Socioeconomic Status and Reading Disability: Neuroanatomy and Plasticity in Response to Intervention

Citation

Romeo, Rachel R. et al. “Socioeconomic Status and Reading Disability: Neuroanatomy and Plasticity in Response to Intervention.” Cerebral Cortex (June 2017): 1–16 © 2017 The Authors

As Published

http://dx.doi.org/10.1093/cercor/bhx131

Publisher

Oxford University Press

Version

Author's final manuscript

Accessed

Sat Feb 09 14:57:47 EST 2019

Citable Link

http://hdl.handle.net/1721.1/113201

Terms of Use

Creative Commons Attribution-Noncommercial-Share Alike

Detailed Terms

http://creativecommons.org/licenses/by-nc-sa/4.0/
Abstract

Socioeconomic Status and Reading Disability:
Neuroanatomy and Plasticity in Response to Intervention

Rachel R. Romeo*1,2
Joanna A. Christodoulou*2,3,4
Kelly K. Halverson2
Jack Murtagh2
Abigail B. Cyr2
Carly Schimmel2
Patricia Chang2
Pamela E. Hook3
John D.E. Gabrieli2,4,5

* These authors contributed equally to this work.

1 Division of Medical Sciences, Harvard Medical School and Harvard University, Boston, MA 02115, USA
2 McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
3 Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA 02129, USA
4 Harvard Graduate School of Education, Cambridge, MA 02138, USA
5 MIT Integrated Learning Initiative, Cambridge, Massachusetts 02139, USA
Although reading disability (RD) and socioeconomic status (SES) are independently associated with variation in reading ability and brain structure/function, the joint influence of SES and RD on neuroanatomy and/or response to intervention is unknown. Sixty-five children with RD (ages 6 to 9) with diverse SES were assigned to an intensive, 6-week summer reading intervention ($n = 40$) or to a waiting-list control group ($n = 25$). Before and after, all children completed standardized reading assessments and magnetic resonance imaging (MRI) to measure cortical thickness. At baseline, higher SES correlated with greater vocabulary and greater cortical thickness in bilateral perisylvian and supramarginal regions—especially in left pars opercularis. Within the intervention group, lower SES was associated with both greater reading improvement and greater cortical thickening across broad, bilateral occipitotemporal and temporoparietal regions following the intervention. Additionally, treatment responders ($n = 20$), compared to treatment non-responders ($n = 19$), exhibited significantly greater cortical thickening within similar regions. The waiting control and non-responder groups exhibited developmentally-typical, non-significant cortical thinning during this time period. These findings indicate that effective summer reading intervention is coupled with cortical growth, and is especially beneficial for children with RD who come from lower-SES home environments.

(Word count: 197)

Keywords: cortical thickness, dyslexia, longitudinal, neuroimaging, SES
Introduction

Reading is the bedrock of early education, and difficulty in reading has widespread and long-term consequences. Two major factors associated with difficulty in learning to read are reading disability (RD) and socioeconomic status (SES). Reading disability (RD) is the most prevalent type of learning disability (Shaywitz et al. 2008), and is estimated to affect about 10% of school-age children (Shaywitz 1998). Developmental dyslexia describes children with RD who demonstrate difficulty with single word reading accuracy or fluency in the context of intact cognitive skills and adequate educational opportunity (Lyon et al. 2003). SES is a common conceptualization of the social and economic status of an individual or group that is often measured by some combination of parental educational attainment, income, and occupation. Higher SES is associated with better reading outcomes (Peterson and Pennington 2015), but unlike RD, SES is associated with environmental factors such as home language environment (Hoff et al. 2002) and quality of school instruction (Lee and Burkam 2002). Here we asked whether there are neuroanatomical brain differences in young children (ages 6 to 9) with RD from varying SES backgrounds, and whether a reading intervention yields similar or dissimilar reading benefits and brain plasticity in children across the SES continuum.

Neuroimaging studies of children and adults with RD have revealed both structural and functional differences as compared to typical readers (for review, see Norton et al. 2015). Structurally, RD is typically associated with cortical gray matter reductions in bilateral temporoparietal regions underlying phonological processing (Brown et al. 2001; Eckert 2004; Silani et al. 2005; Vinckenbosch et al. 2005; Hoeft et al. 2007) and left occipitotemporal regions underlying visual whole-word recognition (Eckert 2004; Silani et al. 2005; Kronbichler et al. 2008; Steinbrink et al. 2008), as well as parts of the cerebellum bilaterally (Brown et al. 2001; Eckert et al. 2003; Brambati et al. 2004; Kronbichler et al. 2008; for meta-analyses, see Linkersdörfer et al. 2012; Richlan et al. 2013). These gray matter disparities are evident even in
young children with a family history of dyslexia who have yet to learn how to read (Raschle et al. 2011), suggesting that these differences are not purely a consequence of reading difficulty. Some studies have found additional gray matter reductions in canonical language regions, including left inferior frontal cortex (including Broca’s area) and left superior temporal cortex (including Wernicke’s area) in both children with RD (Eckert et al. 2003; Hoeft et al. 2007) and adults with RD (Brown et al. 2001; Brambati et al. 2004; Steinbrink et al. 2008). These neuroanatomical differences in left-hemisphere language areas are consistent with evidence that a weakness in a specific component of language, namely some aspects of phonological processing, is one of the most common pre-reading predictors and continuing correlates of RD (Melby-Lervåg et al. 2012).

There is also evidence for brain plasticity following intervention in both children and adults with RD. Most neuroimaging studies examining intervention-induced plasticity have measured functional changes, and often report normalization of pre-intervention hypoactivation in left-hemisphere regions associated with reading and language, as well as increased activation in right-hemisphere homologues interpreted as compensatory plasticity (for reviews, see Gabrieli et al. 2010; Barquero et al. 2014). The two studies examining intervention-induced structural plasticity have reported bilateral changes in the hippocampal region, left precuneus, and right cerebellum (Krafnick et al. 2011) and in white-matter microstructure of the left anterior centrum semiovale that correlated with improvement in phonological decoding ability (Keller and Just 2009).

Socioeconomic status is also strongly associated with reading skill (White 1982; Bowey 1995; Hecht et al. 2000). The disproportionate influence of SES on reading and language skills, as compared to other cognitive domains, is thought to arise from variation in the quantity and complexity of early language exposure (for reviews, see Hoff 2006; Perkins et al. 2013; Schwab
and Lew-Williams 2016). SES-related differences in brain structure are evident as early as one month of age (Betancourt et al. 2015), and appear to increase with age (Noble et al. 2012; Hanson et al. 2013). Specifically, lower SES is correlated with reduced activation in left perisylvian regions during language-related tasks (Raizada et al. 2008) and reduced gray matter in both left perisylvian regions (Noble et al. 2012; Noble et al. 2015) and bilateral occipitotemporal regions (Jednorog et al. 2012; Mackey et al. 2015), among many other regions (see Brito and Noble 2014 for review).

Although separate lines of evidence have revealed neuroanatomical differences in left-hemisphere language areas in relation to RD and SES, these two lines of evidence have yet to be integrated. This is an important gap in knowledge to address, because children from lower-SES backgrounds disproportionately meet RD criteria (Peterson and Pennington 2015) and are diagnosed with specific learning disabilities at significantly higher rates than children from higher SES backgrounds (Shifrer et al. 2011). In 2015, 4th and 8th grade students eligible for the National School Lunch Program (indicating low family income) were 2.5 times more likely to read at a “below proficient” level than students from higher-income families (U.S. Department of Education 2015). This may be related to gene × environment interactions, such that a genetic risk for RD is amplified by decreased access to reading/literary resources in lower-SES environments (Mascheretti et al. 2013), and/or that potentially typical-readers are not achieving their potential due to decreased resources (Friend et al. 2008). Therefore, it is important to understand whether RD arises from similar brain differences and responds similarly or differently to interventions for lower- and higher-SES students.

