal cortex as children become more experienced readers.

We did not find any evidence showing that early reading skill predicted later activation in the temporal or frontal cortex, arguing against the idea that learning to read refines the nature of phonological representations or access to them. This is consistent with the Wagner et al. (1997) study which found that word reading skill did not predict later phonological awareness. They proposed that substantial individual differences of reading ability occur too late, at around first or second grade, to have an impact on phonological awareness, which is relatively mature and stable at that time. However, in support of the refinement hypothesis, Wagner et al. (1997) did find that letter knowledge in kindergarten or first grade predicted later phonological awareness at grade 2 or 3. Letters roughly correspond to phonemes, thereby providing a useful concrete referent for children to learn phonemic awareness. The refinement effect of letter knowledge may be due to the fact that substantial individual differences of letter knowledge occur earlier in children when

phonological processing has not yet stabilized. Behavioral studies on even younger children (e.g. Burgess and Lonigan, 1998; Lerner and Lonigan, 2016) supported this idea by showing that letter knowledge at approximately 4-years-old was predictive of phonological awareness 6-months or one year later. Even though we showed that individual differences in word reading skill did not predict later activation during phonological processing, the refinement effects may occur earlier when children learn early reading skills such as letter knowledge.

Even though we did not find evidence for the refinement effect on phonological processing in STG and IFG, it is still likely that word reading influences phonological awareness tasks as is proposed by Ziegler and Goswami (2005) and as is evidenced in some behavioral studies (e.g. Perfetti et al., 1987). One possible reason for failing to find evidence for the refinement hypothesis is that the effect is too subtle to be captured by univariate analysis. Future studies could use a more sensitive analytical approach, such as multivoxel pattern analysis, which evaluates the activation pattern rather than the averaged activation level (increase or decrease). Another possibility is that the effect on phonological awareness occurs through enhancing the automatic activation of orthographic representations rather than refining phonological representations. In fact, Ziegler and Goswami (2005) propose that reading helps with phoneme awareness by mapping the acoustically inseparable sounds to letters. fMRI research also supports this idea. Dehaene et al. (2010) found that people who were illiterate did not activate their vOT during an auditory lexical decision task whereas people who were literate did. This suggests that reading acquisition impacts spoken language processing partly by automatically activating orthographic representations in vOT. Therefore, future studies should explore whether the effect of reading on phonological awareness tasks is due to enhancing connections with orthographic representations in vOT.

In conclusion, by using fMRI and a cross-lagged panel design, we found that phonological processing at small and large grain sizes predicted reading gains in 5- to 7-year-old children. Our results suggest that phonological representations in the temporal cortex are important in scaffolding reading. The implication of our study is that phonological awareness intervention programs should focus on improving young children's representational quality of both small and large grain sizes of phonology in order to maximize children's reading development. Future research on even younger children is needed to examine the refinement hypothesis, specifically looking at how early reading skills such as letter knowledge may refine phonological awareness in the brain as suggested by behavioral studies (e.g. Wagner et al., 1997; Burgess and Lonigan, 1998; Lerner and Lonigan, 2016). It would also be meaningful to investigate if phonological access by frontal cortex would play a more important role as children get older, as suggested by previous studies (Corina et al., 2001; Kovelman et al., 2012).

Author contribution section

Jin Wang: Conceptualization, Methodology, Formal analysis, Original draft;

Marc F. Joanisse: Review & Editing;

James Booth: Conceptualization, Methodology, Review & Editing; Supervision; Funding acquisition.

Acknowledgment

This research was supported by an NIH grant (R01 DC013274) to James R. Booth.

References

- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R., 2007. The English Lexicon project. *Behav. Res. Methods* 39, 445–459.
- Bentin, S., Leshem, H., 1993. On the interaction between phonological awareness and reading acquisition: it's a two-way Street. *Ann. Dyslexia* 43 (1), 125–148.

