Behavioral/Systems/Cognitive

# Prediction of Reading Skill Several Years Later Depends on Age and Brain Region: Implications for Developmental Models of Reading

**Chris McNorgan, Aubrey Alvarez, Annum Bhullar, Jessica Gayda, and James R. Booth**
Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

We investigated whether brain activity was predictive of future reading skill and, if so, how this brain–behavior correlation informs developmental models of reading. A longitudinal study followed 26 normally developing human children ranging in age from 9 to 15 years who were initially assessed for reading skill and performed a rhyming judgment task during functional magnetic resonance imaging. Patterns of brain activation in this task predicted changes between initial and a follow-up assessment of nonword reading skill administered up to 6 years later. Brain activity in areas typically active during imaging studies of reading was found to predict future nonword reading ability, but the predictive ability of these areas depended on age. Increased activity relative to peers in neural circuits associated with phonological recoding (i.e., inferior frontal gyrus and basal ganglia) was predictive of greater gains in reading fluency in younger children, whereas increased activity relative to peers in orthographic processing circuits (i.e., fusiform gyrus) was predictive of smaller gains in fluency for older children. Interpreted within the context of a connectionist model of reading, these results suggest that younger children who are more sensitive to higher-order phonological word characteristics (e.g., coarticulations) may make greater reading proficiency gains, whereas older children who focus more on whole-word orthographic representations may make smaller proficiency gains.

## Introduction

Because reading is both an acquired and later developing skill, understanding and predicting reading development contributes importantly to our understanding of general cognitive development. At its core, reading involves the transformation of visual (orthographic) into auditory (phonological) representations, and our investigation accordingly focuses only on orthographic and phonological processes. A large literature argues for a distributed network of functionally specialized brain regions contributing to different aspects of reading. Orthographic processing is generally attributed to the left fusiform gyrus (FG) (BA 37), which has been reliably shown to be responsive to visually presented words (McCandliss, 2003) in skilled readers. Left inferior parietal lobule (l IPL) (BA 40) is implicated in cross-modal mapping between orthographic and phonologic representations. On the basis of its sensitivity to conflicting orthography and phonology in children (Bitan et al., 2007b), left dorsal inferior frontal gyrus (dIFG) (BA 44/9) has been described as part of the anterior reading circuit associated with high-level phonological recoding during reading (Pugh et al., 2001) along with basal ganglia (BG),

which has been shown in skilled readers to modulate the phonological processing performed in the dIFG (Booth et al., 2007b). A dorsal reading circuit in posterior regions of left superior temporal gyrus (STG) has also been implicated as a second critical node in processing phonological word forms (Pugh et al., 2001; Booth, 2002). Because of their ties to normal and impaired reading development, we investigate how activity in these regions predicts future reading and how this informs models of reading development.

Previous studies have shown that brain activity predicts short- and long-term reading ability using measurements of event related potentials (ERPs) (Espy et al., 2004; Guttorm et al., 2005, 2010; Maurer et al., 2009; Lemons et al., 2010) taken during various auditory processing tasks. The limited spatial resolution of ERPs, however, precludes strong arguments concerning the role of specific brain areas. Functional magnetic resonance imaging (fMRI) has discriminated good from poor readers in a short-term longitudinal study (Hoeft et al., 2007), but the authors' data-driven approach made it difficult to relate their results to models of reading development. Although Hoeft et al. (2011) found that right inferior frontal activation predicted subsequent reading skill in dyslexics 2.5 years later, they failed to find associations in control children.

We use fMRI in a longitudinal study of reading development to answer two related questions. First, how do various regions implicated in language processing predict reading ability? Second, based on what we know about the role these regions play in language processing, how can this inform models of reading development? Many behavioral studies show the early importance

**Table 1. Means (and SDs) for subject characteristics for younger and older groups in the experiment**

|  | Younger ($n = 14$) | Older ($n = 12$) | $t_{(25)}$ |
|---|---|---|---|
| Age at $T_1$ (years) | 9.3 (0.8) | 13.5 (0.9) |  |
| Age at $T_2$ (years) | 14.2 (1.6) | 18.4 (1.2) |  |
| Follow-up interval (months) | 59 (4) | 60 (3) |  |
| WASI nonverbal IQ | 118.1 (15.4) | 105.6 (11.7) | 1.98 |
| WASI verbal IQ | 122.0 (14.4) | 107.8 (7.9) | 2.20* |
| $T_1$ TOWRE-SWE | 109.4 (2.8) | 105.3 (2.2) | 1.06 |
| $T_2$ TOWRE-SWE | 103.0 (3.3) | 101.7 (1.9) | 0.39 |
| $T_1$ TOWRE-PDE | 105.6 (3.4) | 96.6 (1.8) | 2.35* |
| $T_2$ TOWRE-PDE | 105.9 (4.0) | 98.1 (1.8) | 1.68 |
| $T_1$ WJ-III Word ID | 113.1 (2.9) | 109.3 (2.8) | 1.04 |
| $T_2$ WJ-III Word ID | 114.5 (12.3) | 106.3 (10.9) | 1.81 |
| $T_1$ WJ-III Word Attack | 106.5 (3.0) | 103.3 (2.4) | 0.89 |
| $T_2$ WJ-III Word Attack | 108.0 (10.9) | 102.2 (9.2) | 1.48 |
| Score delta | 16.6 (8.4) | 12.0 (4.8) | 1.69 |

IQ, WASI, TOWRE, and WJ-III standardized scores have $\mu = 100$, $\sigma = 15$. *$p < 0.05$, between-groups difference significant.

of phonological skills and the later importance of orthographic skills as predictors of subsequent reading ability (Badian, 1995; Share, 1995; Sprenger-Charolles et al., 1998; Sprenger-Charolles, 2003). Thus, an important hypothesis we want to test was whether the predictive ability of regions involved in phonological and orthographic processing depend on a child's age.