There are currently no studies examining RD and SES interactions in regard to brain structure, but two functional magnetic resonance imaging (fMRI) studies examining this interaction have yielded conflicting results. One study investigated the effects of SES on the relationship
between phonological awareness, word decoding, and brain activation in children (Noble et al. 2006). Participants (6-9 years old) were recruited based on a history of reading difficulty and, on average, scored in the low- to below-average range on standardized assessments of pseudoword reading skills and phonological awareness. Among children with the lowest phonological awareness scores, higher-SES children exhibited an increased response, versus lower-SES children, in left fusiform and perisylvian regions while viewing pseudowords versus a fixation cross (Noble et al. 2006). The other study investigated the effects of SES on brain activation while both typical readers and children with a diagnosis of dyslexia (8-10 years old) viewed words (versus houses, faces, checkerboard, and blank screen) and listened to speech (versus foreign language and silence) (Monzalvo et al. 2012). Although there were SES-related activation differences for the speech task in right-hemisphere perisylvian regions, there were no SES-related differences during the visual word task (Monzalvo et al. 2012). These two functional studies reached conflicting conclusions about the relation between SES and functional activation in response to print, which could be explained by any number of methodological differences, including language (English vs. French), participant age (6-9 vs. 8-10 years), sample size (38 vs. 23 non-control children), inclusion criteria (history of reading difficulty vs. externally diagnosed dyslexia), SES measurement (continuous variable based on parental education, occupation, income-to-needs ratio vs. categorical variable based on school districts), and/or print stimuli (pseudowords vs. short familiar real words).

The most important goal of understanding RD is to help people overcome reading difficulties, to the extent possible, through educational intervention. Although SES is relatively easy to measure and known to be associated with reading skill, very few studies have asked whether participant response to an intervention varies in relation to SES. A review of 14 studies reported behavioral factors predicting responsiveness to literacy interventions (Lam and McMaster 2014). Although the majority of studies collected some sort of SES information, only 4 studies analyzed
SES as a predictive factor; of these, 2 found that higher SES predicted better treatment response on reading outcome measures (Hatcher et al. 2006; Morris et al. 2012). The other studies either treated SES as a nuisance variable or as a descriptive characterization of their overall sample. Similarly, neuroimaging studies examining brain plasticity associated with intervention rarely consider the SES of participants. In a review of functional neuroimaging studies of reading interventions (Barquero et al. 2014), only 4 of 22 studies reported participant SES information. Of these, only one study (Bach et al. 2013) examined the relationship between SES and intervention outcomes, albeit only in behavioral outcomes. Although this specific study of Swiss-German children did not reveal SES as predictive of intervention efficacy, relationships between SES and academic achievement appear to be stronger in individuals from the United States (Tucker-Drob and Bates 2016), potentially due to greater SES variability in educational quality in the United States. Given how strong the effects of U.S. SES are on both children’s reading ability and their neural architecture, it may be that SES is related to behavioral and neuroanatomical intervention response sensitivity in U.S. children.

In the present study, we recruited young children with RD from a broad SES range and assigned them either to an intensive reading intervention during the school summer break or to a waiting-list control group. First, we asked whether cortical thickness varied by SES at baseline, because it is unknown whether there is such a relation between RD and SES neuroanatomically. Based on the literature linking SES to brain structure in typically developing children (e.g., Noble and Brito 2014), we hypothesized that higher-SES children with RD would exhibit thicker cortex, especially in inferior frontal and posterior temporal regions canonically associated with language and reading. Second, we asked whether SES was related to intervention efficacy in relation to reading outcomes and structural brain plasticity. While there is some evidence of structural brain plasticity associated with reading intervention (Keller and Just 2009; Krafnick et al. 2011), the specific effect on cortical thickness and the relations of plasticity
to SES and treatment response are unknown. Behaviorally, we hypothesized that higher-SES children would respond more positively to the intervention, based on previous intervention response findings (e.g., Torgesen et al. 1999; Hatcher et al. 2006; Morris et al. 2012).

Furthermore, we predicted that the children who exhibited greater behavioral improvement would also exhibit greater gains in cortical thickness.

**Materials and Methods**

**Participants**

Children (n = 65, 22 female) with RD who were between the ages of 6 and 9 years (M = 7.75 years, SD = 0.64 years) and completing grade 1 or 2 were recruited from communities in an SES-diverse Northeast region around a major urban center. Specifically, children were recruited both from the community at-large (n = 50) and from a local partner school (n = 15), which was an SES-diverse urban charter school.

Inclusion criteria required participants to have a history of reading difficulty based on parental report, a current demonstration of reading difficulty, and no neurological or psychiatric impairments or associated medications with the exception of Attention Deficit Hyperactivity Disorder (ADHD). Eleven children carried a diagnosis of ADHD, a disorder highly co-morbid with reading disability (Germano et al. 2010), and 6 of these children received daily medication. However, they did not differ from the remaining participants on any behavioral measures or demographic variables (all p > 0.13), so all 11 were included in the final sample. Additionally, all participants were native English speakers, although 6 participants were simultaneous bilinguals (natively acquired English and another language from birth), and 5 others had exposure to a second language outside of typical foreign language class at school. There was no relationship
between bilingualism and any demographic variable, assessment score, group assignment, or intervention response (all \( p > 0.05 \)). Behavioral findings from a subset of these children (\( n = 47 \)) who participated in the first phase of the intervention study were previously reported (Christodoulou et al. 2015). Findings reported here are from all children who participated in the intervention study except for those whose neuroimaging data were problematic (described below). Written informed consent was obtained from parents, and written assent was obtained from all child participants. All procedures were approved by the Institutional Review Board at the Massachusetts Institute of Technology.

Demographics

Participants’ SES was determined by a composite of maternal education and occupational prestige, as calculated by the Barratt Simplified Measure of Social Status (BSMSS; Barratt 2006). Maternal factors were chosen because they are the most frequently used SES measure (Ensminger and Fothergill 2003), are considered to have stronger relation than paternal factors to cognitive development in younger children (Mercy and Steelman 1982), and because 13 participants lived in single-mother homes. For the 4 participants whose mothers were full-time homemakers, paternal occupation was substituted and combined with maternal education. The BSMSS scale yields possible scores ranging from 8 (lower SES) to 66 (higher SES); participants’ scores ranged from 17 to 66 (\( M = 47.35, SD = 11.75 \)). Maternal education and occupation scores were highly correlated (Pearson’s \( r = 0.76, p < 1^{12} \)), supporting their combination into a composite measure. Additionally, 48 participants (74%) optionally reported their annual gross family income, which ranged from $15,000 to >$120,000 (\( M = $77,400, SD = $33,550 \)). Income was highly correlated with maternal education (\( r = 0.53, p < 0.001 \)), maternal occupation (\( r = 0.47, p < 0.001 \)), and total BSMSS scores (\( r = 0.51, p < 0.001 \)); thus, BSMSS scores were judged to be a valid index of SES. Unless otherwise noted, SES was treated as a continuous variable for all analyses.
Behavioral Assessments

Screening Session

Participants were first invited to a screening session, at which they completed a battery of tests. Non-verbal cognition was assessed with the Matrices subtest of the *Kaufman Brief Intelligence Test, 2nd edition* (KBIT-2; Kaufman and Kaufman 2004). Core reading subskills were assessed with the Elision and Nonword Repetition subtests of the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner et al. 1999), and the Objects, Letters, and 2-set Letters and Numbers subtests of the *Rapid Automatized Naming and Rapid Alternating Stimulus Tests* (RAN/RAS; Wolf and Denckla 2005). Reading was assessed by the Oral Reading Fluency subtest of the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good and Kaminski 2002).

Participants were included in the final RD sample (*n* = 65) if they (1) scored ‘At Risk’ or ‘Some Risk’ on the DIBELS (Good and Kaminski 2002), a criterion-referenced benchmark assessment (*n* = 56), and/or (2) scored below the 25th percentile on at least three of five phonological processing (CTOPP; Wagner et al. 1999) and rapid naming (RAN/RAS; Wolf and Denckla 2005) subtests—skills that are highly associated with reading ability (*n* = 32). Twenty-three participants met both criteria, and there were no demographic differences between the two inclusion criteria (all *p* > 0.23). Additionally, all participants were required to score at or above the 16th percentile on a measure of non-verbal cognitive ability (Matrices subtest, KBIT-2; Kaufman and Kaufman 2004). Twenty-four participants (37%) possessed an external diagnosis of dyslexia or a reading-based learning disability.

Pre-Intervention Characterization
After meeting inclusion criteria, participants completed additional assessments of language skills. Two additional CTOPP subtests (Blending Words and Memory for Digits) were administered to better characterize phonological processing (Wagner et al. 1999). Receptive vocabulary was assessed with the *Peabody Picture Vocabulary Test, 4th edition* (PPVT-4; Dunn and Dunn 2007).