- Boets, B., Op de Beeck, H.P., Vandermosten, M., Scott, S.K., Gillebert, C.R., Mantini, D., Bulthe, J., Snaert, S., Wouters, J., Ghesquiere, P., 2013. Intact but less accessible phonetic representations in adults with dyslexia. *Science* 342 (6163), 1251–1254.
- Bolger, D.J., Hornickel, J., Cone, N.E., Burman, D.D., Booth, J.R., 2008. Neural correlates of orthographic and phonological consistency effects in children. *Hum. Brain Mapp.* 29 (12), 1416–1429.
- Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2004. Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* 16 (7), 1234–1249.
- Burgess, S.R., Lonigan, C.J., 1998. Bidirectional relations of phonological sensitivity and prereading abilities: evidence from a preschool sample. *J. Exp. Child Psychol.* 70 (2), 117–141.
- Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12 (4), 679–690.
- Burton, M.W., LoCasto, P.C., Krebs-Noble, D., Gullapalli, R.P., 2005. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage* 26 (3), 647–661.
- Cao, F., Khalid, K., Lee, R., Brennan, C., Yang, Y., Li, K., Bolger, D., Booth, J.R., 2011. Development of brain networks involved in spoken word processing of Mandarin Chinese. *Neuroimage* 57 (3), 750–759.
- Castles, A., Coltheart, M., 2004. Is there a causal link from phonological awareness to success in learning to read? *Cognition* 91 (1), 77–111.
- Corina, D.P., Richards, T.L., Serafini, S., Richards, A.L., Steury, K., Abbott, R.D., Echelard, D.R., Maravilla, K.R., Berninger, V.W., 2001. fMRI auditory language differences between dyslexic and able reading children. *Neuroreport* 12 (6), 1195–1201.
- Davis, M.H., Johnsruide, I.S., 2003. Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23 (8), 3423–3431.
- Debska, A., Luniewska, M., Chyl, K., Banaszkiewicz, A., Żelechowska, A., Wypych, M., Marchewka, A., Pugh, K.R., Jednoróg, K., 2016. Neural basis of phonological awareness in beginning readers with familial risk of dyslexia—results from shallow orthography. *Neuroimage* 132, 406–416.
- Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. How learning to read changes the cortical networks for vision and language. *Science* 330 (6009), 1359–1364.
- Démonet, J.-F., Price, C., Wise, R., Frackowiak, R.S.J., 1994. A PET study of cognitive strategies in normal subjects during language tasks: influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain* 117 (4), 671–682.
- Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* 113 (28), 7900–7905.
- Fedorenko, E., Hsieh, P.J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104 (2), 1177–1194.
- Fegen, D., Buchsbaum, B.R., D’Esposito, M., 2015. The effect of rehearsal rate and memory load on verbal working memory. *Neuroimage* 105, 120–131.
- Frith, U., 1985. Beneath the surface of developmental dyslexia. *Surf. Dyslexia* 32, 301–330.
- Gathercole, S.E., Baddeley, A.D., 1993. Phonological working memory: a critical building block for reading development and vocabulary acquisition? *Eur. J. Psychol. Educ.* 8 (3), 259.
- Hatcher, P.J., Hulme, C., Snowling, M.J., 2004. Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *JCPP (J. Child Psychol. Psychiatry)* 45 (2), 338–358.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4 (4), 131–138.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92 (1–2), 67–99.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8 (5), 393–402.
- Hoef, F., McCandliss, B.D., Black, J.M., Gantman, A., Zakerani, N., Hulme, C., Lyttinen, H., Whitfield-Gabrieli, S., Glover, G.H., Reiss, A.L., Gabrieli, J.D., 2011. Neural systems predicting long-term outcome in dyslexia. *Proc. Natl. Acad. Sci.* 108 (1), 361–366.
- Hulme, C., 2002. Phonemes, rimes, and the mechanisms of early reading development. *J. Exp. Child Psychol.* 82 (1), 58–64.
- Kaufman, A.S., Kaufman, N.L., 2004. Kaufman Brief Intelligence Test, second ed. Pearson, Bloomington, MN.
- Kearney, M.W., 2003. Cross-lagged panel analysis. In: *The SAGE Encyclopedia of Communication Research Methods* (SAGE, Pp. 1–6). Thousand, Oaks, CA.
- Klaus, J., Hartwigsen, G., 2019. Dissociating semantic and phonological contributions of the left inferior frontal gyrus to language production. *Hum. Brain Mapp.* 40, 3279–3287.
- Kovelman, I., Norton, E.S., Christodoulou, J.A., Gaab, N., Lieberman, D.A., Triantafyllou, C., Wolf, M., Whitfield-Gabrieli, Gabrieli, J.D., 2012. Brain basis of phonological awareness for spoken language in children and its disruption in dyslexia. *Cerebr. Cortex* 22 (4), 754–764.
- Leonard, M.K., Chang, E.F., 2014. Dynamic speech representations in the human temporal lobe. *Trends Cogn. Sci.* 18 (9), 472–479.
- Lerner, M.D., Lonigan, C.J., 2016. Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: a growth curve analysis of the relation between two code-related skills. *J. Exp. Child Psychol.* 144, 166–183.
- Lervåg, A., Bråten, I., Hulme, C., 2009. The cognitive and linguistic foundations of early reading development: a Norwegian latent variable longitudinal study. *Dev. Psychol.* 45 (3), 764–781.
- Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. *Cerebr. Cortex* 15 (10), 1621–1631.
