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 Morris[10]

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

Received April 4, 2017; revised Dec. 30, 2017; accepted Jan. 9, 2018.

Author contributions: J.G.M., K.R.P., S.J.F., F.H., N.L., W.E.M., A.K., P.J.M., R. Sevcik, and R.M. designed research; J.G.M., K.R.P., B.B., S.J.F., F.H., N.L., W.E.M., A.K., R. Staples, P.J.M., and R.M. performed research; J.G.M., W.E.M., A.K., R. Staples, and P.J.M. contributed unpublished reagents/analytic tools; J.G.M., B.B., W.E.M., and R. Staples analyzed data; J.G.M., K.R.P., S.J.F., F.H., N.L., W.E.M., A.K., R. Sevcik, and R.M. wrote the paper.

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (Grant P01HD070837 to Georgia State University, Grants P01HD001994,

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 (Armbruster-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 intrasubject 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, pseudoword decoding, and passage comprehension from the Woodcock-Johnson III Tests of Achievement (WJ-III; Woodcock et al., 2001); the Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn, 2007), which measures receptive vocabulary; the Comprehensive Test of Phonological 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 multifactorial nature of reading skill, as well as the pitfalls of grouping children into diagnostic categories using cutoff scores (Pennington et al., 2012; Branum-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

R01HD086168, and R37HD090153 to Haskins Laboratories). F.H. was supported by the National Institutes of Health (Grants R01HD078351, R01HD067254, P50HD052120, and R01HD044073), University of California Office of the President (Grant MRP-17-454926), the Oak Foundation (Grant ORIO-16-012), and the National Science Foundation (Grant 1540854). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the children and their parents, as well as their schools and teachers, for participation in this study and Candice Goerger for her extensive role in recruitment and data collection at the Atlanta site.

The authors declare no competing financial interests.

Correspondence should be addressed to Jeffrey G. Malins, Haskins Laboratories, 300 George Street Suite 900, New Haven, CT, 06511. E-mail: jeffrey.malins@yale.edu.

DOI:10.1523/JNEUROSCI.0907-17.2018

 0270-6474/18/382982-09$15.00/0

Table 1. Descriptive information concerning the two groups of children who performed the fMRI experiment

		Discovery sample ($n = 44$; 18 female)			Confirmation sample ($n = 32$; 14 female)		
Assessment	Measure	Mean	SD	Range	Mean	SD	Range
—	Age	9.3	0.6	7.8–11.3	9.4	1.1	7.5–11.3
WJ-III Letter Word ID–raw score	Single word reading	43.0	9.0	31–61	46.8	12.1	23–69
WJ-III Letter Word ID–standard score	Single word reading	95.5	13.9	67–124	103.1	14.9	76–133
WASI Full-Scale IQ-2[a]	Intelligence	99.6	15.8	80–140	110.8	14.6	76–138
CTOPP phonological awareness– composite score[b]	Phonological awareness	85.5	13.4	65–112	104.4	15.2	67–145
PPVT–standard score[c]	Receptive vocabulary	103.3	16.9	73–135	114.9	12.5	84–135

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

[a]Full-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.

[b]CTOPP scores are missing from two participants in the confirmation sample.

[c]PPVT scores are missing from one participant in the discovery sample and one participant in the confirmation sample.

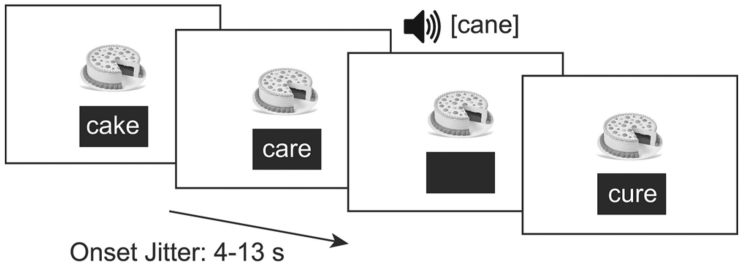

Figure 1. A sample trial sequence for the fMRI task.