*Pre- & Post-Intervention Outcome Measures*

Four *a priori* outcome measures were administered before (pre-test) and after (post-test) the intervention/waiting period: untimed word reading [Word Identification subtest (Word ID), *Woodcock Reading Mastery Test, 3rd edition* (WRMT-3); Woodcock 2011], untimed pseudoword reading [Word Attack subtest (Word Attack), WRMT-3; Woodcock 2011], timed word reading [Sight Word Efficiency subtest (SWE), *Test of Word Reading Efficiency, 2nd edition* (TOWRE-2); Torgesen et al. 2012], and timed pseudoword reading [Phonemic Decoding Efficiency subtest (PDE), TOWRE-2; Torgesen et al. 2012]. For all four subtests, Form A was administered at pre-test, and Form B was administered at post-test to avoid practice/familiarity effects. High alternate form reliability has been reported for standardized tests scores on both the WRMT-3 subtests (Word ID \( r = 0.93 \), Word Attack \( r = 0.76 \); Woodcock 2011) and the TOWRE-2 subtests (SWE \( r = 0.90 \), PDE \( r = 0.92 \); Torgesen et al. 2012). Thus we report changes in standard scores, because changes in raw scores are difficult to interpret. The primary outcome measure was a composite reading score obtained by averaging the standard scores from all four subtests.

Confirming inclusion criteria from screening, all participants either (1) scored below the 25th percentile on at least 2 of the 4 reading subtests (\( n = 52 \)), and/or (2) possessed a discrepancy of 15 or more standard points between the reading composite score and the non-verbal cognitive ability score (\( n = 43 \)). Thirty participants met both descriptions. There were no
demographic differences (including age, grade, gender, bilingual status, diagnoses, and SES) between these two descriptions (all $|t| < 1.15$, all $\chi^2 < 1.71$, all $p > 0.17$).

**Group Assignment**

After all pre-test assessments were completed, the fifty children recruited from the community at-large were randomly assigned to either receive an intensive summer reading intervention ($n = 25$) or to a waiting-list control group ($n = 25$), who received equal access to services after post-test assessments. For the intervention-assigned participants in this community sample, intervention was based in Cambridge, MA in dedicated space at MIT (“Site 1”). Children recruited from the local partner school ($n = 15$), were all assigned to the intervention group as a condition of school participation, and instruction was delivered on-site at the school (“Site 2”). The overall intervention group was therefore oversubscribed with 15 non-randomly assigned students, which allowed for better investigation of individual differences in response to treatment. After pre-test, one participant from the community at-large who had been randomly assigned to the intervention did not continue study participation, leaving 39 participants in the intervention group.

Intervention and control groups did not differ significantly on any demographic or assessment measures, including age, grade, gender, portion with comorbid ADHD, bilingualism, SES, nonverbal cognition, vocabulary, and all reading skills (all $|t| < 0.76$, all $\chi^2 < 1.93$, all $p > 0.16$). Within the full treatment group ($n = 39$), there was a marginal difference in SES by site [$t(37) = 1.87$, $p = 0.07$], which was driven by one outlier from the partner school with an SES 2.6 standard deviations below the sample mean. If excluded, no significant SES difference remained between assignment sites [$Site 1 M = 48$, $SD = 12.6$, $Site 2 M = 42$, $SD = 11.4$, $t(36) < 1.50$, $p > 0.14$]. There were no differences between sites on any other demographics (age, grade, gender, ADHD, bilingualism), pre-test reading scores (all $|t| < 0.26$, all $\chi^2 < 0.83$, all $p >$
0.36), or intervention response (see Results below). Thus, participants from both sites who completed the intervention were combined into a single treatment group ($n = 39$).

**Intervention**

The intervention is described in detail in a prior publication (Christodoulou et al. 2015). In brief, intervention participants (at both sites) received an intensive version of the Lindamood-Bell *Seeing Stars: Symbol Imagery for Fluency, Orthography, Sight Words, and Spelling* (Bell 2007) program in small groups (3-5 children) by trained Lindamood-Bell teaching staff. *Seeing Stars* is a multisensory remedial approach with a primary focus on training orthographic and phonological processing to improve reading accuracy, fluency, and comprehension. The program was held 4 hours per day, 5 days per week, for 6 weeks during the summer break from school, with a high rate of attendance ($M = 113$ total hours, $SD = 7.5$). Total number of hours of attendance was not correlated with any demographic variable, intervention site, pre-test or post-test assessment score, or treatment response (all $|r| < 0.23$, all $p > 0.20$).

**Neuroimaging Data Acquisition**

Participants completed neuroimaging sessions at pre-test and post-test. First, children were acclimated to the MRI environment and practiced lying still in a mock MRI scanner. Data were then acquired on a 3 Tesla Siemens MAGNETOM Trio Tim scanner equipped for echo planar imaging (EPI; Siemens, Erlangen, Germany) with a 32-channel phased array head coil. First, an automated scout image was acquired, and shimming procedures were performed to optimize field homogeneity. Then a whole-head, high-resolution T1-weighted multi-echo MPRAGE (Van Der Kouwe et al. 2008) structural image was acquired using a protocol optimized for movement-prone pediatric populations ($TR = 2530$ ms, $TE = 1.64$ ms, FoV = 220mm, and flip angle = 7°; yielding 176 slices with 1-mm isotropic resolution (Tisdall et al. 2012). All neuroimaging took place at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research,
Assessment Timeline

Behavioral testing and MRI scanning took place on two separate days to avoid child fatigue. All pre-test behavioral assessments occurred within the 5 weeks prior to the start of the intervention ($M = 18$ days prior to start of intervention, $SD = 12$ days). Given constraints of MRI availability during early summer, baseline neuroimaging occurred over a longer timespan within the 10 weeks prior to the start of intervention ($M = 39$ days prior to start of intervention, $SD = 19$ days). There were no differences in the timing of pre-test assessments or scanning between intervention and control groups ($t(59) = 0.82$, $p > 0.4$) nor was there a relationship with any demographic variable (all $|r| < 0.19$, all $p > 0.14$). Similarly, all post-test assessments (both behavioral and MRI scanning) occurred within the 6 weeks immediately following the conclusion of the intervention (behavioral: $M = 15$ days after intervention conclusion, $SD = 8$ days; MRI: $M = 11$ days after intervention conclusion, $SD = 8$ days). The average time difference between pre and post behavioral assessments was 2.16 months ($SD = 0.32$), and the average time difference between pre and post MRI scanning was 2.71 months ($SD = 0.69$). Again, there was no relationship between date of post-testing and any demographic variable (all $|r| < 0.13$, all $p > 0.3$). Unintentionally, intervention and control groups differed marginally in the timespan between intervention conclusion and post-test MRI scanning ($t(56) = 1.99$, $p = 0.052$), although the average difference between groups was only 4.3 days (intervention group $M = 11.4$ days after conclusion; control group $M = 15.7$ days after conclusion), which is a negligible amount of time for confounding cortical changes to occur. However, to ensure correction for potential timing differences, the time interval between intervention conclusion and post-scanning was added as a nuisance variable to between group longitudinal cortical thickness analysis, which did not affect results.
Behavioral Analyses

Change scores were computed individually for each of the four assessments chosen \textit{a priori} as outcome measures. Additionally, a composite change score was computed subtracting the average pre-test standard score from the average post-test standard score. Repeated measures ANOVAs were used to determine the group effect of the reading intervention, and multiple regressions were used to determine which participant-level factors were associated with treatment response.

Structural Image Analyses

T1-weighted images were visually inspected for image quality. Two trained observers, who were blind to participant SES and behavioral measures, rated each image on a scale of 1 (perfect) to 5 (unusable) based on a visual guide of artifacts associated with motion created in-house. If ratings differed, the two observers discussed their ratings until a consensus was reached. Three participants were excluded from pre-test neuroimaging analyses because of poor image quality (pre-test \( n = 62 \)), and six additional participants (3 from each assigned group) had unusable images at post-test. The remaining 55 participants (19 waiting control, 36 intervention (18 each treatment responders/non-responders)) had images of equivalent quality at both time points, which is necessary for accurate measurement of cortical changes. Quality ratings were not correlated with SES, any behavioral measures, or intervention group (all \( p > 0.2 \)).

Cortical reconstruction was conducted with FreeSurfer Version 5.3.0 (Fischl 2012). First, a semi-automated processing stream (recon-all) with default parameters completed motion and intensity correction, surface-based registration, spatial smoothing, subcortical segmentation, and parcellation of cortical white and gray matter boundaries. Pial and white matter surfaces were then manually edited as needed. An observer blind to participant SES and behavioral measures confirmed the accuracy of the final surfaces.
All T1 images from both time points were resampled to a standard brain (fsaverage) and smoothed with a 10-mm full-width half-maximum kernel. Cortical thickness was defined at each location as the distance between the white and pial surfaces (Dale et al. 1999; Fischl and Dale 2000). To examine cross-sectional differences at the pre-test time point, general linear models were constructed to test the whole brain for correlations between cortical thickness and SES, with participant gender and age as nuisance variables. Whole-brain analyses were corrected for multiple comparisons using a Monte Carlo simulation with 10,000 repetitions and Bonferroni adjusted for both hemispheres (cluster-forming $p < .05$, cluster-wise $p < .05$; Hagler et al. 2006).