## Materials and Methods

The research protocol was approved by the Institutional Review Board of Northwestern University. Parental consent and the child's assent were obtained for all evaluation procedures, and children were paid for their participation in the study.

*Participants.* Twenty-six healthy children (9–15 year olds, 14 females) recruited from the Chicago metropolitan area participated in the study. Parents of children were given an interview to ensure their children met the inclusionary criteria for the study. Children were all right handed (mean, 80; range, 55–90) according to a nine-item Likert-scale questionnaire (90 to 90, positive scores indicate right-hand dominance), native English speakers, with normal hearing and normal or corrected-to-normal vision. All children were reported by their parents to be free of neurological diseases or psychiatric disorders and were not taking medication affecting the CNS. Children were reported by their parents to not have a history of intelligence, reading, attention, or oral-language deficits. Because we expected to find differences between younger and older readers, participants were grouped on the basis of chronological age on admission to the study into younger (9–11 years, $n = 14$) and older (13–15 years, $n = 12$) readers to facilitate the analysis and discussion of these differences. The ratio of males to females did not differ between the younger and older groups ($\chi^2_{(1)} = 1.90$, $p > 0.1$).

*Standardized testing.* Children were assessed on entering the study ($T_1$) and after a follow-up period of between ~3 and 5 years ($T_2$) with a battery of standardized tests to ensure that all participants were of at least average IQ and reading ability and to provide a means of measuring change in reading skill. Mental ability was measured with the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) with two verbal subtests (vocabulary, similarities) and two performance subtests (block design, matrix reasoning). Word and nonword reading accuracy was measured with the Woodcock Johnson III Tests of Achievement (WJ-III) (Woodcock et al., 2001). Word and nonword reading speed was measured by the Tests of Word Reading Sight Word Efficiency (TOWRE-SWE) and Pseudoword Decoding Efficiency (TOWRE-PDE) subtests, respectively (Torgesen et al., 1999). See Table 1 for a summary of performance on these tasks for younger ($T_1$ ages 9–11 years) and older ($T_1$ ages 13–15 years) children. The TOWRE subtests measure the number of consecutively presented written items that individuals can pronounce in 45 s, with later list items increasing in difficulty. Although the TOWRE-PDE is a measure of nonword reading ability, we note that

words not yet learned are essentially pseudowords: within the theoretical framework we adopt in our interpretation of the results, nonword reading relies on the same mechanism that encodes regularities learned when mapping between orthographic and phonological representations during normal word reading (Seidenberg et al., 1994). Moreover, in a behavioral study of 87 children of at least 9 years of age (the mean age of our younger group of readers), Barker et al. (1992) found that word and nonword reading measures were highly correlated ($r = 0.87$). Thus, the fluency with which children process pseudowords provides a reasonable measure of a child's capacity to learn to read. Because it assesses reading skill in terms of both speed and accuracy, varies highly between individuals relative to other measures, and assesses reading relatively independently of the child's vocabulary, the TOWRE-PDE test was used as our measure of reading skill for the regression analyses that follow. Within our sample, TOWRE-PDE raw scores ranged from 18 to 51 (of a possible 63) for $T_1$ and ranged from 17 to 62 for $T_2$.

Pearson's product-moment correlation coefficients calculated between all pairs of standardized measures found all reading measures to be significantly and positively correlated with one another at both $T_1$ and $T_2$, with correlations ranging between $r_{(25)} = 0.51$ and 0.83. $T_1$ and $T_2$ measures were also significantly and positively correlated, ranging from $r_{(25)} = 0.54$ for $T_1$ and $T_2$ WJ-III Word Attack scores, to 0.88 for WJ-III Word ID scores.

Younger and older children were generally comparable across the 10 standardized measures (see Table 1), with the younger group having greater verbal IQ and TOWRE-PDE scores. The groups, however, fell within 1 SD of one another and at or above average, even for the measures that differed significantly. Moreover, because the measures that differed were included as regressors in our analyses, these differences were accounted for by our model. Finally, we note that the presence of children with high standardized reading measures in the younger group, if anything, make it more similar to the older group by virtue of including more advanced readers. Because we predict skill-related differences between the groups, this would decrease our experimental power, making more likely null results.

*Task description.* The task, to which we will refer as the rhyming task for the remainder of the paper, included lexical and nonlexical trials. For lexical trials, word pairs were presented visually in a sequential order and participants had to determine whether the foveally read words rhymed. Each word was presented for 800 ms, followed by a 200 ms blank interval. A red fixation cross appeared on the screen after the second word, indicating the need to make a response during the subsequent 2600 ms interval. Twenty-four word pairs were presented in each of four lexical conditions that independently manipulated the orthographic and phonological similarity between words. In the two nonconflicting conditions, paired words were either similar in both orthography and phonology (O+P+, e.g., dime–lime) or different in both orthography and phonology (O−P−, e.g., staff–gain). In the two conflicting conditions, paired words had either similar orthography but different phonology (O+P−, e.g., pint–mint) or different orthography but similar phonology (O−P+, e.g., jazz–has). Participants were instructed to press a button with their right index finger for rhyming word pairs and to press a different button with their middle finger for nonrhyming word pairs. There were three nonlexical conditions. For the fixation condition, which served as our baseline, there were 72 trials, during which a black fixation cross was presented for the same duration as the stimuli in the lexical and perceptual conditions (see below) and participants were instructed to press a button when the black fixation cross turned red, after 1800 ms, corresponding to the onset of the red fixation cross during lexical trials. Additionally, there were 24 trials each of a simple and complex perceptual condition that served as an additional baseline condition for a related study. For these trials, participants were instructed to indicate via button press whether sequentially presented abstract single symbols (simple) or symbol strings (complex) matched. The stimuli were identical to those used by Bitan et al. (2007b), and stimuli characteristics are described in detail therein.