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.

Acquisition of MRI data. Images were acquired using a 3 T Siemens Trio scanner with a 12-channel head coil located at the GSU/GaTech Center for Advanced Brain Imaging in Atlanta, Georgia. T2*-weighted images were acquired in an axial-oblique orientation parallel to the intercommissural line (32 slices; 4 mm slice thickness; no gap) using single-shot echo planar imaging (matrix size = 64 × 64; voxel size = 3.438 × 3.438 × 4 mm; FoV = 220 mm; TR = 2000 ms; TE = 30 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 the same orientation as the functional volumes (MPRAGE; matrix size = 256 × 256; voxel size = 1 × 1 × 1 mm; FoV = 256 mm; TR = 2530 ms; TE = 2.77 ms; flip angle = 7°); these were acquired either after or between the functional runs. In total, participants completed two runs of the functional task, which had a combined duration of 7 min 32 s (226 volumes). Across all trials in the experiment, the time between trial onsets was jittered between 4 and 13 s; trial order and intertrial intervals were optimized by an in-house MATLAB program that balanced intertrial intervals and null trials across conditions and minimized the variability of the measured response in Monte Carlo simulations.

Preprocessing. Data were analyzed using AFNI (Cox, 1996; RRID: SCR_005927). Functional images were preprocessed by first correcting for slice acquisition time (*3dTshift*). After this, functional images were aligned with anatomical images, corrected for motion using a six-parameter rigid-body transform (*3dvolreg*), and normalized to the Colin27 brain in Talairach space using an affine transform (*@auto_tlrc*). These three steps were combined into a single transform that also forced a 3 mm isotropic voxel size on the data. All images were then smoothed (*3dmerge*) using a Gaussian kernel with a full width at half maximum of 8 mm (i.e., twice the between-plane distance of 4 mm; Skudlarski et al., 1999) and data were scaled (*3dcalc*) so that each voxel's time series had a mean of 100 for each run. During this scaling step, values in excess of 200 were clipped; this is the default value for scaling in AFNI and was selected to retain the precision of scaled short values.

We elected to use the Talairach atlas for normalization because Burgund et al. (2002) have shown that, relative to the resolution of fMRI data, there are minimal anatomical differences between children ages 7 and 8 compared with adults. Given that the children in the current study were even older than the children in the Burgund et al. (2002) study (i.e., between 8 and 12 years of age), our view is that use of the Talairach atlas should allow for broader comparability between our study and others, including developmental investigations with adult samples.

GLM analysis. Single trial β estimates were obtained using a single General Linear Model (GLM) including nuisance regressors for the six motion parameters as well as a separate regressor for each trial (Rissman et al., 2004; Mumford et al., 2012; least-squares all, or LS-A). This model was specified using the *-stim_times_IM* flag for *3dDeconvolve* in AFNI. The hemodynamic response function was approximated using a gamma function. Because we were interested in intrinsic neural variability as opposed to variability related to individual differences in behavioral performance on the task, we included reaction times (RTs) for each trial as duration modulators in the GLM (Grinband et al., 2008; Yarkoni et al., 2009). For trials in which a participant either did not respond, responded with a RT <200 ms (i.e., invalid anticipation), or responded with an RT >1.5 times the interquartile range above the third quartile for a participant's distribution of RTs, overall mean RT for that participant was used as a duration modulator; however, these trials were not considered in further analyses.