Volumetric analyses were conducted on the 35 parcellations of the Desikan-Killiany Atlas (Desikan et al. 2006) automatically segmented in FreeSurfer. All volumetric analyses were controlled for gender, age, and estimated intracranial volume (ICV; Buckner et al. 2004) and Bonferroni-corrected for multiple comparisons.

Both T1 images from all participants with two usable images were processed with FreeSurfer’s longitudinal stream (Reuter et al. 2012). This process estimates average participant anatomy by creating an unbiased within-participant template space (Reuter and Fischl 2011) using a robust, inverse consistent registration (Reuter et al. 2010). After all templates were manually edited and checked (as above), information from both the templates and individual T1 images were combined to calculate longitudinal changes in individual anatomy, and surfaces were again resampled to a standard brain and smoothed with a 10-mm full-width half-maximum kernel. General linear models were constructed with symmetrized percent change (SPC) as the dependent variable and controlled for gender. SPC is the rate of change at each location with respect to the average thickness across both time points. This approach is more robust than rate of change or simple percent change, which refer to change only in terms of the first measurement. Whole-brain analyses were cluster-corrected for multiple comparisons using a
Monte Carlo simulation with 10,000 repetitions and were Bonferroni adjusted for both hemispheres (cluster-forming $p < .05$, cluster-wise $p < .05$; Hagler et al. 2006).

Results

Relation of SES to Behavioral Measures at Pre-Test
At pre-test, participants scores on tests of phonological awareness, phonological memory, and rapid naming ranged from average to below-average (Table 1). Single word and pseudoword reading skills clustered at borderline low average to below average scores. Higher SES correlated significantly with higher scores on vocabulary (PPVT-4, Pearson’s $r = 0.37$, $p = 0.002$) and marginally with higher nonverbal cognitive ability scores (KBIT-2, $r = .023$, $p = 0.065$), despite these mean standard scores being within or above the average range. SES was not correlated with scores on any subtests assessing phonological awareness, phonological memory, or rapid naming (all $|r|< 0.08$, all $p > 0.50$). Higher SES was only correlated with higher scores on one of the four single-word reading subtests (WRMT-3 Word Attack: $r = 0.26$, $p = 0.036$; all other reading subtests $r < 0.17$, $p > 0.2$), and consequently was marginally correlated with higher reading-composite scores ($r = 0.24$, $p = 0.05$). Neither of these SES-reading relationships retained significance when controlling for KBIT-2 scores (both $r < 0.2$, both $p > 0.1$). In contrast, when controlling for reading scores, SES and vocabulary maintained a relationship as strong as the zero-order correlation ($r = 0.36$, $p = 0.003$).

Relation of SES to Cortical Thickness and Reading Scores at Pre-Test
Confirming our hypothesis, higher SES correlated significantly with greater pre-test cortical thickness in several clusters spanning both hemispheres (Figure 1, Table 2). In the left hemisphere, these clusters included parts of (1) pars opercularis (the posterior portion of Broca’s area), (2) supramarginal and postcentral regions, and (3) insula, transverse temporal
gyrus, and superior and middle temporal regions. In the right hemisphere, significant clusters included (1) middle and superior temporal regions, (2) supramarginal and postcentral regions, (3) lateral occipital/fusiform regions, and (4) paracentral regions (see Supplementary Figure 1 for scatterplots by region). Nearly identical clusters emerged when additionally controlling for composite reading score. While cortical thickness and SES showed significant associations, cortical thickness was not correlated with any individual reading assessment scores or the reading composite score.

Although the smallest by area, the left opercular cluster’s cortical thickness exhibited the strongest correlation with SES \((p = 10^{-6})\). Using the pre-defined cortical parcellations, higher SES also correlated significantly with greater volume of the entire left pars opercularis (lpOp, partial \(r = 0.33, p = 0.005\)). Given that SES was also strongly correlated with receptive vocabulary scores, we undertook a mediation analysis (Figure 2). By adding lpOp to the regression model, the relationship between SES and vocabulary scores was rendered insignificant, indicating a full mediation. To confirm, a bootstrapping method with 10,000 iterations (Hayes 2013) was employed. There was a significant indirect effect of SES on vocabulary score through lpOp volume, [indirect effect = 0.15, bootstrapped 95% CI (0.06, 0.29), indirect/total effect = 0.37]. This indicates that the volume of the left pars opercularis could account for 37% of the total effect of SES on vocabulary scores.

**Effect of Remediation Program on Reading Scores**

When examining changes on behavioral assessments (i.e., response to intervention), repeated measures ANOVAs revealed group by time-point interactions on the composite reading score \([F(62,1) = 21.87, p < 0.001]\) and on 3 of the 4 reading subtests (meeting a Bonferroni-adjusted significance criterion), indicating a benefit of the intervention (Table 3, Figure 3). These included untimed word reading [WRMT-3 Word Identification, \(F(62,1) = 8.00, p = 0.006\], untimed
psedoword reading [WRMT-3 Word Attack, $F(61,1) = 10.97, p = 0.002$], and timed psedoword reading [TOWRE-2 Phonemic Decoding Efficiency, $F(56,1) = 12.27, p = 0.001$]. Post-hoc paired t-tests for all significant interactions revealed that children with RD who received intervention maintained their scores across time points (all $p > 0.6$), while children with RD in the waiting control group significantly declined (all $p < 0.001$). Both groups declined on the TOWRE-2 Sight Word Efficiency subtest (both $p < 0.005$). Overall, the relative benefit of the intervention was expressed as maintenance of scores for the intervention group relative to a loss of skills for the control group (see Christodoulou et al. 2015 for further information).

**Differences Between Children Who Responded More or Less to Intervention**

To examine variation within the intervention group, we classified participants based on the change in composite scores (Figure 4). Of the 39 participants who completed the intervention, approximately half had positive composite change scores ("responders": $n = 20, M = 4.28, SD = 3.41$), indicating pre-to-post improvement, and half had negative composite change scores ("non-responders": $n = 19, M = -4.91, SD = 4.74$). For comparison, the waiting control group had a mean change score of -6.72 ($SD = 4.02$). Independent t-tests revealed that non-responders did not differ from waiting controls on pre-to-post change scores for the composite or any subtest (all $p > 0.17$), whereas responders differed from both non-responders and waiting controls on all pre-to-post change scores with exception of the Sight Word Efficiency subtest from the TOWRE-2 (all $p < 0.006$; Table 3 and Figure 3).

Contrary to our hypothesis, responders had a significantly lower SES ($M = 39.9, SD = 12.7$) than non-responders [$M = 51.2, SD = 11.1$; $t(37) = 2.96, p = 0.005$; Figure 5a]. When children were divided by median SES, 14 of the 20 responders were in the lower-SES half, and 13 of the 19 non-responders were in the higher-SES half [$\chi^2(1, n = 39) = 5.76, p = .016$]. Treatment response was not significantly related to any other demographic variable, including age, grade,
gender, bilingualism, presence of an ADHD diagnosis and/or use of ADHD medication, vocabulary scores, non-verbal cognitive ability scores, hours of intervention attendance, timing of pre-test or post-test assessments, or whether participants met a low score or discrepancy inclusion criterion (all |t| < 1.4, all p > 0.17).

At baseline, responders also had significantly lower composite reading scores than non-responders [responders: \(M = 80.28, SD = 6.64\); non-responders: \(M = 87.54, SD = 10.01\); \(t(37) = 2.68, p = 0.01\)]. To control for the effect of baseline scores, a regression analysis was performed to examine the relative relations of all potential predictive variables to intervention response. A model including age, grade, gender, bilingual status, ADHD diagnosis, ADHD medication, which inclusion criterion was met, total hours of intervention attendance, intervention site, SES, and RD severity (reverse of pre-test composite score) revealed that only SES and RD severity were significant predictors of binary-coded improvement, with SES explaining 26% of the variance in improvement \((\beta = -0.019, p < 0.003;)\) and RD severity explaining 14% of the variance in improvement \((\beta = 0.022, p < 0.05)\). Removing all non-significant predictors yielded the same pattern (SES: \(\beta = -0.015, p = 0.01, R^2 = 0.17\); RD severity: \(\beta = 0.019, p = 0.02, R^2 = 0.14\)), and these results held when using change scores as the dependent variable instead (Figure 5b; SES: \(\beta = -0.164, p = 0.02, R^2 = 0.14\); RD severity: \(\beta = 0.248, p = 0.01, R^2 = 0.16\)). These findings indicate that both more severe RD and lower SES, two risk factors, were independently associated with greater response to intervention.

Analogous results were seen in pre-to-post cortical thickness changes. On average, there were no significant differences in thickness changes between the intervention and waiting control groups. However, there were large differences in thickness changes within the intervention group. Responders exhibited significantly greater thickening than non-responders bilaterally in
several large clusters spanning (1) middle/inferior temporal regions (extending into fusiform region on the right), (2) supramarginal/angular regions, (3) precentral regions, and (4) paracentral/posterior cingulate regions (Figure 6, Table 4, see Supplementary Figure 2 for group differences by region). An additional cluster spanned a large portion of the right superior temporal gyrus extending into insula. The greatest longitudinal between-group difference occurred in the left middle temporal cluster, where responders’ cortices thickened by an average of 31 micrometers (μm) per month (1% gain), and non-responders’ cortices thinned by an average of 11 μm per month (0.37% loss). For comparison, the waiting control group on average exhibited 7 μm of thinning per month (0.26% loss) in this region, although this thinning was not statistically significant. There were no clusters in which non-responders exhibited greater thickening or thinning than the waiting-control group. When the three participant groups (responders, non-responders, and controls) were analyzed separately, responders exhibited significant thickening over most of the cortical surface, whereas non-responders and controls exhibited no regions of significant thickening or thinning.

Also commensurate with behavioral results, lower SES and greater RD severity were independently correlated with cortical thickening in neighboring but non-overlapping regions. Lower SES (controlling for RD severity) correlated with greater thickening in the bilateral middle temporal and paracentral/cingulate regions, as well as left precentral and right lateral orbitofrontal/pars orbitalis regions (Figure 7 cool colors, Table 5). Greater RD severity (controlling for SES) correlated with greater thickening in a right lateral occipital cluster (Figure 7 warm colors, Table 5). To further evaluate whether the apparent neuroanatomical dissociations in cortical thickening related to lower SES and greater RD severity were independent as opposed to being secondary to statistical thresholding, we examined in several main clusters the correlations between changes in cortical thickness and both baseline SES and RD. These analyses supported the conclusion that regional changes in cortical thickness were related
distinctly to either SES or RD (Supplementary Figure 3). There were no significant correlations between thickness changes and SES or RD severity in the waiting control group.

**Discussion**

The present study yielded three novel discoveries about the relations between SES and RD, including behavioral and neuroanatomical responses to reading intervention. First, among a group of children with RD, higher SES was associated with thicker cortex in multiple neocortical regions, including bilateral perisylvian and supramarginal regions associated with language and reading; this extends, for the first time, the well-documented SES-neuroanatomy relationship to children with RD. Moreover, the strongest correlation occurred in Broca’s area in left inferior frontal cortex, the volume of which fully mediated the relationship between SES and vocabulary, commonly known as the “vocabulary gap.” Second, whereas children who did not receive intervention or who did not respond to intervention exhibited no significant cortical changes, children who responded to intervention (i.e., whose reading improved) exhibited pre-to-post thickening of cortex across broad bilateral occipitotemporal and temporoparietal regions, most notably in the middle temporal gyri. Third, children from lower-SES families and children with more severe RD were more likely to benefit from the intervention than children from higher-SES families or children with less severe RD, both behaviorally and neurally.

**Relation of SES, Cortical Thickness, and Vocabulary in Children with RD**

In accordance with our hypothesis, higher SES (independent of RD severity) was associated with thicker cortex in bilateral perisylvian and supramarginal regions at pre-test, with the strongest association in left pars opercularis (posterior portion of Broca’s area). Prior studies have reported similar relations between SES and neuroanatomical characteristics of language
cortex in typically developing children (see Brito and Noble 2014 for review). However, this is the first study to demonstrate the same relationship among children with RD. The combined influences of RD and lower SES on neuroanatomy may make children especially vulnerable for academic challenges in reading.

Further, there was a strong relation between higher SES and larger vocabulary, a finding consistent with many studies reporting striking relations between SES, home language environment, and vocabulary (e.g., Hart and Risley 1995; Hoff 2003; Fernald et al. 2013). The present study offers an initial insight into a brain mechanism that may be involved in the relation of childhood SES to vocabulary, albeit specifically in children with RD. A mediation model revealed that the volume of the left pars opercularis (the posterior portion of the canonical Broca’s area) fully mediated the strong correlation between SES and vocabulary size, accounting for over a third of the relation between SES and vocabulary.

The present findings are consistent with prior evidence associating SES, neuroanatomy, and performance on a measure of verbal academic achievement (Hair et al. 2015). That study reported that lobar gray matter volumes mediated the relation between growing up near or below the federal poverty line and broader linguistic achievement in children and young adults, with frontal-lobe volume explaining 11% of the income-related differences in language scores. The authors suggested that their results might underestimate the true effects of SES because of strict exclusion criteria that selected for a typical sample of participants and little variation in the other socio-educational domains comprising SES (Hair et al. 2015). Indeed, by defining SES continuously and more broadly within a sample of young children with RD, the present study found that a more specific portion of the left frontal lobe (pars opercularis) explained 37% of the effect of SES on children’s vocabulary knowledge. This relation between Broca’s area and vocabulary is consistent with the known major role of Broca’s area to language development.
(Hagoort 2005; Grodzinsky and Santi 2008; Hagoort 2014). This not only proposes a more focal locus for the previous study’s whole-brain global effects, but also suggests an even tighter relationship between SES, brain, and achievement in a group of children with RD.

Despite notable SES-related variation in vocabulary and the neural structure of core language regions, there were little SES-related behavioral differences in children's reading profiles, including word and pseudoword reading accuracy and fluency, as well as reading-related phonological processing and rapid naming skills. This may in part reflect a restricted range of poor reading ability for all of the children, who all had to have evidence for substantial RD regardless of SES. The similarity of reading scores across SES makes more salient the differences in vocabulary and in neuroanatomy.

**Individual Differences in Intervention Response and Brain Plasticity**

We found considerable variation in response to intervention, with about half of the children with RD exhibiting significant gains in reading after intervention, and about half exhibiting essentially no gains (i.e., they did not differ significantly from the children with RD who received no intervention). Although intervention programs are typically evaluated on the basis of an average, overall response, frequently some portion of children with RD (anywhere from 3-80%, depending on response criteria) fail to respond to intervention programs that are effective for other children (for reviews, see Al Otaiba and Fuchs 2002; Nelson et al. 2003; Lam and McMaster 2014). The present finding of 50% non-response is consistent with previous studies of similar-age children with persistent reading difficulties (Al Otaiba and Fuchs 2002; Lam and McMaster 2014).

There were striking developmental differences in brain plasticity between the children with RD who did respond to the intervention versus the other two groups of children with RD who either
did not respond to the intervention or who received no intervention. Responders exhibited
greater cortical thickening across broad bilateral occipitotemporal and temporoparietal regions.
The greatest group difference was evident in the middle temporal gyrus, where responders’
cortices *thickened* by an average of 31 μm/month, and non-responders’ and waiting controls’
cortices *thinned* by 11 and 7 μm/month, respectively. For reference, this region thins by an
average of 5 μm/month in typically developing, similarly aged children (Sowell et al. 2004). This
suggests that non-responders and waiting controls exhibited a typical cortical trajectory of
developmental thinning during this study, whereas children with RD who responded by
improving their reading exhibited a noteworthy thickening of cortex.

Two other studies have examined neuroanatomical plasticity, one in gray matter and one in
white matter, following reading intervention with children. Our left hemisphere findings are
consistent with a study that used the same *Seeing Stars* intervention in 11 children ages 7-11
years and found increased gray matter volume in left occipitotemporal and medial parietal
regions (Krafnick et al. 2011). Another study reported intervention related changes in white
matter microstructure in children ages 8-12 (Keller and Just 2009). A difference between the
prior and present studies is that only the present study reports a specific relation between
individual differences in treatment response and structural plasticity.