*Experimental procedure.* After informed consent was obtained and the standardized tests were administered, participants were invited for a two-part practice session. First, participants were trained to minimize head

movement using feedback from an infrared tracking device. Second, participants performed one run of the rhyming task in a simulator scanner to ensure they understood the task and to become familiar with the scanner environment. Different stimuli were used in the practice and in the scanning sessions, which took place within 1 week of each other. In the scanning session, the task was administered in two 108 trial runs, in which the order of conditions was optimized for event-related design (Burock et al., 1998). The order of stimuli within task was fixed for all subjects.

*MRI data acquisition.* Images were acquired using a 1.5 T GE scanner, using a standard head coil. Head movement was minimized using vacuum pillow (Bionix). The stimuli were projected onto a screen and viewed through a mirror attached to the inside of the head coil. Participants' responses were recorded using an optical response box (Current Designs). The BOLD functional images were acquired using the echo planar imaging method. The following parameters were used for scanning: TE, 35 ms; flip angle, 90°; matrix size, 64 × 64; field of view, 24 cm; slice thickness, 5 mm; voxel size, 3.75 × 3.75 × 5 mm³; number of slices, 24; TR, 2000 ms. Two runs, with 240 repetitions each, were administered for the functional images. In addition, structural T1-weighted 3D image were acquired (spoiled gradient-recalled acquisition in a steady state; TR, 21 ms; TE, 8 ms; flip angle, 20°; matrix size, 256 × 256; field of view, 22 cm; slice thickness, 1 mm; number of slices, 124), using the functional image orientation.

*Behavioral analysis.* Because interstimulus conflict has been shown to influence behavioral performance and BOLD activity for the task (Bitan et al., 2007b), a two-way ANOVA was conducted with conflict (conflicting vs nonconflicting) as a within-subjects independent variable and age group (young vs old) as a between-subjects independent variable. The dependent variables were decision latency and square root of the accuracy rate (Myers, 1979).

*Image analysis.* Data analysis was performed using statistical parametric mapping (SPM5) (http://www.fil.ion.ucl.ac.uk/spm). Images were spatially realigned to the first volume to correct for head movements. No individual runs had >4 mm maximum displacement. Sinc interpolation was used to minimize timing errors between slices (Henson et al., 2002). The functional images were coregistered with the anatomical image and normalized to the standard T1 template volume (MNI), with a voxel size of 3 × 3 × 3 mm³. The data were then smoothed with a 10 mm isotropic Gaussian kernel. The first four volumes of each run, during which a fixation cross was presented, were dropped from the statistical analyses. Statistical analyses at the first level were calculated using an event-related design with four lexical conditions, two perceptual conditions, and the fixation condition as conditions of interest. A high-pass filter with a cutoff period of 128 s was applied. Word pairs were treated as individual events for analysis and modeled using a canonical hemodynamic response function. Group results were obtained using random-effects analyses by combining subject-specific summary statistics across the group as implemented in SPM5.

The main effect of all lexical conditions versus the baseline fixation condition was tested using a one-sample *t* test. Regions that were more active for the lexical condition at a $p < 0.01$ (uncorrected) voxel-level significance threshold and containing clusters of at least equal to 10 voxels were included in the language network, which served as a mask for the voxelwise multiple regression analyses.

The voxelwise multiple regression analyses included as regressors of interest $T_1$ age (age), $T_2$–$T_1$ change in raw TOWRE-PDE reading score (score delta), and the age × score delta interaction term formed by calculating the product of each subject's age and score delta. We included as nuisance regressors the $T_2$–$T_1$ change in age (age delta), initial raw TOWRE-PDE reading score (score), and nonverbal IQ score. The dependent measure was the signal for each voxel in the brain across all lexical trials.

*Region of interest analysis.* Previous literature suggested four a priori anatomical regions of interest (ROIs)—left dIFG, BG, FG, and IPL—from among the anatomical regions included in the reading network based on their association with processing phonological and orthographic representations. Individually based ROIs were created for these regions in two steps. In the first step, we identified in each of these

**Table 2. Means (and SDs) for accuracy and decision latencies on lexical trials**

| | Younger | | Older | |
|---|---|---|---|---|
| | Accuracy | Decision latency (ms) | Accuracy | Decision latency (ms) |
| Conflict | 0.71 (0.13) | 1562 (344) | 0.82 (0.07) | 1264 (307) |
| Nonconflict | 0.91 (0.05) | 1413 (302) | 0.96 (0.03) | 1141 (301) |

**Table 3. Table of coordinates for whole-brain group analysis for lexical minus fixation condition reaching cluster size > 10 and $p < 0.05$ (FDR corrected) significance threshold**

| Region | Size | FDR | Max Z | x | y | z |
|---|---|---|---|---|---|---|
| L dorsal inferior frontal gyrus (BA 9) | 4006 | <0.001 | 5.81 | −54 | 15 | 30 |
| R ventral inferior frontal gyrus (BA 47) | 392 | <0.001 | 4.66 | 39 | 24 | −6 |
| L fusiform gyrus (BA 37) | 270 | <0.001 | 4.61 | −45 | −54 | −15 |
| R parahippocampal gyrus | 1018 | <0.001 | 4.58 | 24 | −45 | 3 |
| R precentral gyrus (BA 6) | 15 | 0.005 | 3.48 | 39 | −12 | 63 |
| L superior temporal gyrus (BA 22) | 43 | 0.008 | 3.27 | −60 | −42 | 12 |
| R postcentral gyrus (BA 43) | 10 | 0.015 | 3.03 | 54 | −12 | 18 |
| R red nucleus | 18 | 0.017 | 2.97 | 3 | −24 | −18 |
| R claustrum | 15 | 0.017 | 2.97 | 36 | −21 | 3 |
| R hippocampus | 30 | 0.019 | 2.94 | 30 | −15 | −18 |

