Behavioral/Cognitive

**Individual Differences in Reading Skill Are Related to Trial-by-Trial Neural Activation Variability in the Reading Network**

Jeffrey G. Malins,1,2 Kenneth R. Pugh,1,3,4,5 Bonnie Buis,1 Stephen J. Frost,1 Fumiko Hoeft,1,6,7,8 Nicole Landi,1,3 W. Einar Mencl,1,4 Anish Kurian,1,3 Ryan Staples,1 Peter J. Molfese,1,3,9 Rose Sevcik,10 and Robin Morris10

1Haskins Laboratories, New Haven, Connecticut 06511, 2Department of Pediatrics, Yale University School of Medicine, New Haven, Connecticut 06520, 3Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut 06269, 4Department of Linguistics, Yale University, New Haven, Connecticut 06511, 5Department of Diagnostic Radiology, Yale University School of Medicine, New Haven, Connecticut 06520, 6Department of Psychiatry and Weill Institute for Neurosciences, University of California, San Francisco, California 94143, 7Dyslexia Center, University of California, San Francisco, California 94143, 8UC6-Stanford Precision Learning Center Neurosciences, San Francisco, California 94143, 9Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institutes of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, and 10Department of Psychology, Georgia State University, Atlanta, Georgia 30303

Recent work has suggested that variability in levels of neural activation may be related to behavioral and cognitive performance across a number of domains and may offer information that is not captured by more traditional measures that use the average level of brain activation. We examined the relationship between reading skill in school-aged children and neural activation variability during a functional MRI reading task after taking into account average levels of activity. The reading task involved matching printed and spoken words to pictures of items. Single trial activation estimates were used to calculate the mean and standard deviation of children’s responses to print and speech stimuli; multiple regression analyses evaluated the relationship between reading skill and trial-by-trial activation variability. The reliability of observed findings from the discovery sample (n = 44; ages 8–11; 18 female) was then confirmed in an independent sample of children (n = 32; ages 8–11; 14 female). Across the two samples, reading skill was positively related to trial-by-trial variability in the activation response to print in the left inferior frontal gyrus pars triangularis. This relationship held even when accounting for mean levels of activation. This finding suggests that intrasubject variability in trial-by-trial fMRI activation responses to printed words accounts for individual differences in human reading ability that are not fully captured by traditional mean levels of brain activity. Furthermore, this positive relationship between trial-by-trial activation variability and reading skill may provide evidence that neural variability plays a beneficial role during early reading development.

Key words: β series; BOLD variability; event-related fMRI; individual differences; reading disability; trial-by-trial variability

**Significance Statement**

Recent work has suggested that neural activation variability, or moment-to-moment changes in the engagement of brain regions, is related to individual differences in behavioral and cognitive performance across multiple domains. However, differences in neural activation variability have not yet been evaluated in relation to reading skill. In the current study, we analyzed data from two independent groups of children who performed an fMRI task involving reading and listening to words. Across both samples, reading skill was positively related to trial-by-trial variability in activation to print stimuli in the left inferior frontal gyrus pars triangularis, even when accounting for the more conventional measure of mean levels of brain activity. This finding suggests that neural variability could be beneficial in developing readers.

**Introduction**

A growing body of neuroimaging research has linked reading skill to variation in structural and functional circuitry in the brain (Norton et al., 2015). Investigations concerning the functional neuroanatomy of reading have focused on mean levels of activa-
tion across trials while children are engaged in different reading tasks. However, an emerging literature suggests that mean differences in activation and connectivity reflect only part of the complex neural foundation of reading ability. Recent studies have linked reading skill to the stability of neural responses to speech sounds (Hornickel and Kraus, 2013). In addition, animal work has shown that expression of the rat homolog of the dyslexia susceptibility gene KIAA0319 is linked to increased trial-by-trial variability in speech sound responses (Centanni et al., 2014a,b).

Together, these studies have helped to support the neural noise hypothesis of reading disability, which postulates that levels of neural noise can influence timing mechanisms that affect signal variability and thereby affect reading performance (Hancock et al., 2017).