- Lundberg, I., Frost, J., Petersen, O.-P., 1988. Effects of an extensive program for stimulating phonological awareness in preschool children. *Read. Res. Q.* 23 (3), 263–284.
- Maurer, U., Bucher, K., Brem, S., Benz, R., Kranz, F., Schulz, E., van der Mark, S., Steinhausen, H.-C., Brandeis, D., 2009. Neurophysiology in preschool improves behavioral prediction of reading ability throughout primary school. *Biol. Psychiatry* 66 (4), 341–348.
- Mazaika, P., Hoef, F., Glover, G.H., Reiss, A.L., 2009. Methods and software for fMRI analysis for clinical subjects. In: *Human Brain Mapping Conference*.
- McCandliss, B.D., Noble, K.G., 2003. The development of reading impairment: a cognitive neuroscience model. *Ment. Retard. Dev. Disabil. Res. Rev.* 9 (3), 196–205.
- McNorgan, C., Awati, N., Desroches, A.S., Booth, J.R., 2013a. Multimodal lexical processing in auditory cortex is literacy skill dependent. *Cerebr. Cortex* 24 (9), 2464–2475.
- McNorgan, C., Randazzo-Wagner, M., Booth, J.R., 2013b. Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Front. Hum. Neurosci.* 7, 388.
- Melby-Lervåg, M., Lyster, S.-A., Hulme, C., 2012. Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138 (2), 322.
- Muter, V., Hulme, C., Snowling, M., 1998. Segmentation, not rhyming, predicts early progress in learning to read. *J. Exp. Child Psychol.* 71, 3–27.
- Muter, V., Hulme, C., Snowling, M.J., Stevenson, J., 2004. Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Dev. Psychol.* 40 (5), 665–681.
- Nelson, D.L., McEvoy, C.L., Schreiber, T.A., 1998. *The University of South Florida word association, rhyme, and word fragment norms*. <http://www.usf.edu/FreeAssociation/>.
- Nithart, C., Demont, E., Metz-Lutz, M.N., Majerus, S., Poncellet, M., Leybaert, J., 2011. Early contribution of phonological awareness and later influence of phonological memory throughout reading acquisition. *J. Res. Read.* 34 (3), 346–363.
- Okada, K., Matchin, W., Hickok, G., 2018. Phonological feature repetition suppression in the left inferior frontal gyrus. *J. Cogn. Neurosci.* 30 (10), 1549–1557.
- Perfetti, C.A., Beck, I., Bell, L.C., Hughes, C., 1987. Phonemic knowledge and learning to read are reciprocal: a longitudinal study of first grade children. *Merrill-Palmer Q.* 283–319.
- Peterson, R.L., Pennington, B.F., 2015. Developmental dyslexia. *Annu. Rev. Clin. Psychol.* 11, 283–307.
- Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62 (2), 816–847.
- Ramus, F., 2014. Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends Cogn. Sci.* 18 (6), 274–275.
- Raschle, N.M., Zuk, J., Gaab, N., 2012. Functional characteristics of developmental dyslexia in left-hemispheric posterior brain regions predate reading onset. *Proc. Natl. Acad. Sci.* 109 (6), 2156–2161.
- Roth, F.P., Speece, D.L., Cooper, D.H., 2002. A longitudinal analysis of the connection between oral language and early reading. *J. Educ. Res.* 95 (5), 259–272.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Huber, W., 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci.* 105 (46), 18035–18040.
- Schatschneider, C., Fletcher, J.M., Francis, D.J., Carlson, C.D., Foorman, B.R., 2004. Kindergarten prediction of reading skills: a longitudinal comparative analysis. *J. Educ. Psychol.* 96 (2), 265–282.
- Seymour, H.N., Roper, T., Villiers, De, J.G., 2003. DELV: Diagnostic Evaluation of Language Variation: Screening Test. *PsychCorp*.
- Simos, P.G., Fletcher, J.M., Bergman, E., Breier, J.I., Foorman, B.R., Castillo, E.M., Davis, R.N., Fitzgerald, M., Papanicolaou, A.C., 2002. Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology* 58 (8), 1203–1213.
- Suárez-Pellicioni, M., Prado, J., Booth, J.R., 2018. Lack of improvement in multiplication is associated with reverting from verbal retrieval to numerical operations. *Neuroimage* 183, 859–871.
- Tong, Y., Chen, Q., Nichols, T.E., Rasetti, R., Callicott, J.H., Berman, K.F., Weinberger, D.R., Mattay, V.S., 2016. Seeking optimal region-of-interest (ROI) single-value summary measures for fMRI studies in imaging genetics. *PLoS One* 11 (3), e0151391.
- Torgesen, J.K., Morgan, S.T., Davis, C., 1992. Effects of two types of phonological awareness training on word learning in kindergarten children. *J. Educ. Psychol.* 84 (3), 364.
- Treiman, R., Zukowski, A., 1991. Levels of phonological awareness. Phonological processes in literacy: a tribute to Isabelle Y. Liberman, pp. 67–83.
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., Richmond-Welty, E.D., 1995. The special role of rimes in the description, use, and acquisition of English orthography. *J. Exp. Psychol. Gen.* 124 (2), 107–136.
- Vandermosten, M., Correia, J., Vanderauwera, J., Wouters, J., Ghesquiere, P., Bonte, M., 2019. Brain activity patterns of phonemic representations are atypical in beginning readers with family risk for dyslexia. *Dev. Sci.* <https://doi.org/10.1111/desc.12857>.
- Vitevitch, M.S., Luce, P.A., 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instrum. Comput.* 36, 481–487.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., Hecht, S.A., Barker, T.A., Burgess, S.R., Donahue, J., Garon, T., 1997. Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Dev. Psychol.* 33 (3), 468–479.

- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., Pearson, N.A., 2013. *Comprehensive Test of Phonological Processing—Second Edition*.
- Wang, J., Joannis, M.F., Booth, J.R., 2018. Reading skill related to left ventral occipitotemporal cortex during a phonological awareness task in 5–6-year old children. *Dev. Cogn. Neurosci.* 30, 116–122.
- Wiig, E.H., Semel, E., Secord, W.A., 2013. *Clinical Evaluation Language Fundamentals (CELF-5)*, fifth ed. Pearson, San Antonio.
- Wilke, M., Altay, M., Holland, S.K., CMIND Authorship Consortium, 2017. CerebroMatic: a versatile toolbox for spline-based MRI template creation. *Front. Comput. Neurosci.* 11, 5.
- Woodcock, R.W., McGrew, K.S., Mather, N., 2001. *Woodcock-Johnson III*. Itasca, IL. Riverside Publishing.
- Xie, X., Myers, E., 2018. Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: a comparison of clear and conversational speech. *J. Cogn. Neurosci.* 30 (3), 267–280.
- Younger, J.W., Lee, K.W., Demir-Lira, O.E., Booth, J.R., 2019. Brain lateralization of phonological awareness varies by maternal education. *Dev. Sci.* 22, e12807.
- Ziegler, J.C., Goswami, U., 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131 (1), 3–29.