L, Left; R, right; FDR, FDR-corrected significance level; Max, maximum. Size is measured in voxels. Coordinates reflect standard MNI space.

regions, as specified by the aal template (Tzourio-Mazoyer et al., 2002) definitions of l dIFG (pars triangularis and pars operculi), l BG (caudate and putamen), l FG, and l IPL, the voxel with the highest Z-score in the statistical map of the age-related interaction (see Table 5), around which a 10 mm radius sphere was drawn. In the second step, these spheres were intersected with each subject's activation map for the lexical condition minus fixation contrast.

The ROI multiple regression analysis of the interaction included as regressors of interest $T_1$ age (age) and $T_2$–$T_1$ change in raw TOWRE-PDE reading score (score delta). This analysis was intended to determine the direction of the first-order correlations within these ROIs and thus explain the source of the interaction. For reasons explained below, interpreting the significance of partial correlations obtained for these regions is difficult, and thus nuisance regressors were omitted from the model. The dependent measure was the mean beta weight for voxels included in each ROI.

## Results

### Behavioral performance
Mean and SDs for decision latency and accuracy for lexical trials are presented in Table 2. For accuracy, there was a main effect of conflict ($F_{(1,24)} = 86.02$, $p < 0.001$) and age ($F_{(1,24)} = 8.06$, $p < 0.01$), but these factors did not interact ($F_{(1,24)} = 2.99$, NS). The results for decision latency were similar, with a main effect of conflict ($F_{(1,24)} = 56.1$, $p < 0.001$) and age ($F_{(1,24)} = 5.42$, $p < 0.05$) but no interaction between the two factors ($F_{(1,24)} < 1$, NS). Thus, decisions made for conflicting pairs were slower and less accurate, and younger children were slower and less accurate than older children, but both age groups were similarly sensitive to the conflict effect.

### Neuroimaging results: overall effects
The reading network was defined by a group analysis of the lexical condition minus fixation baseline contrast. Table 3 lists the statistically significant clusters containing at least 10 voxels for the reading network at the $p < 0.05$ significance threshold [false discovery rate (FDR) corrected at the whole-brain level]. Activation was apparent in both hemispheres but was left-hemisphere dominant. Figure 1 shows the left-hemisphere clusters at the same significance threshold. The largest cluster in this activation map is expansive, ranging from left prefrontal areas (BA 44/45) to

parietal areas, and includes two of our a priori regions of interest. To more accurately describe the activity within each of these regions, we reanalyzed the data within individual ROIs anatomically defined by the aal template (Tzourio-Mazoyer et al., 2002). As indicated in Table 4, clusters containing at least 10 voxels reaching FDR-corrected significance (small-volume correction applied for the volume of each ROI) were found within each of these regions.

### Neuroimaging results: voxelwise analysis of interaction between age and score delta

A multiple regression analysis performed on areas outside of the reading network mask found no clusters that significantly predicted score delta, even at a very liberal uncorrected threshold of $p < 0.01$, indicating that only voxels that were associated with the lexical task were reliable predictors of reading ability. Within the reading network mask, a number of clusters containing 10 or more voxels showed significant correlations at the $p < 0.05$ (FDR corrected) threshold between $T_1$ lexical minus fixation activity and score delta. Importantly, the statistical map for which the age $\times$ score delta interaction term (Fig. 2) reached the same threshold was nearly identical in extent, indicating that the predictive ability of activity of nodes in the reading network depends on age. Table 5 lists the clusters demonstrating an age-related interaction with changes in reading proficiency at the $p < 0.05$ threshold (FDR corrected, small volume correction applied for the mask volume).

Because the mean verbal IQ score for younger children was higher than that for older children, we repeated the analysis, replacing nonverbal IQ with verbal IQ as a nuisance regressor. The results were very similar, likely because verbal and nonverbal IQ were significantly correlated ($r_{(25)} = 0.75$, $p < 0.001$).

### Neuroimaging results: ROI analysis of interaction between age and score delta

We further analyzed the imaging data separately for younger and older participants to determine the nature of the age-related interaction observed throughout the reading network. The interaction described above indicates that the relationship between brain activity in the language network and score delta differs by age. Thus, it is important to determine within each age group how the intensity of activation a child exhibits in this network relative to peers relates to score delta. Each of the regions for which ROIs were created demonstrated an age-related interaction at the whole-brain level. Although implicated in phonological processing of word stimuli, the STG was not included as a region of interest because it neither reliably predicted reading in the voxelwise regression analysis nor demonstrated an age-related interaction. Mean beta values for the lexical minus fixation contrast were calculated over the voxels included in each ROI for each subject, and the lexical minus fixation beta value was correlated with score delta across participants to determine the direction of the correlations for each age group in each region. Figure 3a–d show the
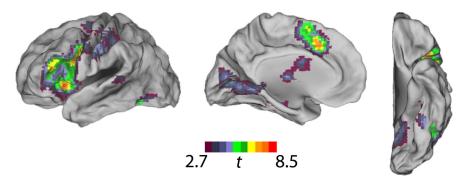


**Figure 1.** Whole-brain group analysis for lexical minus fixation condition showing left-hemisphere clusters of $k > 10$ at $p < 0.05$ (FDR corrected).