This previous work leads to an expectation that reading skill in children is related to within-subject measures of neural activation variability. However, to date, neural activation variability has been evaluated with respect to reading skill in children only by examining brainstem electrophysiological responses to speech sounds (Hornickel and Kraus, 2013). In the current study, we instead asked children to perform a task involving word reading and examined trialwise variability in cortical activation using fMRI, a technique that has been used successfully to examine the relationship between neural activation variability and behavioral performance in multiple domains outside of reading (Garrett et al., 2013).

Across these other domains, there exists some debate concerning whether increased variability in the blood oxygen level-dependent (BOLD) signal confers a positive or negative impact on behavior. The directionality of the effect appears in part to be related to the extent to which a task entails cognitive versus sensory processing. For example, increased BOLD signal variability has been associated with faster and more consistent reaction times (RTs) in younger versus older adults during cognitive tasks including attentional cueing and delayed match-to-sample (Garrett et al., 2011), whereas in a study examining cognitive flexibility and stability (Armbuster-Genc et al., 2016), the direction of the relationship between BOLD signal variability and cognitive performance has been characterized as positive or negative depending on the task. Conversely, for sensory processing, increased BOLD signal variability has been associated with increased behavioral variability in older compared with younger adults during audiovisual speech perception (Baum and Beauchamp, 2014) and has been considered maladaptive in adults with autism, who showed greater trial-to-trial variability compared with matched controls in primary sensory areas during a low-level sensory task (Haigh et al., 2016). Given that reading involves both sensory and cognitive components, the direction of the relationship between BOLD signal variability and reading skill therefore remains an open question.

In the current study, we addressed the following novel questions: (1) Does trial-by-trial neural activation variability account for variance in reading skill in children above and beyond differences in mean activation? (2) If so, what is the direction of the relationship between neural activation variability and reading skill? To address these questions, we first conducted analyses on fMRI data from a discovery sample of children who performed a task in which they judged whether printed or spoken words matched pictures of items. We then confirmed whether observed effects held in a separate, independent sample of children. Analyses focused on using single trial β estimates to quantify mean activation across trials, as well as trial-by-trial variability, in the evoked response to print and speech within regions of the reading network. These mean and variability measures were then entered into multiple regression models characterizing the manner in which trial-by-trial activation variability is associated with reading skill after accounting for mean task-related activation as well as predictors of noninterest such as subject age.

Materials and Methods

Discovery sample

Participants. Children were selected from a larger study examining response to intervention for reading disability; the data presented here correspond to baseline scans before the onset of any intervention. Of this larger sample of 82 children, 44 were selected who met the following inclusion criteria: in third or fourth grade (71/82; the other 11 participants belonged to a cohort of seventh and eighth graders who participated in the larger study), an average Euclidean movement of 0.25 mm or less (58/71), and at least 70% accuracy in each of the auditory and visual mismatch conditions (44/58). Euclidean movement was calculated per volume by first computing point-to-point change for each the six motion parameters (i.e., three translation and three rotation) and then taking the square root of the sum of squares of these measures; average Euclidean movement was calculated by taking the mean value of this measure across all volumes of data collection. The accuracy cutoffs were selected to have a sufficient number of correct trials per participant to calculate dependable standard deviation (SD) measurements. The motion cutoff was selected to increase power to detect effects related to differences in trial-by-trial variability that are unrelated to motion because we expected that intransubject variability would be affected by movement in the scanner (Lund et al., 2005). We acknowledge that the percentage of data lost is larger than comparable fMRI studies with pediatric populations; however, data quality criteria were particularly stringent for this investigation given our concerns regarding participant motion as well as the requirement of having a sufficient number of trials to calculate valid SD measurements.