Several prior studies have reported functional brain differences between individuals with RD
who did or did not respond to intervention at either a single pre-intervention (Odegard et al.
2008; Davis et al. 2011; Farris et al. 2011; Molfese et al. 2013) or post-intervention time point
(Rezaie et al. 2011a, 2011b), although few have reported longitudinal neural changes. One
study using magnetic source imaging (MSI) found that responders, but not non-responders,
exhibited increased duration of activity and a shift in activation timing in a broad left
temporoparietal region during a phonological decoding task, such that their neural profiles
matched typical readers post-intervention (Simos et al. 2007). Another study used evoked response potentials (ERPs) during a German phonological lexical decision task to compare functional plasticity between responders and non-responders (Hasko et al. 2014). Treatment responders, but not non-responders, exhibited an increase in the post-test amplitude of the N400 component, thought to underlie orthographic processing. Although locating the source of ERP components is difficult, the N400 is thought to arise from bilateral superior/middle temporal gyri and temporoparietal regions (Kutas and Federmeier 2011). Thus, the prior functional studies align with the present structural study in suggesting that plasticity in temporoparietal regions distinguishes children with RD who do versus do not respond to specific interventions.

**Relation of SES and RD Severity to Intervention Response and Plasticity**

Contrary to our hypotheses, both lower SES and greater RD severity at baseline were independently associated with greater response to intervention in regard to both reading ability and brain plasticity. Importantly, because analysis models controlled for baseline reading scores, this result cannot solely be attributed to a regression-to-the-mean explanation. SES and RD severity, however, appeared to have differential relations between reading gains and structural plasticity, suggesting that the two factors influenced treatment response via cortical growth in different brain regions. Children from lower SES families exhibited greater thickening across broad bilateral occipitotemporal regions, largely corresponding with the left hemisphere reading network and its right hemisphere homologues. Children with more severe RD exhibited greater thickening in a right lateral occipital region that may provide compensatory support for the visual component of reading.

The finding that children with lower SES and more severe RD responded more strongly to this specific intervention is a notable difference from prior studies. Most reading intervention studies examining these factors have largely found that lower-SES children (Torgesen et al. 1999;
Hatcher et al. 2006; Morris et al. 2012) and children with lower word-reading and decoding skills (Hatcher et al. 2006; Vellutino et al. 2007; Compton et al. 2012) tended to exhibit a worse response to interventions. However, these studies largely utilized in-school remediation programs focused on phonological awareness with short instructional sessions distributed across many weeks during the academic year, whereas the present study employed an intensive, short-term intervention with a small teacher to student ratio (1:3-5) during the non-academic summer.

Several interpretations are possible for the greater effect of the intervention on lower-SES than higher-SES children with RD. One possibility concerns the nature of the present intervention; specifically, the pronounced focus on visual and orthographic imagery. Given that lower-SES, above-average readers exhibit greater white matter tract coherence in the right inferior longitudinal fasciculus (Gullick et al. 2016), which supports visuospatial processing, it is possible that this visual approach stimulated greater neural plasticity in right hemisphere areas, and, in turn, a more positive treatment response. Another possibility is that the combination of intervention intensity, duration, and small group size (all of which predict greater response frequency (Denton 2012), was particularly potent for lower-SES children. A limitation on interpretation of these findings is that the waiting-control group served as a passive control condition, thus precluding the separate effects of intensive, small-group attention and interaction from the specific academic content of the intervention.

In any case, another possible explanation concerns the timing of the intervention. The particular benefits of the reading intervention during summer for the lower-SES children with RD may be related to evidence that lower-SES children in general are vulnerable to academic regression during the summer, a phenomenon known as “summer slump” or “summer slide.” During the summer months, lower-SES students tend to regress in their reading skills, while higher-SES
students tend to maintain or gain reading skills (Cooper et al. 1996; McCoach et al. 2006; Alexander et al. 2007). This is frequently attributed to decreased access to books and reduced experiences with or emphasis on literacy in the homes of lower-SES children. Thus it is plausible that the present access to an intensive reading program provided the lower-SES children with precisely the literacy access they would otherwise be missing, and presumably had missed in previous summers.

Finally, the better intervention response for lower-SES participants could be related to variations in RD etiology. Reading deficits can occur for many reasons, and it is possible that the origins of RDs could vary in relation to SES. Differences in environmental factors, such as home literacy, access to reading material, and school quality may be responsible for systematic heterogeneity in the root cause of RD across children from varying SES (Ursache and Noble 2016). Consistent with this possibility is the finding that higher-SES and lower-SES environments interact differently with genetic factors related to RD (Friend et al. 2008; Mascheretti et al. 2013). A large study of twins with reading difficulty revealed differences in the heritability of reading disability across SES, such that environmental factors accounted for more of the variance in reading deficits in children from lower-SES families than higher-SES families (Friend et al. 2008; but see Kirkpatrick et al. 2011 for conflicting findings). This raises the possibility that the neurobiological bases of RD could vary with SES. In such a case, environmentally-driven neurobiological differences in children from lower-SES families may be more amenable to an (environmental) treatment intervention. In contrast, genetically-driven neurobiological differences in children from higher-SES families may be more resistant to treatment intervention. By this view, children with RD of varying SES could respond differently to interventions due to variation in environmental versus genetic contributions to the etiologies of their behaviorally similar RD.
Several limitations of the present study are noted. One limitation involves the non-random treatment group assignment of 15 participants, as a condition of school participation. Ideally all participants would have been randomly assigned to the treatment or control group, so that the post-intervention response can only be explained by the intervention itself. It is, however, unlikely that alternative participant characteristics (such as school or intervention site) contributed to either between-group or within-group treatment differences, because participants were similar across sites in their demographics, assessment scores, and treatment response. A second limitation is that we lacked information to characterize the quality of school reading instruction for all participants. On average, it is likely that lower-SES children may receive less high-quality instruction in lower performing schools, though this was not evaluated in our current study. In order to promote SES diversity of children in this study, the children attended a wide variety of public, private, and public-charter schools from a large metro region. Future studies may attempt to control for this by enrolling participants from a single school with a SES-diverse population, or characterizing the quality of reading instruction in diverse schools. Third, we could not dissociate potentially separable effects of SES dimensions of parental education, parental occupation, and income on any outcomes. There is evidence for dissociations among these dimensions on both behavioral (e.g., Duncan and Magnuson 2012) and neural (e.g., Brito and Noble 2014) child outcomes. In our sample, parental education, parental occupation, and parental income were highly correlated, which precluded any dissociations. Future research with larger, more diverse participant samples will be required to untangle these correlated dimensions of SES when considering treatment response in children with RD.

In summary, this study investigated how the brain structure of young students with RD varies by SES, and explored SES-related differences in their behavioral and neural response to intervention. Despite reduced cortical thickness in canonical language regions at baseline, lower-SES children responded more favorably to an intensive summer reading invention than
their higher-SES peers, both in terms of reading scores and structural plasticity throughout the neural reading networks. Taken as a whole, this suggests that intensive summer reading intervention might be even more effective for these dually at-risk children.
**Funding**

This work was supported by the Ellison Medical Foundation (to J.D.E.G.); the Halis Family Foundation (to J.D.E.G.); Lindamood-Bell Learning Processes (which supported the intervention); and the National Institutes of Health (T32-DC000038 and F31-HD086957 to R.R.R.).
Acknowledgements

We thank the participants and their families. We thank the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research (MIT) and Atshusi Takahashi, Steve Shannon, and Sheeba Arnold for data collection technical support. We thank Allyson Mackey for assistance with cortical thickness analyses, and we thank Camila Caballero for comments on the manuscript.

Address correspondence to:

Rachel Romeo, MIT (Office 46-4037), 43 Vassar Street, Cambridge, MA 02139, USA,
rromeo@mit.edu

or Joanna Christodoulou, 36 First Avenue, Boston, MA 02129, USA,
jchristodoulou@mghihp.edu
References


Hecht SA, Burgess SR, Torgesen JK, Wagner RK, Rashotte CA. 2000. Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade:
the role of phonological awareness, rate of access, and print knowledge. Read Writ 12:99-127.


Torgesen JK, Wagner R, Rashotte C. 2012. Test of word reading efficiency, 2nd ed. Austin, TX: PRO-ED.


### Tables

**Participant Test Scores and Relation to Socioeconomic Status**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>M (SD)</th>
<th>Zero-order Correlation with SES</th>
<th>Partial Correlation with SES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonverbal Cognition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KBIT-2 Matrices</td>
<td>103.09 (14.05)</td>
<td>$r = 0.23^\dagger$</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Oral Language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT-4 (Receptive Vocabulary)</td>
<td>107.03 (12.50)</td>
<td>$r = 0.37^{**}$</td>
<td>$r = 0.32^{**}$</td>
</tr>
<tr>
<td><strong>Phonological Awareness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTOPP Elision</td>
<td>8.49 (2.02)</td>
<td>$r = 0.02$</td>
<td>$r = -0.02$</td>
</tr>
<tr>
<td>CTOPP Blending Words</td>
<td>10.12 (2.27)</td>
<td>$r = -0.08$</td>
<td>$r = -0.13$</td>
</tr>
<tr>
<td><strong>Phonological Memory</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTOPP Nonword Repetition</td>
<td>7.92 (1.18)</td>
<td>$r = -0.02$</td>
<td>$r = -0.07$</td>
</tr>
<tr>
<td>CTOPP Memory for Digits</td>
<td>9.23 (2.33)</td>
<td>$r = 0.05$</td>
<td>$r = 0.02$</td>
</tr>
<tr>
<td><strong>Rapid Automatized Naming</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAN/RAS Objects</td>
<td>89.79 (13.65)</td>
<td>$r = 0.02$</td>
<td>$r = 0.07$</td>
</tr>
<tr>
<td>RAN/RAS Letters</td>
<td>95.33 (11.49)</td>
<td>$r = 0.08$</td>
<td>$r = 0.09$</td>
</tr>
<tr>
<td>RAN/RAS 2-Set Letters and Numbers</td>
<td>96.23 (11.90)</td>
<td>$r = -0.03$</td>
<td>$r = -0.04$</td>
</tr>
<tr>
<td><strong>Single Word/Nonword Reading Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRMT-3 Word Identification</td>
<td>85.11 (9.82)</td>
<td>$r = 0.13$</td>
<td>$r = 0.04$</td>
</tr>
<tr>
<td>WRMT-3 Word Attack</td>
<td>86.72 (11.69)</td>
<td>$r = 0.26^*$</td>
<td>$r = 0.15$</td>
</tr>
<tr>
<td><strong>Single Word/Nonword Reading Fluency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOWRE-2 Sight Word Efficiency</td>
<td>84.54 (10.67)</td>
<td>$r = 0.09$</td>
<td>$r = 0.05$</td>
</tr>
<tr>
<td>TOWRE-2 Phonemic Decoding Efficiency</td>
<td>80.80 (9.89)</td>
<td>$r = 0.17$</td>
<td>$r = 0.11$</td>
</tr>
<tr>
<td><em>Reading Composite</em></td>
<td>84.08 (8.71)</td>
<td>$r = 0.24^\dagger$</td>
<td>$r = 0.19$</td>
</tr>
</tbody>
</table>

$^\dagger p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$

**Table 1.** List of standardized assessments administered before intervention, participants’ average standard scores, and correlations with SES. Partial correlations between SES and reading scores control for standardized KBIT-2 scores. All assessments have a mean standard score of 100, with a standard deviation of 15, except CTOPP which has a mean scaled score of 10 and a standard deviation of 3. Note: KBIT-2 = *Kaufman Brief Intelligence Test, 2nd edition*, PPVT-4 = *Peabody Picture Vocabulary Test, 4th edition*, CTOPP = *Comprehensive Test of Phonological Processing*, RAN/RAS = *Rapid Automatized Naming and Rapid Alternating*

### Pre-test Correlations between Socioeconomic Status and Cortical Thickness

<table>
<thead>
<tr>
<th>Region of Cluster</th>
<th>Brodmann Areas</th>
<th>Area of cluster (mm²)</th>
<th>Peak Significance (–log₁₀ p)</th>
<th>Peak MNI Coordinates</th>
<th>Cluster-wise p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left pars opercularis</td>
<td>44</td>
<td>916.13</td>
<td>6.262</td>
<td>-46.9 12.0 18.2</td>
<td>0.04547</td>
</tr>
<tr>
<td>Left supramarginal + postcentral</td>
<td>40, 3, 1, 2</td>
<td>2581.8 6</td>
<td>4.782</td>
<td>-53.5 -42.4 45.6</td>
<td>0.00020</td>
</tr>
<tr>
<td>Left insula + superior/middle temporal</td>
<td>41, 42, 21, 22</td>
<td>1710.6 2</td>
<td>4.056</td>
<td>-36.1 -15.6 11.3</td>
<td>0.00020</td>
</tr>
<tr>
<td>Right middle/superior temporal</td>
<td>21, 22</td>
<td>1927.5 7</td>
<td>4.134</td>
<td>56.0 1.1 -29.4</td>
<td>0.00020</td>
</tr>
<tr>
<td>Right supramarginal + postcentral</td>
<td>40, 3, 1, 2</td>
<td>2486.6 3</td>
<td>3.388</td>
<td>44.7 -17.7 20.0</td>
<td>0.00020</td>
</tr>
<tr>
<td>Right lateral occipital + fusiform</td>
<td>18, 19, 37</td>
<td>1526.0 7</td>
<td>3.042</td>
<td>35.2 -80.2 -12.0</td>
<td>0.00080</td>
</tr>
<tr>
<td>Right paracentral</td>
<td>4, 3, 1, 2</td>
<td>139722</td>
<td>2.951</td>
<td>9.4 -32.3 51.4</td>
<td>0.02504</td>
</tr>
</tbody>
</table>

*Table 2.* Regions where SES was significantly correlated with cortical thickness, controlling for age and gender. Note: MNI = Montreal Neurological Institute.
**Post-test Reading Scores**

<table>
<thead>
<tr>
<th>Post-Test Assessment</th>
<th>Waiting Control (n = 25)</th>
<th>Intervention Total (n=39)</th>
<th>Intervention Non-Responders (n=19)</th>
<th>Intervention Responders (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Identification</td>
<td>80.08 (8.22)</td>
<td>84.69 (9.76)</td>
<td>84.32 (10.22)</td>
<td>85.05 (9.55)</td>
</tr>
<tr>
<td>Change Score</td>
<td>-5.32 (6.26)***</td>
<td>-0.46 (6.97)</td>
<td>-4.32 (5.38)**</td>
<td>3.20 (6.39)*</td>
</tr>
<tr>
<td>Word Attack</td>
<td>79.96 (9.83)</td>
<td>87.67 (9.79)</td>
<td>87.16 (10.07)</td>
<td>88.15 (9.75)</td>
</tr>
<tr>
<td>Change Score</td>
<td>-7.32 (8.61)***</td>
<td>0.66 (9.81)</td>
<td>-3.58 (9.34)</td>
<td>4.89 (8.52)*</td>
</tr>
<tr>
<td>Sight Word Efficiency</td>
<td>79.16 (11.28)</td>
<td>80.34 (12.79)</td>
<td>79.44 (14.09)</td>
<td>81.15 (11.81)</td>
</tr>
<tr>
<td>Change Score</td>
<td>-5.48 (6.88)***</td>
<td>-4.32 (8.63)**</td>
<td>-7.06 (8.17)**</td>
<td>-1.85 (8.48)</td>
</tr>
<tr>
<td>Phonemic Decoding Eff.</td>
<td>73.35 (9.36)</td>
<td>79.87 (8.81)</td>
<td>78.44 (9.90)</td>
<td>81.15 (7.73)</td>
</tr>
<tr>
<td>Change Score</td>
<td>-8.90 (6.12)***</td>
<td>-0.22 (10.37)</td>
<td>-5.67 (9.66)*</td>
<td>4.95 (8.30)*</td>
</tr>
<tr>
<td>Reading Composite</td>
<td>78.13 (8.53)</td>
<td>83.62 (8.55)</td>
<td>82.63 (9.33)</td>
<td>84.56 (7.86)</td>
</tr>
<tr>
<td>Change Score</td>
<td>-6.72 (4.02)***</td>
<td>-0.20 (6.17)</td>
<td>-4.91 (4.74)***</td>
<td>4.28 (3.41)***</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01, ***p < 0.001

Table 3. Group means (and standard deviations) of reading assessment standard scores at post-test. All assessments have a mean standard score of 100, with a standard deviation of 15. Change scores are post-test minus pre-test scores, averaged across participants, with indicated significance from paired t-tests. Word Identification and Word Attack are subtests of the **Woodcock Reading Mastery Test, 3rd edition** (WRMT-3). Sight Word Efficiency and Phonemic Decoding Efficiency are subtests of the **Test of Word Reading Efficiency, 2nd edition** (TOWRE-2). Reading Composite is the average of standard scores on all four subtests.
### Cortical Thickness Changes in Treatment Responders vs. Non-Responders

<table>
<thead>
<tr>
<th>Region of Cluster</th>
<th>Approximate Brodmann Areas</th>
<th>Area of cluster (mm²)</th>
<th>Peak Significance (−log₁₀ p)</th>
<th>Peak MNI Coordinates x</th>
<th>y</th>
<th>z</th>
<th>Cluster-wise p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left middle/inferior temporal</td>
<td>21, 37</td>
<td>1548.41</td>
<td>5.359</td>
<td>-60.1</td>
<td>-29.2</td>
<td>-12.2</td>
<td>0.00020</td>
</tr>
<tr>
<td>Left supramarginal</td>
<td>40</td>
<td>1401.52</td>
<td>3.035</td>
<td>-56.3</td>
<td>-24.6</td>
<td>27.5</td>
<td>0.00020</td>
</tr>
<tr>
<td>Left precentral</td>
<td>4</td>
<td>691.94</td>
<td>3.070</td>
<td>-31.7</td>
<td>-12.9</td>
<td>57.6</td>
<td>0.01514</td>
</tr>
<tr>
<td>Left paracentral + cingulate</td>
<td>4, 3, 1, 2, 31, 24</td>
<td>773.27</td>
<td>3.857</td>
<td>-7.5</td>
<td>-27.0</td>
<td>53.4</td>
<td>0.00619</td>
</tr>
<tr>
<td>Right middle/inferior temporal + fusiform</td>
<td>21, 37, 19</td>
<td>3302.67</td>
<td>4.546</td>
<td>47.8</td>
<td>-59.7</td>
<td>3.1</td>
<td>0.00020</td>
</tr>
<tr>
<td>Right supramarginal + angular</td>
<td>39, 40</td>
<td>926.53</td>
<td>3.490</td>
<td>59.0</td>
<td>-42.5</td>
<td>17.1</td>
<td>0.00140</td>
</tr>
<tr>
<td>Right superior temporal + insula</td>
<td>22</td>
<td>1836.93</td>
<td>3.239</td>
<td>44.0</td>
<td>-33.7</td>
<td>-0.8</td>
<td>0.00020</td>
</tr>
<tr>
<td>Right precentral</td>
<td>4</td>
<td>730.29</td>
<td>3.095</td>
<td>22.1</td>
<td>-10.1</td>
<td>53.0</td>
<td>0.01236</td>
</tr>
<tr>
<td>Right paracentral + posterior cingulate</td>
<td>4, 3, 1, 2, 31</td>
<td>1544.20</td>
<td>4.753</td>
<td>7.4</td>
<td>-20.1</td>
<td>56.3</td>
<td>0.00020</td>
</tr>
</tbody>
</table>

**Table 4.** Regions exhibiting significant differences in cortical thickness changes between children whose reading scores improved after intervention versus children whose scores did not improve. Comparisons are controlled for age and gender. Note: MNI = Montreal Neurological Institute.
Cortical Thickness Changes associated with SES and RD severity

<table>
<thead>
<tr>
<th>Region of Cluster</th>
<th>Approximate Brodmann Areas</th>
<th>Area of cluster (mm²)</th>
<th>Peak Significance (-log₁₀ p)</th>
<th>Peak MNI Coordinates</th>
<th>Cluster-wise p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation with SES, controlling for RD severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left middle temporal (anterior)</td>
<td>21, 20</td>
<td>2705.29</td>
<td>-3.755</td>
<td>-64.5</td>
<td>-26.4</td>
</tr>
<tr>
<td>Left middle temporal (posterior)</td>
<td>21, 20</td>
<td>623.14</td>
<td>-3.311</td>
<td>-60.6</td>
<td>-54.9</td>
</tr>
<tr>
<td>Left precentral</td>
<td>4</td>
<td>774.27</td>
<td>-3.783</td>
<td>-33.0</td>
<td>-18.2</td>
</tr>
<tr>
<td>Left posterior cingulate + paracentral</td>
<td>31, 4, 5, 3, 1, 2</td>
<td>611.57</td>
<td>-3.345</td>
<td>-6.4</td>
<td>-17.7</td>
</tr>
<tr>
<td>Right middle/superior temporal</td>
<td>21, 22</td>
<td>3913.28</td>
<td>-5.150</td>
<td>52.4</td>
<td>-24.5</td>
</tr>
<tr>
<td>Right paracentral</td>
<td>11, 47</td>
<td>963.37</td>
<td>-3.770</td>
<td>8.4</td>
<td>-10.6</td>
</tr>
<tr>
<td>Right lateral orbitofrontal + pars orbitalis</td>
<td>4, 3, 1, 2</td>
<td>704.96</td>
<td>-3.671</td>
<td>37.5</td>
<td>30.0</td>
</tr>
<tr>
<td><strong>Correlation with RD severity, controlling for SES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right lateral occipital</td>
<td>18, 19</td>
<td>726.64</td>
<td>-5.205</td>
<td>41.2</td>
<td>-73.2</td>
</tr>
</tbody>
</table>

Table 5. Regions exhibiting significant correlations between changes in cortical thickness and SES (controlling for RD severity and gender) or with RD severity (controlling for SES and gender) among all children who received intervention. Note: MNI = Montreal Neurological Institute.
Figures

**Figure 1.** Correlation between SES and cortical thickness, controlling for age and gender. Colored regions exhibited significantly thicker cortex with higher SES at baseline. Outlines represent the cortical parcellations from the Desikan-Killiany gyral-based atlas.
**Figure 2.** Mediation model showing the effect of SES on vocabulary scores as mediated by the volume of the left pars opercularis. Solid arrows represent direct paths, whereas the dotted arrow represents the indirect (mediated) path. β coefficients represent standardized regression coefficients. Each regression controls for participant age and gender, and all models involving the left pars opercularis (lpOp) also control for head size (estimated intracranial volume). Thus, vocabulary is represented by raw scores on the *Peabody Picture Vocabulary Test, 4th edition* (PPVT-4), to avoid adjusting for age twice. *p < 0.05, **p < 0.01.
Figure 3. Pre-to-post changes in standard scores on reading composite and subtests and composite by group. Positive scores indicate a score increase, while negative scores indicate a score decrease. “Intervention Total” combines Intervention Non-Responders and Intervention Responders. Reading Composite is the average of standard scores on all four subtests. Word Identification and Word Attack are subtests of the Woodcock Reading Mastery Test, 3rd edition (WRMT-3). Sight Word Efficiency and Phonemic Decoding Efficiency are subtests of the Test of Word Reading Efficiency, 2nd edition (TOWRE-2). Error bars represent standard errors. **p < 0.01.
**Figure 4.** Histogram of pre-to-post changes in the composite reading score for all participants in the intervention group only (n = 39). Positive scores indicate a score increase and classification as an intervention “Responder,” while negative scores indicate a score decrease and classification as an intervention “Non-Responder.”
Figure 5. Relation between SES and response to treatment. (a) Boxplot of SES as a factor of treatment response. Intervention response (improvement) was operationalized as a positive change score when averaging standard scores on four reading subtests: WRMT-3 Word Identification and Word Attack, and TOWRE-2 Sight Word Efficiency and Phonemic Decoding Efficiency. (b) Partial residual plot showing the amount of improvement (change in composite reading score, controlled for baseline score) as a function of SES.
Figure 6. Regions where treatment responders exhibited significantly greater cortical thickening versus treatment non-responders following an intensive summer intervention, controlling for gender. Outlines represent the cortical parcellations from the Desikan-Killiany gyral-based atlas.
Figure 7. Regions exhibiting significant correlations between changes in cortical thickness and SES (cool colors) or RD severity (warm colors) among all children who received intervention, controlling for gender. Outlines represent the cortical parcellations from the Desikan-Killiany gyral-based atlas.
Supplementary Figure 1. Correlations between SES and cortical thickness at pre-test. Cortical thickness is controlled for gender and age in months, and then averaged over the area of each significant cluster. Each scatter plot contains the best-fitting linear regression line. Cluster heat maps are shown in Figure 1, and cluster statistics are shown in Table 1.
Supplementary Figure 2. Change in cortical thickness between pre-test and post-test in regions that exhibit significant differences between treatment responders ($n = 18$) and treatment non-responders ($n = 18$). Symmetrized percent change (SPC) is the rate of change at each vertex with respect to the average thickness across both time points, and is averaged over the area of each significant cluster. Cluster heat maps are shown in Figure 6, and cluster statistics are shown in Table 4.
Supplementary Figure 3. Partial regression plots showing the change in cortical thickness between pre-test and post-test as a function of SES and reading disability (RD) severity ($n = 36$ children who received intervention). Regressions additionally include gender as a nuisance variable. Each scatter plot contains the best-fitting linear regression line. Symmetrized percent change (SPC) is the rate of change at each vertex with respect to the average thickness across both time points, and is averaged over the area of each significant cluster. SES was significantly correlated, and RD severity uncorrelated, with SPC in each region presented with exception of the right lateral occipital cluster, in which the reverse was true. Cluster heat maps are shown in Figure 7, and cluster statistics are shown in Table 5.