**Table 4. Table of coordinates for clusters within the anatomically defined reading network ROIs in the group analysis for lexical minus fixation condition reaching $p < 0.05$ (FDR corrected) significance**

| Region | Size | FDR | Max Z | x | y | z |
|---|---|---|---|---|---|---|
| L dorsal inferior frontal gyrus (BA 9/47/13) | 659 | <0.001 | 5.81 | −54 | 15 | 30 |
| L putamen, caudate body | 477 | <0.001 | 4.66 | −18 | 3 | 6 |
| L fusiform gyrus (BA 37/20) | 122 | <0.001 | 4.59 | −45 | −54 | −18 |
| L superior temporal gyrus (BA 22) | 124 | <0.001 | 4.52 | −48 | 9 | 0 |
| L inferior parietal lobule (BA 2/40) | 279 | <0.001 | 4.18 | −51 | −21 | 36 |
| L superior temporal gyrus (BA 22) | 45 | 0.002 | 3.27 | −60 | −42 | 12 |

L, Left; FDR, FDR-corrected significance level; Max, maximum. Size is measured in voxels. Coordinates reflect standard MNI space.
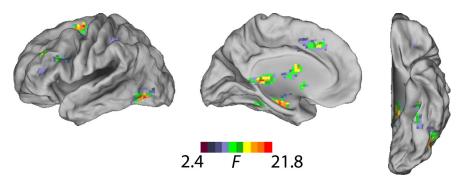


**Figure 2.** Voxels within the reading network demonstrating an age-related interaction with changes in reading proficiency.

**Table 5. Table of coordinates for clusters within the reading network demonstrating an age-related interaction with changes in reading proficiency**

| Region | Size | FDR | Max Z | x | y | z |
|---|---|---|---|---|---|---|
| L caudate body | 405 | 0.009 | 4.05 | −12 | 3 | 18 |
| L inferior occipital gyrus/fusiform gyrus (BA 19/37) | 341 | 0.009 | 3.58 | −45 | −72 | −9 |
| L thalamus | 153 | 0.009 | 3.58 | −9 | −36 | 6 |
| L middle frontal gyrus (BA 6) | 301 | 0.009 | 3.55 | −21 | −6 | 66 |
| L middle frontal gyrus (BA 9) | 44 | 0.009 | 3.42 | −33 | 39 | 33 |
| L dorsal inferior frontal gyrus (BA 9) | 90 | 0.009 | 3.29 | −54 | 18 | 33 |
| L precuneus (BA 7) | 91 | 0.009 | 3.01 | −24 | −60 | 54 |
| L inferior parietal lobule (BA 40) | 55 | 0.01 | 2.68 | −42 | −39 | 54 |
| L insula (BA 13) | 25 | 0.011 | 2.52 | −48 | −33 | 21 |

L, Left; FDR, FDR-corrected significance level; Max, maximum. Size is measured in voxels. Coordinates reflect standard MNI space.

volumetric union across subjects and the regression lines for each of the ROIs. Overall, interactions in these regions at the whole-brain level were driven mainly by positive correlations between activation and performance for young children in caudate and
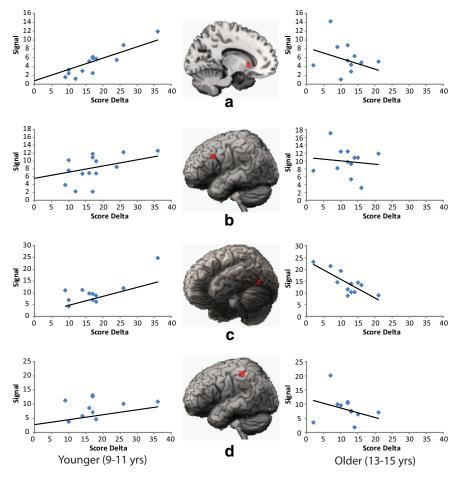
**Figure 3.** Regression lines and ROIs for left caudate (*a*), left dorsal inferior frontal gyrus (*b*), left fusiform gyrus (*c*), and left inferior parietal lobule (*d*).
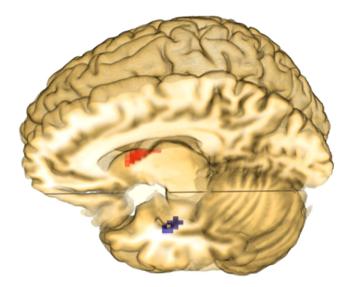


**Figure 4.** Voxels within the left caudate nucleus and putamen (red) demonstrate a positive correlation for younger children and those within the left fusiform gyrus (blue) demonstrate a negative correlation for older children between activation and change in reading fluency.

left dIFG and negative correlations between activation and performance for older children in left FG. Note, however, that correlation coefficients calculated for regions defined, in part at least, by voxels demonstrating a significant interaction may be inflated (Vul et al., 2009) and are therefore not reported.

## Neuroimaging results: voxelwise analysis within younger and older children

To obtain unbiased correlation measures within our regions of interest, we repeated the voxelwise analysis on anatomically defined ROIs separately for the younger and older groups, masked by the aal template (Tzourio-Mazoyer et al., 2002) definitions of the left BG (caudate and putamen), dIFG (pars triangularis and pars operculus), IPL, and FG regions. The ROI regression analysis suggested that any correlations observed in these regions would be positive for younger children and negative for older children. Figure 4 presents the activation map for these contrasts at a $p < 0.005$ (uncorrected) significance threshold showing only clusters of 10 or more contiguous voxels, and Table 6 presents the associated coordinates and statistical values. Of the correlations observed in the ROI analysis, only the positive correlation found in left basal ganglia (caudate and putamen) for younger children and the negative correlation found in the left fusiform gyrus for older children were reliable at this significance level. Younger participants showed no negative correlations and older participants showed no positive correlations in the selected regions at the $p < 0.05$ (uncorrected) significance threshold, indicating that these regions are relatively homogenous with respect to predictive ability.