β estimates corresponded to the amplitude assigned to each regressor in the GLM, and the set of β estimates across trials for a given voxel constituted that voxel's β series. When performing the GLM, any volume that exceeded the thresholds of 0.3 mm Euclidean movement and/or 10% outliers were censored from further analysis, resulting in an average loss of less than one trial in each of the auditory and visual mismatch conditions. It should be noted that this approach gave rise to some extreme outlier β values due to rare spikes that were still present in the data even after censoring these volumes. To handle these, outlier β values were identified for each participant using the program *3dToutcount* in AFNI, which flags outliers using an algorithm based on median absolute deviation. Trials with outliers in >10% of voxels in the brain were cen-

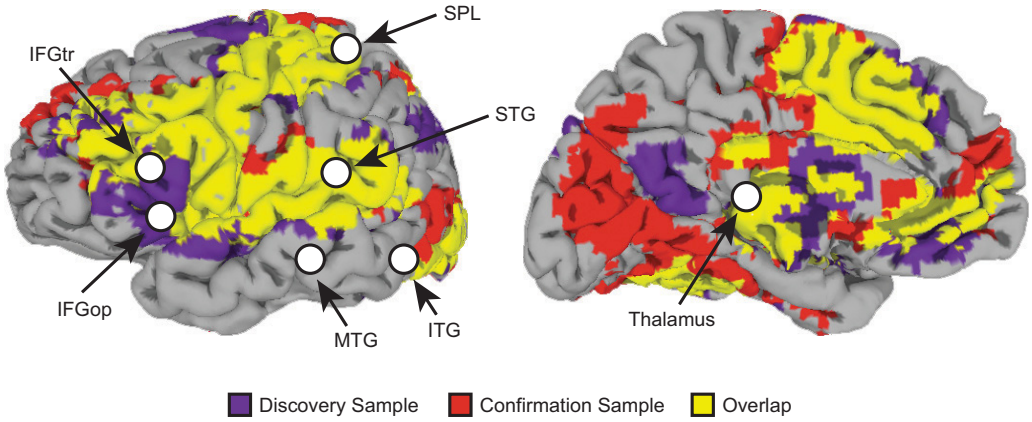

Figure 2. Lateral and medial views of the left hemisphere showing the location of each of the regions of interest (ROIs) tested in the multiple regression analyses. These ROIs, shown in white, are peak coordinates from the Martin et al. (2015) meta-analysis and are overlaid on a conjunction map that shows the overlap in task-related activation between the discovery and confirmation samples (each corrected at a false discovery rate (FDR) < 0.01; for more details on how this map was constructed, refer to the Materials and Methods section). IFGtr, Inferior frontal gyrus pars triangularis; IFGop, inferior frontal gyrus pars opercularis; SPL, superior parietal lobule; MTG, middle temporal gyrus; ITG, inferior temporal gyrus.

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 zeroes 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 were specifically 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 print specific 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 (*3dROIstats*), 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 R^2 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; Jasińska 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,

Table 2. MRI quality control parameters and performance for the in-scanner picture cue–target word identification task

	Discovery sample			Confirmation sample		
Measure	Mean	SD	Range	Mean	SD	Range
Average motion per brain volume (mm/TR)	0.12	0.05	0.04–0.23	0.13	0.05	0.05–0.25
Visual mismatch condition (print)						
Number of trials in the beta series	19.6	2.8	12–25	26.7	3.9	19–37
Percent accuracy	90.8	7.3	76–100	91.3	7.4	75–100
Mean reaction time for correct trials	1700	348	1159–3013	1479	355	989–2125
SD of reaction time for correct trials	506	203	181–1001	442	166	169–700
Auditory mismatch condition (speech)						
Number of trials in the beta series	21.0	2.1	17–24	27.3	3.7	20–36
Percent accuracy	96.5	4.7	80–100	93.0	4.9	81–100
Mean reaction time for correct trials	1698	296	1121–2335	1419	237	1072–1933
SD of reaction time for correct trials	456	194	135–1015	373	123	171–633

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: $t_{(55)} = -1.51$, $p = 0.14$; age: $t_{(47)} = -0.30$, $p = 0.77$), although standard single word reading scores, phonological awareness, vocabulary, and IQ were lower in the discovery sample compared with the confirmation sample (WJ-III LWID standard scores: $t_{(64)} = -2.26$, $p = 0.03$; CTOPP phonological awareness composite standard scores: $t_{(57)} = -5.51$, $p < 0.001$; PPVT standard scores: $t_{(72)} = -3.39$, $p = 0.001$; WASI FSIQ-2 standard scores: $t_{(68)} = -3.15$, $p < 0.01$), with the caveat that children in the confirmation sample were tested using CTOPP (Wagner, Torgesen, & Rashotte, 1999) instead of CTOPP-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 reran 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

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

					Discovery sample				Confirmation sample			
		Centre of mass (Talairach)[a]			Print		Speech		Print		Speech	
Region	Abbreviation	*x*	*y*	*z*	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Left inferior frontal gyrus pars opercularis	IFGop	−49	22	3	1.11	2.83	1.24	2.56	0.41	3.30	0.88	3.02
Left inferior frontal gyrus pars triangularis	IFGtr	−51	24	15	0.58	1.83	0.61	1.70	0.49	2.10	0.54	1.96
Left superior parietal lobule	SPL	−22	−48	50	0.16	1.13	0.12	1.05	0.03	1.18	0.07	1.16
Left middle temporal gyrus	MTG	−57	−25	−4	0.24	1.69	0.47	1.52	−0.17	1.84	0.37	1.79
Left inferior temporal gyrus	ITG	−52	−59	−9	0.66	2.28	0.07	2.44	0.55	3.18	0.19	2.97
Left superior temporal gyrus	STG	−55	−30	16	0.45	1.91	1.61	1.89	0.11	1.89	1.53	1.82
Left thalamus	—	−12	−25	10	0.59	1.29	0.45	1.23	0.27	1.15	0.45	1.13

Each region of interest was 891 mm^3 in size (33 voxels; 3 mm isotropic).

[a]LPI orientation.

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

		Print							Speech						
Region	Regressor	β	SE	ΔAIC	ΔBIC	ΔR^2_{adj}	*F*	*p*	β	SE	ΔAIC	ΔBIC	ΔR^2_{adj}	*F*	*p*
Left IFGop	Mean activation	−0.072	0.146	1.73	3.52	−0.015	0.246	0.623	−0.043	0.158	1.92	3.71	−0.021	0.075	0.786
	Activation variability	0.177	0.146	0.405	2.19	0.009	1.48	0.231	0.208	0.179	0.536	2.32	0.008	1.35	0.252
Left IFGtr	Mean activation	0.042	0.130	1.88	3.67	−0.015	0.106	0.746	−0.043	0.150	1.91	3.69	−0.021	0.083	0.775
	Activation variability	0.329	0.129	−4.66	−2.87	0.095	6.53	0.014	0.150	0.150	0.904	2.69	<0.001	1.01	0.321
Left SPL	Mean activation	−0.006	0.136	2.00	3.78	−0.019	0.002	0.964	−0.011	0.150	1.99	3.78	−0.023	0.006	0.939
	Activation variability	0.221	0.141	−0.644	1.14	0.028	2.48	0.123	0.086	0.166	1.71	3.49	−0.017	0.267	0.608
Left MTG	Mean activation	0.046	0.137	1.88	3.66	−0.017	0.111	0.741	0.057	0.153	1.85	3.63	−0.020	0.138	0.712
	Activation variability	0.211	0.137	−0.537	1.25	0.026	2.37	0.131	0.107	0.153	1.47	3.25	−0.012	0.490	0.488
Left ITG	Mean activation	−0.093	0.138	1.51	3.29	−0.011	0.449	0.507	0.103	0.171	1.61	3.40	−0.015	0.364	0.549
	Activation variability	0.102	0.141	1.42	3.21	−0.009	0.527	0.472	−0.187	0.171	0.739	2.52	0.005	1.19	0.281
Left STG	Mean activation	−0.039	0.140	1.92	3.70	−0.018	0.077	0.782	0.224	0.179	0.353	2.14	0.013	1.56	0.218
	Activation variability	0.102	0.141	1.42	3.21	−0.009	0.530	0.471	−0.399	0.179	−3.01	−1.23	0.088	4.95	0.032
Left thalamus	Mean activation	0.413	0.119	−9.52	−7.74	0.160	12.0	0.001	0.304	0.147	−2.39	−0.604	0.072	4.30	0.044
	Activation variability	0.160	0.120	0.095	1.88	0.011	1.77	0.191	−0.156	0.147	0.802	2.59	0.003	1.13	0.294

In all models, removal of the activation variability term did not affect the significance of the mean activation term.