All children completed a battery of standardized cognitive assessments (Table 1). These included assessments of single word reading, pseudo-word decoding, and passage comprehension from the Woodcock-Johnson III Tests of Achievement (WI-III; Woodcock et al., 2001); the Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn, 2007), which measures receptive vocabulary; the Comprehensive Test of Phono- logical Awareness (CTOPP-2; Wagner et al., 2013), which measures metalinguistic knowledge of the structure of speech, or phonological awareness, by assessing skills including phoneme elision, blending, and isolation; and the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011), which measures verbal and nonverbal intelligence. As can be observed in Table 1, the range of reading scores was very broad and some children in the sample would be considered typically developing, whereas others would be classified as having reading disability using traditional diagnostic criteria. However, we treated reading skill as a continuous dimension, consistent with recent views concerning the multifaceted nature of reading skill, as well as the pitfalls of grouping children into diagnostic categories using cutoff scores (Pennington et al., 2012; Brannon-Martin et al., 2013).

fMRI task. Functional volumes were acquired while participants completed a task in which they judged whether picture cues matched auditory and visual target words (Frost et al., 2009; Landi et al., 2013; Jasińska et
Table 1. Descriptive information concerning the two groups of children who performed the fMRI experiment

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Measure</th>
<th>Discovery sample (n = 44; 18 female)</th>
<th>Confimation sample (n = 32; 14 female)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>WI-III Letter Word ID—raw score</td>
<td>Single word reading</td>
<td>43.0</td>
<td>9.0</td>
</tr>
<tr>
<td>WI-III Letter Word ID—standard score</td>
<td>Single word reading</td>
<td>95.5</td>
<td>13.9</td>
</tr>
<tr>
<td>WASI Full-Scale IQ—2</td>
<td>Intelligence</td>
<td>99.6</td>
<td>15.8</td>
</tr>
<tr>
<td>CTPP phonological awareness–composite score</td>
<td>Phonological awareness</td>
<td>85.5</td>
<td>13.4</td>
</tr>
<tr>
<td>PPVT—standard score</td>
<td>Receptive vocabulary</td>
<td>101.3</td>
<td>16.9</td>
</tr>
</tbody>
</table>

WI-II — Woodcock-Johnson III Tests of Achievement; WASI — Wechsler Abbreviated Scale of Intelligence; CTPP — Comprehensive Test of Phonological Processing; PPVT, Peabody Picture Vocabulary Test. For the discovery sample, the following versions were used: WASI-II, CTPP-2, and PPVT-4. For the confirmation sample, the following versions were used: WASI-I, CTPP, and PPVT-III.

aFull-Scale IQ-4 was measured instead of Full-Scale IQ-2 for one participant in the discovery sample. Furthermore, Full-Scale IQ-2 is missing from one participant in the confirmation sample.

bCTPP scores are missing from two participants in the confirmation sample.

cPPVT scores are missing from one participant in the discovery sample and one participant in the confirmation sample.

Malins et al., Reading Skill is Related to fMRI BOLD Variability
J. Neurosci., March 21, 2018 • 38(12):2981–2989 • 2983

al., 2016; Preston et al., 2016). In this task, participants were presented with pictures of common items (e.g., “cake”) that remained on the screen for 40–65 s, corresponding to between seven and eight trials, before being replaced by another picture. This procedure encouraged participants to generate strong expectations of target items, thereby maximizing responses to mismatches, and also obviated the need to associate targets with a new picture on every single trial, which could have been overly taxing. Whereas each picture remained on the screen, participants were presented with target items in an event-related fashion; specifically, printed words appeared in a box below the picture (presented for 3000 ms in 40-point Arial font), or auditory words were presented via headphones. Importantly, in 1/6 of trials, the printed or spoken word matched the picture, whereas in the other 5/6 of trials, the printed or spoken word mismatched the picture. Participants were asked to indicate via button press whether the printed or spoken word matched the picture. In total, participants completed 25 trials in each of the auditory (spoken) and visual (print) mismatch conditions. A sample trial sequence is illustrated in Figure 1.

Figure 1. A sample trial sequence for the fMRI task.
sored from analysis; this occurred for an average of two trials in each of the auditory and visual mismatch conditions. In all other trials, outlier values were replaced with zeros and ignored when calculating average β values within regions of interest (the mean number of voxels with outlier values across all trials and participants was less than one in both the auditory and visual mismatch conditions in each of the ROIs detailed in the next section).