## Neuroimaging results: effects of conflict

Previous literature has shown activation for our task to be sensitive to orthographic and phonological conflict between words (Bitan et al., 2007b), and thus we hypothesized that one or more lexical conditions may be responsible for driving these effects. That is, perhaps the presence (or absence) of certain types of conflict makes some items more predictive of subsequent reading than others. To address this possibility, we performed a 2 (orthographic conflict) × 2 (phonological conflict) × 4 (region) repeated-measures analysis of covariance, separately for younger and older children. We entered the previously extracted beta values for each condition as our independent variables, score delta as our dependent variable, and age delta, initial score, and nonverbal IQ as covariates of noninterest. There were no significant interactions involving score delta and either lexical factor (all $F$ values <1), indicating that activation in response to orthographic or phonological conflict was not differentially predictive of reading ability.

## Discussion

Measurements of neural activity in language network taken using fMRI during a rhyming task predicted reading fluency. This work advances the literature establishing a predictive relationship between brain activity and future reading skill using ERPs (Espy et al., 2004; Maurer et al., 2009; Lemons et al., 2010) and fMRI (Hoeft et al., 2007, 2011). Our focus on future (up to 6 years after scan) rather than concurrent reading ability allows examination

of developmental changes to the contributions of neural substrates supporting language toward reading and how the interaction between reading processes shape language representations in the brain.

The language network contains multiple functionally distinct regions developing at different rates with exposure to spoken and written language. Accordingly, our hypothesis that predictive ability of these regions is age dependent was supported by the age-related interaction. Specifically, activity in phonological processing areas predicted reading ability for third- to fifth-grade children, whereas activity in orthographic processing areas was predictive for seventh- to ninth-grade adolescents.

Early reliance on phonological skill is consistent with the timeline of language exposure: early experience with speech teaches infants the allowable set of phonemes in their native language (Kuhl, 2004), which are later associated with graphemes. Evidence that phonological skill predicts early reading skill is found in many longitudinal studies. For example, Badian's (1995) study of early readers showed an early predictive role of phonemic awareness giving way in later years to a larger influence of orthographic skill. This transition from phonological to orthographic dependence has been framed in terms of stage-based (Frith, 1985) and experientially based (Share, 1995) accounts, in which phonological recoding provides the error signal used in a self-teaching mechanism. Sprenger-Charolles (2003, p 196) argue on the basis of their own and previously published findings that it is "well established" that early reading relies primarily on phonological processing but that orthographic processing takes on a larger role over time, a point on which both stage-based and experientially based accounts of reading development agree. This conclusion agrees with literature spanning multiple methodologies, including ERPs (Ashby, 2010), fMRI (Desroches et al., 2010), and computational modeling (Rueckl and Seidenberg, 2009), showing that phonological and orthographic processes work together during skilled word recognition.

The nature of the brain-based interactions provides insight into how the dynamics of the reading network changes over time and how these changes relate to the developmental changes observed empirically. We interpret these interactions in the context of Harm and Seidenberg's (1999) parallel distributed processing reading model. Simulations presented therein explore how orthographic, phonologic, and cross-modal mapping systems support reading. Figure 5 presents a sketch of their connectionist model. Reading requires orthographic representations in the visual system (FG) to be mapped via a hidden layer (IPL) to associated phonological processing regions (STG, IFG, and BG). Orthographic to phonologic mapping is arbitrary and further complicated in languages such as English, which maintain context-dependent one-to-many letter-to-sound mappings. Thus, cross-modal mapping in reading requires pattern separation, a process by which similar signals are recoded into more discriminable representations.

Pattern separability depends on available processing resources. At one extreme, the completely overlapping representations generated by a lone cross-modal mapping unit (i.e., all inputs activate the one mapping unit) would be less distinct than the input and fail to differentiate inputs. At the other extreme, an

**Table 6. Table of coordinates for clusters within the a priori defined ROIs demonstrating positive correlations for younger and negative correlations for older children**

| Age group | Region | Size | Uncorrected p | Max Z | x | y | z |
|---|---|---|---|---|---|---|---|
| Younger | L caudate head | 53 | <0.001 | 3.67 | −9 | 6 | 6 |
| Older | L fusiform gyrus | 18 | <0.001 | 4.76 | −36 | −6 | −33 |

L, Left; FDR, FDR-corrected significance level; Max, maximum. Size is measured in voxels. Coordinates reflect standard MNI space.
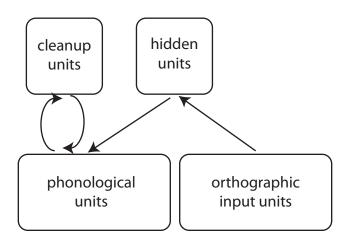


**Figure 5.** Harm and Seidenberg's (1999) attractor network model maps orthographic input to phonological representations via a hidden layer. Cleanup units representing higher-order phonological relationships fine-tune phonological representations.
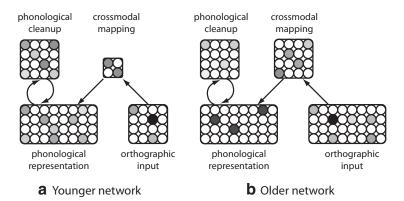


**Figure 6.** Fewer resources allotted to cross-modal mapping in the younger network constrain the number of distinct mapping patterns, necessitating greater feedback from phonological cleanup to fine-tune phonological representations.

infinite number of mapping units would be a perfect, albeit impractical, pattern separator because any input could be assigned a unit associated with a unique mapping representation. Between these extremes, the degree of overlap between mapping representations is inversely related to the number of mapping units. The phonological system overcomes the fuzziness inherent to overlapping mappings by using learned higher-order phonological relationships (e.g., coarticulatory probabilities) to constrain and correct generated phonological representations (Share, 1995). Rather than directly map every orthographic–phonologic relationship, the mapping system needs only get the phonological representation "in the right neighborhood," allowing the phonological cleanup units to refine the representation to one allowed by the language.