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

Experimental condition	Region	Regressor	β	SE	ΔAIC	ΔBIC	ΔR^2_{adj}	*F*	*p*[a]
Print	Left IFGtr	Mean activation	−0.062	0.139	1.76	3.23	−0.014	0.200	0.658
		Activation variability	0.465	0.142	−8.70	−7.23	0.171	10.7	0.003
	Left thalamus	Mean activation	−0.003	0.141	2.00	3.47	−0.021	0.001	0.982
		Activation variability	−0.037	0.145	1.92	3.39	−0.019	0.065	0.801
Speech	Left STG	Mean activation	0.214	0.152	−0.189	1.28	0.022	1.98	0.170
		Activation variability	−0.179	0.152	0.461	1.93	0.009	1.38	0.250
	Left thalamus	Mean activation	0.178	0.156	0.545	2.01	0.007	1.30	0.263
		Activation variability	−0.121	0.156	1.32	2.79	−0.009	0.598	0.446

In all models, removal of the activation variability term did not affect the significance of the mean activation term.

[a]For this confirmatory analysis, we adopted a Bonferroni-corrected alpha threshold for significance of 0.0125 (0.05 divided by 4, the total number of models tested).

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 β 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 β 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

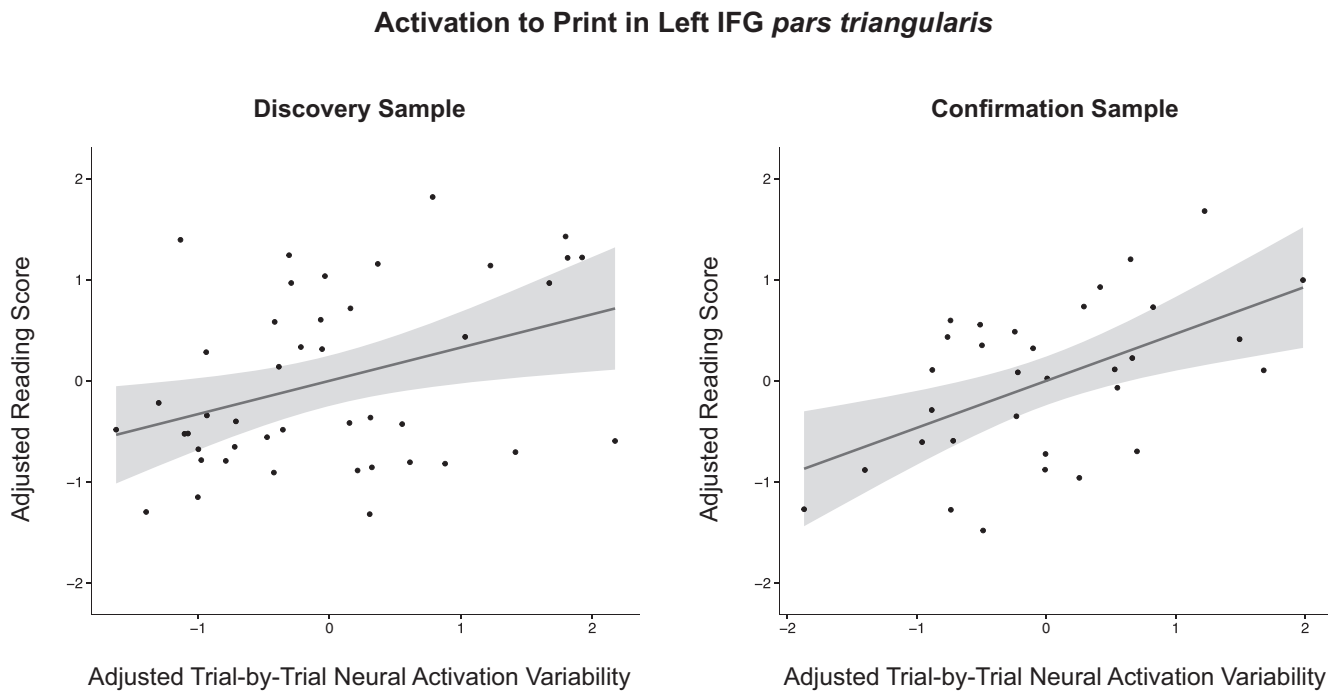

Figure 3. Partial correlation results for the multiple regression analyses performed for the left inferior frontal gyrus pars triangularis region of interest. The *x*-axis specifies adjusted trial-by-trial neural activation variability, which corresponds to residuals from a model where the standard deviation 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.

relation to reading skill (Hornickel and Kraus, 2013), as well as the potential impact of neural noise on the timing and systematic variability of processes important for reading (Hancock et al., 2017). Based on this previous work, we hypothesized that reading skill would be related to trial-by-trial variability in neural activation even after accounting for intrasubject differences in mean levels of task-related activation. However, the direction of this relationship remained an open question because domains outside of reading have shown different relationships between behavioral performance and variability in the fMRI BOLD response, whether measured from moment-to-moment within blocks (Garrett et al., 2011) or from trial-to-trial in experiments using event-related designs (Baum and Beauchamp, 2014; Armbruster-Genc et al., 2016; Haigh et al., 2016).

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. Armbruster-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 intrasubject 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

Armbruster-Genç DJ, Ueltzhöffer K, Fiebach CJ (2016) Brain signal variability differentially affects cognitive flexibility and cognitive stability. J Neurosci 36:3978–3987. CrossRef Medline

Baum SH, Beauchamp MS (2014) Greater BOLD variability in older compared with younger adults during audiovisual speech perception. PLoS One 9:e111121. CrossRef Medline

Becker R, Reinacher M, Freyer F, Villringer A, Ritter P (2011) How ongoing neuronal oscillations account for evoked fMRI variability. J Neurosci 31:11016–11027. CrossRef Medline

Bitan T, Cheon J, Lu D, Burman DD, Gitelman DR, Mesulam MM, Booth JR (2007) Developmental changes in activation and effective connectivity in phonological processing. Neuroimage 38:564–575. CrossRef Medline

Bookheimer S (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151–188. CrossRef Medline

Branum-Martin L, Fletcher JM, Stuebing KK (2013) Classification and identification of reading and math disabilities: the special case of comorbidity. Journal of Learning Disabilities 46:490–499. CrossRef Medline

Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, Petersen SE, Schlaggar BL (2002) The feasibility of a common stereotactic space for children and adults in fMRI studies of development. Neuroimage 17:184–200. CrossRef Medline

Centanni TM, Booker AB, Sloan AM, Chen F, Maher BJ, Carraway RS, Khodaparast N, Rennaker R, LoTurco JJ, Kilgard MP (2014) Knockdown of the dyslexia-associated gene *Kiaa0319* impairs temporal responses to speech stimuli in rat primary auditory cortex. Cereb Cortex 24:1753–1766. CrossRef Medline

Centanni TM, Chen F, Booker AM, Engineer CT, Sloan AM, Rennaker RL, LoTurco JJ, Kilgard MP (2014) Speech sound processing deficits and training-induced neural plasticity in rats with dyslexia gene knockdown. PLoS One 9:e98439. CrossRef Medline

Che A, Truong DT, Fitch RH, LoTurco JJ (2016) Mutation of the dyslexia-associated gene Dcdc2 enhances glutamatergic synaptic transmission be-

tween layer 4 neurons in mouse neocortex. Cereb Cortex 26:3705–3718. CrossRef Medline
Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 29:162–173. CrossRef Medline
Dunn LM, Dunn LM (1997) PPVT-III: Peabody Picture Vocabulary Test. Bloomington, MN: Pearson Assessments.
Dunn LM, Dunn DM (2007) PPVT-4: Peabody Picture Vocabulary Test. Bloomington, MN: Pearson Assessments.
Efron B (1981) Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 68:589–599. CrossRef
Faisal AA, Selen LP, Wolpert DM (2008) Noise in the nervous system. Nat Rev Neurosci 9:292–303. 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
Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673–9678. CrossRef Medline
Fox MD, Snyder AZ, Vincent JL, Raichle ME (2007) Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. Neuron 56:171–184. CrossRef Medline
Frost SJ, Landi N, Mencl WE, Sandak R, Fulbright RK, Tejada ET, Jacobsen L, Grigorenko EL, Constable RT, Pugh KR (2009) Phonological awareness predicts activation patterns for print and speech. Annals of Dyslexia 59: 78–97. CrossRef Medline
Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N (1985) Developmental dyslexia: four consecutive cases with cortical anomalies. Annals of Neurology 18:222–233. CrossRef
Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2010) Blood oxygen level-dependent signal variability is more than just noise. J Neurosci 30: 4914–4921. CrossRef Medline
Garrett DD, Kovacevic N, McIntosh AR, Grady CL (2011) The importance of being variable. J Neurosci 31:4496–4503. CrossRef Medline
Garrett DD, Samanez-Larkin GR, MacDonald SW, Lindenberger U, McIntosh AR, Grady CL (2013) Moment-to-moment brain signal variability: a next frontier in human brain mapping? Neurosci Biobehav Rev 37:610–624. CrossRef Medline
Grady CL, Garrett DD (2014) Understanding variability in the BOLD signal and why it matters for aging. Brain Imaging and Behavior 8:274–283. CrossRef Medline
Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008) Detection of time-varying signals in event-related fMRI designs. Neuroimage 43: 509–520. CrossRef Medline
Haigh SM, Gupta A, Barb SM, Glass SAF, Minshew NJ, Dinstein I, Heeger DJ, Eack SM, Behrmann M (2016) Differential sensory fMRI signatures in autism and schizophrenia: analysis of amplitude and trial-to-trial variability. Schizophrenia Research 175:12–19. CrossRef Medline
Hancock R, Pugh KR, Hoeft F (2017) Neural noise hypothesis of developmental dyslexia. Trends Cogn Sci 21:434–448. CrossRef Medline
He BJ, Zempel JM (2013) Average is optimal: an inverted-U relationship between trial-to-trial brain activity and behavioral performance. PLoS Computational Biology 9:e1003348.
Ho DE, Imai K, King G, Stuart EA (2011) MatchIt: nonparametric preprocessing for parametric causal inference. Journal of Statistical Software 42:1–28.
Hornickel J, Kraus N (2013) Unstable representation of sound: a biological marker of dyslexia. J Neurosci 33:3500–3504. CrossRef Medline
Jasińska KK, Molfese PJ, Kornilov SA, Mencl WE, Frost SJ, Lee M, Pugh KR, Grigorenko EL, Landi N (2016) The BDNF Val66Met polymorphism influences reading ability and patterns of neural activation in children. PLoS One 11:e0157449. CrossRef Medline
Landi N, Frost SJ, Mencl WE, Preston JL, Jacobsen LK, Lee M, Yrigollen C, Pugh KR, Grigorenko EL (2013) The COMT Val/Met polymorphism is associated with reading-related skills and consistent patterns of functional neural activation. Dev Sci 16:13–23. CrossRef Medline
Leopold DA, Murayama Y, Logothetis NK (2003) Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. Cereb Cortex 13:422–433. CrossRef Medline
Lund TE, Nørgaard MD, Rostrup E, Rowe JB, Paulson OB (2005) Motion or activity: their role in intra- and inter-subject variation in fMRI. Neuroimage 26:960–964. CrossRef Medline
Martin A, Schurz M, Kronbichler M, Richlan F (2015) Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. Hum Brain Mapp 36:1963–1981. CrossRef Medline
McIntosh AR, Kovacevic N, Itier RJ (2008) Increased brain signal variability accompanies lower behavioral variability in development. PLoS Comput Biol 4:e1000106. CrossRef Medline
Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643. CrossRef Medline
Muthukumaraswamy SD, Edden RA, Jones DK, Swettenham JB, Singh KD (2009) Resting GABA concentration predicts peak gamma frequency and fMRI amplitude in response to visual stimulation in humans. Proc Natl Acad Sci U S A 106:8356–8361. CrossRef Medline
Norton ES, Beach SD, Gabrieli JD (2015) Neurobiology of dyslexia. Curr Opin Neurobiol 30:73–78. CrossRef Medline
Pennington BF, Santerre-Lemmon L, Rosenberg J, MacDonald B, Boada R, Friend A, Leopold DR, Samuelsson S, Byrne B, Willcutt EG, Olson RK (2012) Individual prediction of dyslexia by single versus multiple deficit models. Journal of Abnormal Psychology 121:212–224. CrossRef Medline
Pernet CR, Sajda P, Rousselet GA (2011) Single-trial analyses: why bother? Front Psychol 2:322. Medline
Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage 10:15–35.
Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154. CrossRef Medline
Preston JL, Molfese PJ, Frost SJ, Mencl WE, Fulbright RK, Hoeft F, Pugh KR (2016) Print-speech convergence predicts future reading outcomes in early readers. Psychol Sci 1–1.
Pugh KR, Landi N, Preston JL, Mencl WE, Austin AC, Sibley D, Fulbright RK, Seidenberg MS, Grigorenko EL, Constable RT, Molfese P, Frost SJ (2013) The relationship between phonological and auditory processing and brain organization in beginning readers. Brain Lang 125:173–183. CrossRef Medline
Pugh KR, Frost SJ, Rothman DL, Hoeft F, Del Tufo SN, Mason GF, Molfese PJ, Mencl WE, Grigorenko EL, Landi N, Preston JL, Jacobsen L, Seidenberg MS, Fulbright RK (2014) Glutamate and choline levels predict individual differences in reading ability in emergent readers. J Neurosci 34:4082–4089. CrossRef Medline
Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752–763. CrossRef Medline
Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE (2002) Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science 296:1476–1479. CrossRef
Skudlarski P, Constable RT, Gore JC (1999) ROC analysis of statistical methods used in functional MRI: individual subjects. Neuroimage 9:311–329. CrossRef Medline
Tibshirani R, Leisch F (2015) Bootstrap: functions for the book "an introduction to the bootstrap". R package version 2015.2.
Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003) Development of neural mechanisms for reading. Nat Neurosci 6:767–773. CrossRef Medline
Venables WN, Ripley BD (2002) Modern applied statistics with S, Ed 4. New York, NY: Springer.
Wagner RK, Torgesen JK, Rashotte CA (1999) CTOPP: Comprehensive Test of Phonological Processing. Austin, TX: Pro-Ed.
Wagner RK, Torgesen JK, Rashotte CA, Pearson NA (2013) Comprehensive Test of Phonological Processing: CTOPP2. Austin, TX: Pro-Ed.
Wechsler D (1999) Wechsler Abbreviated Scale of Intelligence (WASI-I). San Antonio, TX: The Psychological Corporation.
Wechsler D, Hsiao-pin C (2011) WASI-II: Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II). San Antonio, TX: Pearson Assessments.
Woodcock RW, McGrew KS, Mather N (2001) Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing.
Yarkoni T, Barch DM, Gray JR, Conturo TE, Braver TS (2009) BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4:e4257. CrossRef Medline