ROI selection. Given that mean activation was one of the predictors we aimed to include in the multiple regression models evaluating relationships with reading skill, we elected not to analyze mean activation at the whole brain level because this would have biased selection of ROIs. Moreover, we did not predict perfect concordance between areas in which the reading task resulted in overall levels of activation and areas in which the task resulted in increased levels of variance in activation. Therefore, we instead defined ROIs using a recent meta-analysis that took the results of 20 different imaging studies of reading in children and combined them to identify a set of coordinates which showed convergence across studies (Martin et al., 2015). Because this meta-analysis combined results across tasks examining different aspects of reading, our view was that, by using these coordinates, we were more likely to include regions that may show a relationship between neural activation variability and reading skill even if these regions do not appear in a map of mean activation for the current task; with that said, we acknowledge the limitation that the meta-analysis also used mean activation to define ROIs.

We created spheres with a radius of 6 mm (two voxels) centered on the Martin et al. (2015) coordinates for the following regions: left inferior frontal gyrus (IFG) pars opercularis, left IFG pars triangularis, left middle temporal gyrus, left superior temporal gyrus (STG), left superior parietal lobule, and left inferior temporal gyrus. In addition, we also included an ROI for left thalamus, given extant findings indicating that the thalamus contributes to the reading network (Galaburda et al., 1985; Pugh et al., 2013). The full set of ROIs selected for analysis are detailed in Table 3 and displayed in Figure 2. For reference, we overlaid these ROIs on a conjunction map that shows the extent of overlap in task-related activation between the discovery sample and the confirmation sample (further details below). This map was created by running a standard GLM for each participant with a single regressor per condition; groupwise evoked response maps across all task conditions were then generated for both the discovery and confirmation samples using the program 3dANOVA2 (corrected at a false discovery rate (FDR) < 0.01). The conjunction map was created using step functions (3dcalc) and adding together resultant maps.

Analysis of trialwise variability. For analysis of trialwise variability, we considered correct trials in the auditory and visual word mismatch conditions with RTs within the acceptable range. Our rationale for analyzing both print and speech trials was that, even though we specifically were interested in responses to printed words, the paradigm included spoken words and analyzing neural responses in this condition afforded us the ability to determine whether any potential relationships between neural activation variability and reading skill were present or were instead more general for language. Our rationale for analyzing only the mismatch conditions was that these were the predominant conditions in the experiment in terms of the overall number of trials; related to this, there were too few trials in the match conditions to calculate valid SD measurements. For each trial and each ROI, we calculated the average β weight across the voxels in the ROI (3dROImaps), ignoring outlier voxels that had been replaced with zero.

Next, we calculated the mean and SD of the β series in each ROI. Intrasubject SD measures were calculated by using leave-one-out jackknife estimation in version 2015.2 of the package “bootstrap” (Tibshirani and Leisch, 2015) in the R Project for Statistical Computing (RRID: SCR_001905) and taking the mean across estimates. Jackknife estimation was used to mitigate bias of SD estimates, especially given the relatively small number of measurements from which these SDs were derived (Efron, 1981). Then, in each ROI, we ran separate multiple regression models for the auditory and visual mismatch conditions with reading skill as the dependent variable, which was quantified using raw scores for Letter–Word Identification (LWID) from the WJ-III Tests of Achievement. We started with a full model that included the mean and SD of the β series in either the auditory or visual mismatch condition, as well as the following predictors of noninterest: age in months (McIntosh et al., 2008; Garrett et al., 2011); amount of subject motion, defined as the average point-to-point Euclidean movement across all volumes of data collection (Power et al., 2012); and the number of trials used to calculate the mean and SD of the β series (i.e., the number of correct trials following removal of trials that exceeded motion, outlier, or RT thresholds). Using the program dropterm in version 7.43–45 of the “MASS” package in R (Venables and Ripley, 2002), we removed, in a stepwise fashion, any of the three predictors of noninterest that did not account for significant variance in reading skill (an α criterion of 0.05 was used for backward selection; at each step, the predictor with the largest associated p-value was removed). Then, for the resulting models, change in the Akaike Information Criterion (AIC), change in the Bayesian Information Criterion (BIC), and change in adjusted R² were quantified for both the mean and SD of the β series by comparing final models with models in which each of these respective terms were removed.

Code Accessibility. Analysis scripts can be accessed via the Open Science Framework at: https://osf.io/6zrak/.

Confirmation sample
Participants. Children were selected for this analysis from a large dataset that has been the subject of other reports (Frost et al., 2009; Landi et al., 2013; Jastińska et al., 2016; Preston et al., 2016). From this larger sample of 122 children, we first selected participants whose average Euclidean movement was 0.25 mm or less (81/122). Next, because the distribution of reading ability in the larger sample differed from the discovery sample,
which was weighted toward the lower end of the reading skill distribution, we selected a subset of children who were matched to the discovery sample in age and raw single word reading scores (WJ-III LWID) using version 3.0.1 of the R package “MatchIt” (Ho et al., 2011). From this subset of children, we then selected those who attained at least 70% accuracy in each of the auditory and visual mismatch conditions, which resulted in 32 children in the confirmation sample (14 female). Assessment scores for the confirmation sample are listed in Table 1. As can be noted from the table, mean raw reading scores and mean age were not significantly different across the two samples (WJ-III LWID raw scores: 0.001; PPVT standard scores: 0.01), with the caveat that children in the discovery sample were tested using CTOPP (Wagner, Torgesen, & Rashotte, 1999) instead of CTTOPP. phonological awareness composite standard scores: t(55) = −2.26, p = 0.03; CTOPP phonological awareness composite standard scores: t(57) = −5.51, p < 0.001; PPVT standard scores: t(52) = −3.39, p = 0.001; WASI FSIQ-2 standard scores: t(50) = −3.15, p < 0.01), with the caveat that children in the confirmation sample were tested using CTTOPP (Wagner, Torgesen, & Rashotte, 1999) instead of CTTOPP-2, PPVT-III (Dunn & Dunn, 1997) instead of PPVT-4, and WASI-I (Wechsler, 1999) instead of WASI-II. This caveat notwithstanding, we would argue that such sample differences provide for increased generalizability of results.

fMRI task. Functional volumes were acquired while participants completed the same picture cue–target word identification task as the children in the discovery sample. However, the task in this sample included a larger number of conditions. More specifically, for both the auditory and visual modalities, mismatches were either real words or pseudowords; in addition, for the visual modality, some mismatches were either semantically related words or meaningless consonant strings. As a result of this different design, the match to mismatch ratio was 1:4 instead of 1:5; in addition, printed words were presented for a duration of 2000 ms and in 18-point Verdana font. To keep the confirmatory analyses as comparable as possible to the analyses used for the discovery sample, only the real word conditions in both modalities were considered.

Acquisition of MRI data. Images were acquired using a 1.5 T Siemens Sonata scanner with a one-channel head coil located at the Yale Magnetic Resonance Research Center in New Haven, Connecticut. T2*-weighted images were acquired in an axial-oblique orientation parallel to the intercommissural line (20 slices; 6 mm slice thickness; no gap) using single-shot echoplanar imaging (matrix size = 64 × 64; voxel size = 3.125 × 3.125 × 6 mm; FoV = 200 mm; TR = 2000 ms; TE = 50 ms; flip angle = 80°). To allow for stabilization of the magnetic field, the first four volumes within each run were discarded. Anatomical scans were collected in a sagittal orientation (MPRAGE; matrix size = 256 × 256; voxel size = 1 × 1 × 1 mm; FoV = 256 mm; TR = 2000 ms; TE = 3.65 ms; flip angle = 8°); these were acquired either following or between the functional runs. Participants completed between 7 and 10 functional runs each 3:46 (113 volumes) in length, which corresponded to up to 40 trials in each of the auditory and visual mismatch conditions (i.e., 4 trials in each condition in each run). Across all trials in the experiment, the time between trial onsets was jittered between 4 and 13 s.

Analysis pipeline. All analyses were conducted in the exact same fashion as they were for the discovery sample. Removal of volumes that exceeded the thresholds of 0.3 mm Euclidean movement and/or 10% outliers before the GLM resulted in an average loss of less than one trial in each of the auditory and visual mismatch conditions. After the GLM, removal of trials with outliers in >10% of voxels in the brain resulted in a further loss of three trials on average in the auditory mismatch condition and two trials on average in the visual mismatch condition; subsequently, the mean number of voxels with outlier values across all trials and participants was less than one in both the auditory and visual mismatch conditions in each ROI. Analyses focused solely on the ROIs and experimental conditions for which we observed effects in the discovery sample for either the mean or SD of the β series. When performing these confirmatory analyses, we opted to use a Bonferroni-corrected α threshold for significance of 0.0125, which was calculated by dividing 0.05 by 4, the total number of models tested.

Results

Behavioral performance for the in-scanner task for both samples is summarized in Table 2, along with data concerning average amount of movement in the scanner and the number of trials in the β series in each condition. Average values across participants for the mean and SD of the β series within each ROI are presented in Table 3; results for the multiple regression analysis are detailed in Table 4 for the discovery sample and Table 5 for the confirmation sample.

Across the two samples, we observed a positive relationship between reading skill and trial-by-trial neural activation variability for printed words in the left IFG pars triangularis (Fig. 3). In this region, trial-by-trial neural activation variability for printed words not only accounted for significant variance in reading skill above and beyond mean levels of activation, but actually accounted for a greater proportion of variance in reading skill than did mean activation. The relationship between reading skill and neural activation variability in this region appears to be fairly selective for print because we did not observe a significant relationship between reading skill and trial-by-trial variability in neural activation for spoken words in this region. For spoken words, the only relationship that we observed between reading skill and trial-by-trial neural activation variability was a negative association in the left STG in the discovery sample; however, this finding did not hold in the confirmation sample. These results also appear to be fairly selective for reading ability because we ran a secondary analysis with performance IQ (WASI Matrix Reasoning raw scores) as the dependent variable to assess whether neural activation variability was related to general cognitive ability. These analyses did not reveal any significant relationship between performance IQ and neural activation variability for print or speech in any region of interest. In addition, we also ran the analysis with single word reading skill as the dependent variable, this time including performance IQ as a covariate in the multiple regression models. The positive relationship between reading skill
and neural activation variability for print in the left IFG pars triangularis was still marginally significant in both the discovery and confirmation samples, even when accounting for individual differences in performance IQ (discovery sample: \( p = 0.055 \); confirmation sample: \( p = 0.017 \)).

The above results stand in contrast to those for mean activation, which only showed a significant relationship between reading skill and activation for printed and spoken words in the left thalamus in the discovery sample but not the confirmation sample. We should also note that the alternative predictors did not show systematic patterns across both samples; however, we did observe that average Euclidean motion and the number of trials within the \( \beta \) series accounted for significant variance in reading skill in a number of regions. To further test for the influence of subject motion and the number of trials in the \( \beta \) series on the observed results, we ran a secondary analysis in which we relaxed the subject inclusion criteria to a maximum of 0.40 mm average Euclidean motion and a minimum of 50% accuracy for both the auditory and visual mismatch conditions. This resulted in a sample size of 50 for the discovery sample and 43 for the confirmation sample. The relationship between reading skill and neural activation variability for print in the left IFG pars triangularis was still significant in the same direction in both samples (discovery sample: \( p = 0.012 \); confirmation sample: \( p = 0.002 \)).

### Discussion

Our aim was to assess the relationship between reading skill in school-aged children and trial-by-trial variability in fMRI activation for print or speech. This stemmed from recent advances concerning individual differences in neural response stability in school-aged children and trial-by-trial variability in fMRI activation.

### Table 3. Location of the regions of interest selected for the multiple regression analysis and average mean and standard deviation of the beta series across participants

<table>
<thead>
<tr>
<th>Region</th>
<th>Centre of mass (Talairach)</th>
<th>Discovery sample</th>
<th>Confirmation sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left inferior frontal gyrus pars opercularis</td>
<td>IFGop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left inferior frontal gyrus pars triangularis</td>
<td>IFGtr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left superior parietal lobule</td>
<td>SPL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left middle temporal gyrus</td>
<td>MTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left inferior temporal gyrus</td>
<td>ITG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left superior temporal gyrus</td>
<td>STG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left thalamus</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. Multiple regressions for the discovery sample. Each model quantifies the contribution of mean activation and trial-by-trial activation variability to WJ-III Letter Word Identification raw scores

<table>
<thead>
<tr>
<th>Region</th>
<th>Regressor</th>
<th>Print</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left IFGop</td>
<td>Mean activation</td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td>Activation variability</td>
<td>ΔAIC</td>
<td>BIC</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>ΔAIC</td>
<td>BIC</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5. Multiple regressions for the confirmation sample. Each model quantifies the contribution of mean activation and trial-by-trial activation variability to WJ-III Letter Word Identification raw scores

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>Region</th>
<th>Regressor</th>
<th>Print</th>
<th>Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left IFGtr</td>
<td>Mean activation</td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td>Activation variability</td>
<td>ΔAIC</td>
<td>BIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>ΔAIC</td>
<td>BIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Left thalamus</td>
<td>Mean activation</td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td>Activation variability</td>
<td>ΔAIC</td>
<td>BIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>ΔAIC</td>
<td>BIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Activation to Print in Left IFG pars triangularis

Figure 3. Partial correlation results for the multiple regression analyses performed for the left inferior frontal gyrus pars triangularis region of interest. The x-axis specifies adjusted trial-by-trial neural activation variability, which corresponds to residuals from a model where the standard deviation of the β series is the dependent variable and the mean of the β series, as well as any significant predictors of noninterest (i.e., age, average Euclidean motion, and/or the number of trials in the β series) are the regressors; the y-axis specifies adjusted reading scores, which are the size of the residuals from a model where letter–word identification raw scores are the dependent variable and the regressors are the mean of the β series as well as the same predictors of noninterest. The regression line is superimposed on the plot; the shaded region represents the 95% confidence interval.

Trial-by-trial activation variability versus mean activation

For each of two samples of children who performed an fMRI picture-word matching task, we entered intrasubject means and SDs of single trial β estimates for print and speech trials into multiple regression models predicting reading skill. We observed that in the left inferior frontal gyrus pars triangularis, the SD of the β series for printed words not only accounted for additional variance in reading skill that was not captured by the mean of the β series, but actually accounted for a greater proportion of variance in reading skill than did mean levels of activation. This effect held across the two samples despite differences in participants, scanners, and slight differences in trial context. Moreover, these effects were fairly selective for print because we did not observe a relationship between reading skill and activation variability for spoken words in this region; the only effect we observed for spoken words was a negative relationship between reading skill and activation variability in the left STG in the discovery sample that was not observed in the confirmation sample. In contrast to the findings for activation variability, for mean activation, the only relationships that we observed were positive correlations between reading skill and mean activation for both printed and spoken words in the left thalamus in the discovery sample, supporting previous studies documenting the important contributions of the thalamus to reading (Galaburda et al., 1985; Pugh et al., 2013); however, these relationships were not observed in the confirmation sample.

The left IFG has long been implicated as a critical part of the skilled reading network in adults, with more anterior and lateral subregions of IFG such as pars triangularis thought to be involved in semantic processing (Poldrack et al., 1999; Bookheimer, 2002). Moreover, activation of the left IFG, as well as connectivity between the left IFG and certain components of the reading network, has been associated with age-related increases over the course of reading development and this region has been linked to processes such as phonological segmentation and covert articulation (Schlaggar et al., 2002; Turkeltaub et al., 2003; Bitan et al., 2007). However, the current study is the first time that individual differences in reading ability in children have been associated with variability in fMRI activation for printed words in this region.

Direction of the relationship between reading skill and BOLD signal variability

In the left IFG pars triangularis, trial-by-trial variability in neural activation for print was positively related to reading skill. A positive relationship between BOLD signal variability and behavioral performance has been previously observed by Garrett et al. (2011), who found that increased levels of BOLD signal variability were associated with faster and more consistent behavioral performance in younger versus older adults across a range of cognitive tasks, as well as by Armbruster-Genc et al. (2016), who found that increased levels of BOLD signal variability were associated with
greater cognitive flexibility in adults, which manifested as reduced behavioral switching costs in a task-switching paradigm.

Based on EEG measures in children, McIntosh et al. (2008) suggest that increased neural variability reflects a greater dynamic range of cognitive states as well as a greater ability to transition between them, which perhaps translates to a greater ability to adapt to the environment. It is possible that the increased neural variability observed in the better readers in the current study could reflect greater neural adaptability; however, we interpret this with caution given that we assessed trial-by-trial variability as opposed to moment-to-moment variability at a finer within-trial timescale. Furthermore, increased neural variability may not always have a positive effect on behavioral performance. Armbuster-Genc et al. (2016) observed that one of the brain regions that showed a positive association between neural variability and cognitive flexibility—that is, the left inferior frontal junction—actually showed a negative association with cognitive stability, which manifested as more extensive behavioral costs for distractor inhibition. This dissociation in terms of directionality may depend on the level of hierarchical organization in which a brain region is situated as well as the extent to which a task is weighted toward sensory versus cognitive processing. For example, individuals with autism, a neurodevelopmental disorder that can co-occur with reading disability, have shown increased fMRI BOLD signal variability in primary sensory areas in response to low-level sensory stimulation and this finding has been used to explain why individuals with autism may experience difficulties in highly sensory environments (Haigh et al., 2016). This distinction between sensory and cognitive processing may help to reconcile the current results with the observation that low-level neural responses to speech sounds show greater variability in children with reading disability compared with typically developing children (Hornickel and Kraus, 2013). In addition, these findings, as well as future experiments that more directly tease apart sensory versus cognitive processing, may inform the neural noise hypothesis put forth by Hancock et al. (2017) by elucidating the conditions that promote greater versus lesser neural variability in developing readers as well as how these relationships pattern across different brain regions as a function of reading experience. The differentiation of the role of random neural “noise” versus systemic components that drive “dynamic range” or “adaptability” within such greater neural variability indices may be a critical conceptual and analytic challenge.

In the electrophysiological literature, He and Zempel (2013) have asserted that a certain amount of neural variability is beneficial, but if the level of variability is too high, brain activity could be scattered across too wide a range, which could be detrimental. Therefore, it is possible that the relationship between neural variability and behavioral performance is non-monotonic, and that we are only observing the ascending portion of an inverted U-shaped curve. Future studies could address this possibility by including children who are more severely impaired than the poorest readers in the present sample.

Possible mechanisms of neural variability

The increased neural variability observed in the more skilled readers in the current study could be the result of spontaneous fluctuations in the BOLD signal that are intrinsically generated in the brain and not attributable to specific inputs or outputs (Fox et al., 2005, 2007; Fox and Raichle, 2007). These spontaneous fluctuations may serve to coordinate neuronal activity between distal brain regions and may be the product of changes in the power of high-frequency electrical activity such as the gamma band (Leopold et al., 2003). These changes in gamma oscillation frequency may in turn be associated with differences in GABA concentrations and their resulting influence on the balance of neural excitation and inhibition (Muthukumaraswamy et al., 2009). It is also possible that the balance of excitation and inhibition could have been influenced by glutamatergic inputs, especially given recent findings documenting an association between glutamate concentrations and reading skill (Pugh et al., 2014), as well as similar support from animal models (Che et al., 2016). Based on these findings, future investigations should target the neural mechanisms of BOLD signal variability and their links to other neurobiological indices, including neural oscillations, neurochemistry, indices of neural noise, and neuroanatomical measures (Becker et al., 2011).

Conclusions and future directions

Overall, this investigation lends support to work advocating for the added value of evaluating inrasubject variability in brain signals compared with solely evaluating mean levels of neural activity (Faisal et al., 2008; Garrett et al., 2010, 2011, 2013; Pernet et al., 2011) and highlights the importance of considering individual difference dimensions beyond subject age as contributors to, or reflections of, individual differences in neural activation variability (McIntosh et al., 2008; Grady and Garrett, 2014). Furthermore, the current findings constitute a critical first step in considering the role of adaptability in developing brain systems involved in reading and motivate future investigations concerning the mechanistic link between neural activation variability, neural noise, and reading skill. Finally, from an applied standpoint, these results beg the tantalizing question of whether trial-by-trial activation variability in reading-related brain areas could serve as a useful biomarker for clinically relevant phenotypes such as response to intervention for reading disability.

References


Che A, Truong DT, Fitch RH, LoTurco JJ (2016) Mutation of the dyslexia-associated gene Ddc2 enhances glutamatergic synaptic transmission be-
CrossRef Medline
Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700–711. CrossRef Medline