We assume that each system develops over time, as illustrated in Figure 6 but that the phonological system will have substantially matured by reading age. In younger readers (Fig. 6a), fewer

resources are devoted to cross-modal mapping, which thus generate more overlapping hidden representations and commensurately fuzzier initial phonological representations. Resolution of sharp phonological representations therefore relies on feedback from phonologic cleanup. Because more resources are allotted to cross-modal mapping with development (Fig. 6b), it generates less overlapping representations and places correspondingly smaller demand on phonological cleanup. The assumption of a developing cross-modal mapping system thus predicts a strong role of phonological skill in reading acquisition for younger readers that decreases with age.

We argue that younger children who strongly engage higher-order phonological processing areas more effectively use this knowledge to decode the phonological representations of written words. Graybiel (2005) argues that basal ganglia provide a reinforcement signal for motor learning. One intriguing possibility is that it plays a similar role within the anterior phonological circuit and facilitates the learning and use of phonological representations in reading. Thus, children who best apply the phonological recoding system to reading should more quickly build the knowledge on which more challenging reading depends, as suggested by Share (1995). Because a more developed mapping system can better perform pattern separation, phonological recoding plays a smaller role in shaping word representations in older children—even those who had previously made effective use of phonological processing areas—reflected by a weaker correlation between phonological processing and improvement in reading ability.

Another strategy for overcoming the bottleneck imposed by a developing cross-modal mapping system is to perform pattern separation at the input level. Thus, rather than treat "pint" and "mint" as similar orthographic inputs to be disambiguated, one could instead assign relatively distinct orthographic patterns to similar words by analyzing them at a coarser granularity (e.g., the whole-word approach to reading). One drawback is that, although English has many exception words, it nonetheless possesses a great deal of regularity on which children can capitalize. Thus, early in development, highly overlapping representations in the mapping system poses a challenge to phonological decoding. Later, decreased but moderate overlap among representations in a more developed cross-modal system captures regularities present in language and facilitates decoding challenging words.

FG is beginning to support orthographic processing in younger children and is unlikely to be qualitatively better at pattern separation among orthographic inputs than is the mapping system. Thus, the relationship between FG recruitment and reading skill in younger children should be weak because the burden of decoding should fall on the phonological cleanup system as described above. We argue that FG is predictive of reading in older children because its use influences the development of the cross-modal mapping system. Older children that rely on orthographic pattern separation (i.e., show greater activation in FG) cannot take advantage of statistical regularities between spelling and sound correspondences. Because older children encounter more varied and challenging written words, they will be unable to take advantage of these regularities and thus make smaller gains in reading ability if they rely on pattern separation at the input level. This is reflected in the negative correlation between activation in FG and change in reading score for older children. Note that several authors, whose accounts similarly predict increasing reliance on ventral systems, argue that ventral orthographic word form circuits are tuned by input from dorsal phonological circuits (Pugh et al., 2000; McCandliss, 2003). However, if FG is the source of visual word form input during reading, then the assumption of phonological feedback to this area implies bidirectional connectivity. Thus, phonology-driven tuning of FG and orthographic influence on cross-modal representations are mutually consistent and suggest a dynamic reciprocal influence between the two systems.

To summarize, based on functionality established in the reading literature, we found an age-dependent correlation between activity in brain areas associated with phonological and orthographic processing and future reading ability, consistent with what is known about the maturational timeline of these language processes: early reliance on phonological and later reliance on orthographic processing. A final consideration is that previous neuroimaging studies of reading have noted a developmental increase in activation of FG (Booth et al., 2003, 2004; Brem et al., 2006), IPL (Booth et al., 2003, 2007a; Bitan et al., 2007a), and IFG (Booth et al., 2003, 2004; Gaillard et al., 2003), consistent with the assumption that these regions recruit additional resources through development. We have described above the consequence of adding resources to phonological and cross-modal mapping systems. In light of the negative correlation observed in FG, one might wonder why reading does not therefore worsen with age. One must bear in mind our correlations were within an age group. All older children appear to recruit FG to a greater extent than younger children, but those that rely on it more, relative to their age-matched peers, are those that made the smallest gains in reading. Together, this suggests an optimal recruitment of the systems participating in reading acquisition that varies by age.

## References

Ashby J (2010) Phonology is fundamental in skilled reading: evidence from ERPs. Psychon Bull Rev 17:95–100.

Badian N (1995) Predicting reading ability over the long term: the changing roles of letter naming, phonological awareness and orthographic processing. Ann Dyslexia 45:79–96.

Barker TA, Torgesen JK, Wagner RK (1992) The role of orthographic processing skills on five different reading tasks. Read Res Q 27:335–345.

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

Bitan T, Burman DD, Chou TL, Lu D, Cone NE, Cao F, Bigio JD, Booth JR (2007b) The interaction between orthographic and phonological information in children: an fMRI study. Hum Brain Mapp 28:880–891.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002) Functional anatomy of intra- and cross-modal lexical tasks. Neuroimage 16:7–22.

Booth JR, Burman DD, Meyer JR, Lei Z, Choy J, Gitelman DR, Parrish TB, Mesulam MM (2003) Modality-specific and -independent developmental differences in the neural substrate for lexical processing. J Neurolinguist 16:383–405.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2004) Development of brain mechanisms for processing orthographic and phonologic representations. J Cogn Neurosci 16:1234–1249.

Booth JR, Cho S, Burman DD, Bitan T (2007a) Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. Dev Sci 10:441–451.

Booth JR, Wood L, Lu D, Houk JC, Bitan T (2007b) The role of the basal ganglia and cerebellum in language processing. Brain Res 1133:136–144.

Brem S, Bucher K, Halder P, Summers P, Dietrich T, Martin E, Brandeis D (2006) Evidence for developmental changes in the visual word processing network beyond adolescence. Neuroimage 29:822–837.

Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. Neuroreport 9:3735–3739.

Desroches AS, Cone NE, Bolger DJ, Bitan T, Burman DD, Booth JR (2010) Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. Brain Res 1356:73–84.

Espy KA, Molfese DL, Molfese VJ, Modglin A (2004) Development of auditory event-related potentials in young children and relations to word-level reading abilities at age 8 years. Ann Dyslexia 54:9–38.

Frith U (1985) Beneath the surface of developmental dyslexia. In: Surface dyslexia, pp 301–330. Hillsdale, NJ: Erlbaum.

Gaillard WD, Sachs BC, Whitnah JR, Ahmad Z, Balsamo LM, Petrella JR, Braniecki SH, McKinney CM, Hunter K, Xu B, Grandin CB (2003) Developmental aspects of language processing: fMRI of verbal fluency in children and adults. Hum Brain Mapp 18:176–185.

Graybiel AM (2005) The basal ganglia: learning new tricks and loving it. Curr Opin Neurobiol 15:638–644.

Guttorm TK, Leppänen PH, Poikkeus AM, Eklund KM, Lyytinen P, Lyytinen H (2005) Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. Cortex 41:291–303.

Guttorm TK, Leppänen PH, Hämäläinen JA, Eklund KM, Lyytinen HJ (2010) Newborn event-related potentials predict poorer pre-reading skills in children at risk for dyslexia. J Learn Disabil 43:391–401.

Harm MW, Seidenberg MS (1999) Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol Rev 106:491–528.

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ (2002) Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15:83–97.

Hoeft F, Ueno T, Reiss AL, Meyler A, Whitfield-Gabrieli S, Glover GH, Keller TA, Kobayashi N, Mazaika P, Jo B, Just MA, Gabrieli JD (2007) Prediction of children's reading skills using behavioral, functional, and structural neuroimaging measures. Behav Neurosci 121:602–613.

Hoeft F, McCandliss BD, Black JM, Gantman A, Zakerani N, Hulme C, Lyytinen H, Whitfield-Gabrieli S, Glover GH, Reiss AL, Gabrieli JD (2011) Neural systems predicting long-term outcome in dyslexia. Proc Natl Acad Sci U S A 108:361–366.

Kuhl PK (2004) Early language acquisition: cracking the speech code. Nat Rev Neurosci 5:831–843.

Lemons CJ, Key AP, Fuchs D, Yoder PJ, Fuchs LS, Compton DL, Williams SM, Bouton B (2010) Predicting reading growth with event-related potentials: thinking differently about indexing "Responsiveness." Learn Individ Differ 20:158–166.

Maurer U, Bucher K, Brem S, Benz R, Kranz F, Schulz E, van der Mark S, Steinhausen HC, Brandeis D (2009) Neurophysiology in preschool improves behavioral prediction of reading ability throughout primary school. Biol Psychiatry 66:341–348.

McCandliss BD, Cohen L, Dehaene S (2003) The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn Sci 7:293–299.

Myers JL (1979) Fundamentals of experimental design, Ed 3. Boston: Allyn and Bacon.

Pugh KR, Mencl WE, Jenner AR, Katz L, Frost SJ, Lee JR, Shaywitz SE, Shaywitz BA (2000) Functional neuroimaging studies of reading and reading disability (developmental dyslexia). Ment Retard Dev Disabil Res Rev 6:207–213.

Pugh KR, Mencl WE, Jenner AR, Katz L, Frost SJ, Lee JR, Shaywitz SE, Shaywitz BA (2001) Neurobiological studies of reading and reading disability. J Commun Disord 34:479–492.

Rueckl JG, Seidenberg MS (2009) Computational modeling and the neural bases of reading disorders. In: How children learn to read: current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice (Pugh KR, McCardle P, eds), pp 99–131. London: Psychology Press.

Seidenberg MS, Plaut DC, Petersen AS, McClelland JL, McRae K (1994) Nonword pronunciation and models of word recognition. J Exp Psychol Hum Percept Perform 20:1177–1196.

Share DL (1995) Phonological recoding and self-teaching: sine qua non of reading acquisition. Cognition 55:151–218; discussion 219–226.

Sprenger-Charolles L, Siegel LS, Bonnet P (1998) Reading and spelling acquisition in French: the role of phonological mediation and orthographic factors. J Exp Child Psychol 68:134–165.

Sprenger-Charolles L, Siegel LS, Béchennec D, Serniclaes W (2003) Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling: a four-year longitudinal study. J Exp Child Psychol 84:194–217.

Torgesen JK, Wagner RK, Rashotte CA (1999) TOWRE: test of word reading efficiency. Austin, TX: PRO-ED.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect Psychol Sci 4:274–290.

Wechsler D (1999) Wechsler abbreviated scale of intelligence (WASI). San Antonio, TX: The Psychological Corporation.

Woodcock RW, McGrew KS, Mather N (2001) Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside.