Individual differences in decoding skill, print exposure, and cortical structure in young adults

Clinton L. Johns, Andrew A. Jahn, Hannah R. Jones, Dave Kush, Peter J. Molfese, Julie A. Van Dyke, James S. Magnuson, Whitney Tabor, W. Einar Mencl, Donald P. Shankweiler & David Braze

To cite this article: Clinton L. Johns, Andrew A. Jahn, Hannah R. Jones, Dave Kush, Peter J. Molfese, Julie A. Van Dyke, James S. Magnuson, Whitney Tabor, W. Einar Mencl, Donald P. Shankweiler & David Braze (2018): Individual differences in decoding skill, print exposure, and cortical structure in young adults, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2018.1476727

To link to this article: https://doi.org/10.1080/23273798.2018.1476727

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 22 May 2018.
Individual differences in decoding skill, print exposure, and cortical structure in young adults

Clinton L. Johns, Andrew A. Jahna, Hannah R. Jones, Dave Kush, Peter J. Molfese, Julie A. Van Dyke, James S. Magnuson, Whitney Tabor, W. Einar Mencl, Donald P. Shankweiler, and David Braze

Haskins Laboratories, New Haven, USA; Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, USA; Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway; Department of Health and Human Services, Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institutes of Mental Health, National Institutes of Health, Bethesda, USA; Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs, USA; Brain Imaging Research Center, University of Connecticut, Storrs, USA; Department of Psychological Sciences, University of Connecticut, Storrs, USA

ABSTRACT
This exploratory study investigated relations between individual differences in cortical grey matter structure and young adult readers’ cognitive profiles. Whole-brain analyses revealed neuroanatomical correlations with word and nonword reading ability (decoding), and experience with printed matter. Decoding was positively correlated with grey matter volume (GMV) in left superior temporal sulcus, and thickness (GMT) in right superior temporal gyrus. Print exposure was negatively correlated with GMT in left inferior frontal gyrus (pars opercularis) and left fusiform gyrus (including the visual word form area). Both measures also correlated with supramarginal gyrus (SMG), but in spatially distinct subregions: decoding was positively associated with GMV in left anterior SMG, and print exposure was negatively associated with GMT in left posterior SMG. Our comprehensive approach to assessment both confirms and refines our understanding of the novel relation between the structure of pSMG and proficient reading, and unifies previous research relating cortical structure and reading skill.

ARTICLE HISTORY
Received 18 October 2017
Accepted 4 May 2018

KEYWORDS
Individual differences; reading and literacy; structural MRI; phonological decoding; print exposure

1. Introduction

Proficient reading depends upon the efficient coordination of both language-specific processes (e.g. phonological, lexical, syntactic, semantic) and domain general processes (e.g. working memory, reasoning). Becoming literate entails changes to the brain’s cortical structure, both in grey and white matter (Carreiras et al., 2009; Castro-Caldas et al., 1999; Petersson, Silva, Castro-Caldas, Ingvar, & Reis, 2007). There is substantial evidence for individual behavioural differences in reading comprehension and its components (for reviews see Long, Johns, & Morris, 2006; van den Broek, Mouw, & Kraal, 2015; Wagner, Piasta, & Torgesen, 2006), and that these differences often correlate with differences in functional activity in task-relevant brain regions (e.g. Clements-Stephens et al., 2012; Meyler, Keller, Cherkassky, Gabrieli, & Just, 2008; Shankweiler et al., 2008; Welcome & Joanisse, 2012). However, although it might be assumed that cortical structure may be similarly related to such behavioural differences, there is relatively little evidence available to support this hypothesis (for review see Richardson & Price, 2009). Structural imaging may provide critical complementary information about the neural substrates underlying reading behaviour. Consequently, the goal of the current study is to explore potential relations between cortical grey matter structure and performance measures related to reading comprehension.

There are few studies that have directly assessed whether literacy-related skills correlate with indices of grey matter structure. In a recent example (Jednoróg et al., 2015), two such skills were assessed: rapid automated naming (RAN) and decoding. The goal was to assess potential differences in grey matter volume (GMV) between two large groups of children, one with developmental dyslexia (n = 130), and the other an age-matched control group of unimpaired readers (n = 106). Performance on the RAN task – a speeded response task in which readers name characters that appear in a visual array – is often associated with fluent and efficient processing speed (although this is the subject of ongoing debate; for review see Norton & Wolf, 2012). Decoding
refers to the ability to accurately map orthographic information – i.e. printed words in alphabetic languages – to a corresponding phonological representation. Both RAN and decoding ability are strong predictors of reading comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990; Norton & Wolf, 2012; Shankweiler et al., 1999). Jednoróg and colleagues found no evidence for group differences in cortical grey matter. There was also no evidence that either decoding or RAN were related to grey matter differences in dyslexic readers. However, they did observe individual differences within the control group: in unimpaired readers, decoding and GMV were correlated such that higher word reading accuracy corresponded to greater GMV in left supramarginal gyrus (SMG). Given its power, this study provides robust evidence that a skill that is critical to reading comprehension can also be directly related to variation in the structure of cortical grey matter in non-dyslexic readers.

The approach adopted by Jednoróg et al. (2015) is in some respects emblematic of most investigations of grey matter morphology and reading ability. Such studies are often primarily concerned with describing some specific group of interest (e.g. dyslexic or bilingual readers). Consistent with this, most studies of linguistic relations to cortical grey matter emphasise group-level comparisons with “typical” (i.e. non-dyslexic, or monolingual) readers, rather than specific measures of the participants’ reading-related skills. Furthermore, even when such measures are included, their scope is usually quite limited: it is most common to obtain only a single skill measure (or perhaps two). One consequence of these methodological emphases is that our ability to assess potential relations between distinct components of reading skill and neuroanatomical structure is limited by the reliance on only narrow information about readers’ literacy skills. Another is that we know comparatively little about the possible neurostructural correlates of literacy in so-called “typical” readers. Thus, we conducted an exploratory study of cortical grey matter structure in monolingual young adults without diagnosed reading disabilities. Further, in order to more fully characterise both the processes supporting reading comprehension in our participants and their neuroanatomical correlates, we administered a battery of behavioural tests indexing a wide range of reading-related skills.

There is some evidence that the structural relation to decoding skill that was observed in children (Jednoróg et al., 2015) may also be present in young adults. In one study, brain morphology was related to group differences in college students (Welcome, Chiarello, Thompson, & Sowell, 2011). Group membership was determined by testing not only decoding efficiency, but also participants’ reading comprehension ability (i.e. their global understanding of a text). There were three groups: proficient readers (n = 22), whose scores indicated that their performance on both tests was commensurate with age; poor readers (n = 12), whose performance was below age-based norms; and “resilient” readers (n = 21), who had impaired decoding, but nonetheless exhibited age-appropriate reading comprehension. Decoding ability was related to hemispheric asymmetry in temporo-parietal regions (including SMG): grey matter thickness (GMT) in left hemisphere temporo-parietal regions (including SMG) was greater relative to their homologues in the right hemisphere in proficient readers. However, this asymmetry was reduced in both groups with poor decoding skills. Although reductions in the typical leftward asymmetry are known to be associated with specific reading disability (Eckert, 2004; Heim & Keil, 2004; see also Chiarello, Lombardino, Kacinik, Otto, & Leonard, 2006), this was the first report of such a relation in individuals without a history of reading disability. In contrast to decoding, reading comprehension ability was not related to the structure of the left temporo-parietal region. Instead, it correlated only with structural aspects of right hemisphere brain regions: specifically, radial expansion – a measure of “local brain shape” related to cortical surface area – was smaller in frontal and parietal regions for poor comprehenders relative to the other groups. Overall, these findings provide important neurostructural information about the components of literacy skill: they corroborate the importance of assessing decoding ability; indicate a discrete role for measures of reading comprehension ability; and suggest that the relation between efficient decoding and the structure of cortical regions including SMG may be consistent across developmentally distinct age groups.

Another study of college students without histories of reading difficulty (N = 28) also assessed both decoding ability and reading comprehension skill, and additionally included measures of experience with printed material (Goldman & Manis, 2013). Significant correlations among these three literacy-related skills are well established. Skilled readers typically exhibit greater decoding ability and more extensive print exposure; conversely, decoding difficulty is associated with less print exposure and poorer reading comprehension (Cunningham & Stanovich, 1991, 1997; Shankweiler et al., 1999; Share, 1995; Stanovich, 1986; Stanovich & Cunningham, 1992; for review see Mol & Bus, 2011). Goldman & Manis defined regions of interest in the left hemisphere reading network (e.g. Bolger, Perfetti, & Schneider, 2005; Pugh, 2006; Pugh et al., 2013). Decoding ability was not related to GMT in any of these regions in this sample. However, both print exposure and reading
comprehension were positively correlated with GMT. In separate analyses, print exposure was related to GMT in left SMG, left fusiform gyrus, both pars opercularis and pars triangularis in inferior frontal gyrus (IFG), and angular gyrus (AG); reading comprehension ability, by contrast, was related to GMT only in the latter two regions. A subsequent analysis explored whether the overlapping correlations in IFG and AG indicated that each literacy skill was a unique predictor of GMT, or whether the correlations in the initial analyses might instead be based upon variance shared between the two measures. Although the latter was the case in IFG, print exposure emerged as a unique predictor of GMT in AG. These results corroborate the importance of assessing reading comprehension ability; suggest experience with printed material as a potentially important structural correlate; and highlight the importance of accounting for shared variance among multiple measures of literacy skills in analyses of individual differences and brain morphology.

Two additional studies report significant correlations between cortical grey matter and literacy skills in non-clinical populations. The principal findings concern vocabulary knowledge, which is known to be strongly correlated with reading comprehension (Anderson, Wilson, & Fielding, 1988; Joshi, 2005; Perfetti, 2007; Stanovich, 1986). Measures of vocabulary knowledge often emerge as unique predictors that capture reading-related variance beyond that of other cognitive assessments (for reviews see Braze et al., 2016; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013; Tunmer & Chapman, 2012). Lee et al. (2007) found that greater vocabulary knowledge was positively related to grey matter density in bilateral SMG in adolescents (N = 34). Richardson, Thomas, Filippi, Harth, and Price (2010) confirmed this finding in adolescents, but did not observe the same correlation in either young children or adults. Moreover, in the later study, vocabulary knowledge was also positively associated with grey matter density in two other brain regions: left superior temporal sulcus (STS) and, in adults and adolescents only, in left posterior temporo-parietal cortex. Finally, both studies included some additional indices of individual differences, none of which correlated with cortical structure: Lee et al. included measures of verbal fluency and both verbal and performance IQ, and Richardson et al. included a measure of matrix reasoning ability. These results demonstrate the importance of assessing readers’ word knowledge. In addition, both studies used their measures of general reasoning as nuisance variables in their regression analyses, so that any observed effects could be attributed specifically to linguistic factors, rather than to general cognitive ability.

Taken together, these studies provide evidence that at least a small set of cognitive skills related to reading comprehension may be correlated with cortical grey matter structure. However, it is equally clear that direct comparison of their results is not straightforward. For example, the studies of collegiate young adults included no measure of vocabulary knowledge, obviating a possible extension of the vocabulary findings. Both studies of college students (Goldman & Manis, 2013; Welcome et al., 2011) assessed decoding ability, but comparing their results is complicated: although one tested whether decoding related to cortical structure, the other employed it to differentiate groups of participants (and analysed broadly defined cortical regions rather than specific areas of the brain). Both assessed reading comprehension, but Goldman and Manis did not examine the right hemisphere, precluding confirmation of differences in hemispheric asymmetry found by Welcome and colleagues. The populations tested differ in each study, with two using convenience samples of college students (age 18–24, Goldman & Manis, 2013; age 18–34, Welcome et al., 2011), one using pre-collegiate adolescents (age 12–16; Lee et al., 2007), one using children from three different nations (age 8–13; Jendróg et al., 2015), and one using a wide spectrum of ages (age 7–11, n = 9; age 12–17, n = 17; age 21–72, n = 22; Richardson et al., 2010). These sampling differences are non-trivial for comparing the results of these studies, as there is ample evidence for neurodevelopmental structural changes across these age groups (Giedd et al., 1999; Gogtay et al., 2004; Lu et al., 2009; Paus, 2005; Salat et al., 2004). The consequence of these methodological differences – assessing different indices of grey matter structure, and different measures of reading-related abilities, in different populations – is that each study presents a relatively narrow account of potential neurostructural links to the components of literacy skill. That is, although the available evidence does not contain clear contradictions, it also does not admit clear conclusions, either about the literacy skills that might be related to cortical structure, or the cortical structures to which they might relate.

Despite these inconsistencies, there is one region whose grey matter structure appears to be consistently linked to components of reading comprehension: supra-marginal gyrus. Furthermore, SMG’s structural relations to literacy skills may provide unique information about the neural substrates of reading comprehension. Specifically, both studies of vocabulary knowledge localised the correlation with grey matter density to posterior SMG (pSMG), rather than anterior SMG (aSMG). This distinction is important since, as noted by both Lee et al. (2007) and Richardson et al. (2010), pSMG has not been functionally
related to language-specific processes. Rather, functional relations to language are typically reported in neighboring areas such as aSMG and AG. These regions are associated with functional activity during phonological (aSMG) and semantic (AG) processing (Booth et al., 2002; Démonet et al., 1992; Devlin, Matthews, & Rushworth, 2003; Gathercole, Hitch, Service, & Martin, 1999; Price, Moore, Humphreys, & Wise, 1997; Tan, Laird, Li, & Fox, 2005). Additional support for this structural dissociation comes from a complementary analysis of white matter tractography, which revealed that pSMG has direct connections with both aSMG and AG, but that aSMG and AG are not themselves directly connected (Lee et al., 2007). Although Goldman and Manis (2013) did not specify MNI coordinates for their findings (making it unclear whether print exposure was correlated with a specific subregion of SMG), the other studies reporting links to SMG are broadly consistent with this pattern. The area of SMG in which GMT related to decoding in children was anteriorly located (Jednoróg et al., 2015); and although Welcome et al. (2011) also did not discriminate subregions of SMG, their figures suggest that the effects driven by group differences in decoding were in anterior SMG (see Figures 2 and 3 of Welcome et al., 2011, pp. 1199 & 1201). Importantly, this fine-grained subdivision of the structure of SMG according to specific components of literacy is not possible on the basis of functional associations alone.

This study advances our knowledge of possible links between brain morphology and skills related to literacy achievement in young adult readers. It is notable for at least three reasons. First, we assessed literacy skills using a large battery of cognitive assessments. This battery included measures assessed in previous studies, such as decoding ability, reading comprehension, vocabulary knowledge, and print exposure, as well as additional cognitive abilities that were not. This study simultaneously assessed the unique contributions of a broad range of specific cognitive measures, and is a necessary step for reconciling the diverse findings from previous research. Second, we assessed two indices of cortical structure: grey matter thickness and grey matter volume. Most previous research examining relations between cortical structure and reading ability (especially in clinical populations – see the General Discussion) has focused on only one of these, typically the latter, which is intuitive: GMV is derived from GMT and cortical surface area, and it might therefore be expected that differences observed in GMV would be reflected in its component measures. Yet recent evidence shows that this is only true for cortical surface area. In contrast, differences that manifest in GMT may not be reflected in GMV, and vice versa (Frye et al., 2010; Greve et al., 2013), making it important to assess both in order to clearly characterise grey matter variation (cf. Winkler et al., 2010). Finally, although most previous work has recruited students of one kind or another, we chose to recruit a community-based sample of young adults who were not university students. The neurobiological bases of literacy skill have not been as extensively studied in young adults as in early language learners (Curtis, 2002), and even less is known about young adult readers who are not enrolled in (and may not plan to obtain) post-secondary education. Based on our previous work with this population, we expected a broad range in literacy-related skills across participants (Braze, Tabor, Shankweiler, & Mencl, 2007; Braze, Mencl et al., 2011; Braze, Katz et al., 2016; Johns, Matsuki, & Van Dyke, 2015; Kukona et al., 2016; Kuperman & Van Dyke, 2011; Li et al., 2017; Magnuson et al., 2011; Shankweiler et al., 2008; Van Dyke, Johns, & Kukona, 2014), which confers an advantage in our power to detect individual differences (for discussion see Peterson, 2001).

Given the small number of studies that constitute the current state of the field, we consider this an exploratory investigation. The scarcity of previous research, as well as its methodological heterogeneity and diverse, non-overlapping patterns of results, make it difficult to propose specific hypotheses about potential links between cortical structure and behavioural measures of literacy-related skills. Thus, although we were particularly interested in SMG (see above), we did not define any hypothesis-driven regions of interest a priori. Rather, we conducted a naïve whole brain analysis without assumptions or restrictions based on the size, location, or direction of potential neurostructural correlations with the behavioural battery measures.

2. Method

2.1. Participants

We obtained informed consent from 39 young people recruited from the local community. We recruited participants in several ways, including presentations at adult education centres, advertisements in local newspapers, and flyers placed on adult school campuses, community centres, public transportation hubs, and local retail and laundry facilities. Of the 39 participants, four were left-handed, and their data were excluded from further analysis. The remaining 35 participants (ages 16–24 years, mean 20.44; 17 female) were right-handed native English speakers with normal or corrected-to-normal vision. Participants reported no history of psychiatric or neurological disorder, no active use of psychoactive
medications, and no diagnosed reading or learning disability. Based on the Fast Reading subtest of the Stanford Diagnostic Reading Test (Karlson & Gardner, 1995), all participants demonstrated the ability to read well enough to comprehend basic texts (minimum 70% accuracy on attempted items). Each participant underwent two experimental sessions, each on a separate day. Participants received $80 for one scanning session, which lasted no longer than 60 min, together with one behavioural testing session lasting no longer than three hours. Behavioural testing was completed prior to the MRI scan. The Yale University Human Investigation Committee approved this protocol.

2.2. Literacy-related cognitive assessments

We administered a battery of behavioural tests of literacy-related skills and abilities. Standardised instruments were chosen to optimise construct validity and test-retest reliability. The standardised measures are widely used for clinical assessment and diagnosis, and were administered individually during individual test sessions. Two skills – working memory and print exposure – were not derived from standardised assessments; in these cases, we employed test instruments identical in format to those that are commonly used in experimental research. The skills we examined, and the tests associated with them, included:

- Print exposure: Magazine Recognition Test (Acheson, Wells, & MacDonald, 2008; Cunningham & Stanovich, 1990), in which participants identify real magazine titles from a list that includes real and foil titles. We retained the original format of this test, but updated test items by replacing out-of-print titles with the names of current publications.
- Vocabulary knowledge: Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1997), a test of receptive vocabulary knowledge, in which participants hear a target word and select a picture (from a group of four possibilities) that best depicts its definition.
- Working memory capacity (WMC): assessed with the Sentence Span task (Daneman & Carpenter, 1980). Following the original format of this complex span assessment, participants heard sets of 2–6 sentences (number per set increases linearly), judging each as true or false; after each series, all sentence-final words must be recalled (in any order). We used an auditory variant of the task, permitting us to measure verbal working memory independent of the need to decode printed stimuli; moreover, we modified the sentence materials in order to make them more amenable for use with our community-based sample (for details, see Clark, McRoberts, Van Dyke, Shankweiler, & Braze, 2012).
- Non-linguistic reasoning ability: assessed using the Weschler Abbreviated Scale of Intelligence (WASI; Psychological Corp., 1999). We used the Matrix Reasoning subtest, in which participants completed visual analogical reasoning tasks, as a measure of general cognitive ability.
- Phonological awareness: Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). We used the composite phonological awareness measure, derived from the Elision core subtest (forming words by eliding a phonological segment from spoken word prompts) and the Blending Words core subtest (spoken sounds are combined to form words).
- Rapid naming: CTOPP (Wagner et al., 1999), Rapid Letter Naming core subtest. Scores on this test reflect time to name letters presented in a grid-like array. Because RAN scores are naming times, lower scores indicate better performance.
- Reading comprehension: Peabody Individual Achievement Test-Revised (PIAT-R; Markwardt, 1998). Participants read a series of sentences of increasing difficulty, choosing a picture (from a group of four possibilities) corresponding to the meaning of each. We administered odd numbered items to measure reading comprehension, reserving the even numbered items for a measure of speech sentence comprehension (Leach, Scarborough, & Rescorla, 2003; Spring & French, 1990).
- Listening comprehension: we created a listening comprehension assessment by splitting the PIAT-R (Markwardt, 1998), such that even numbered items were recorded and presented aurally in order to assess listening comprehension (Leach et al., 2003; Spring & French, 1990). The characteristics of the sentences, and the behavioural response task (i.e. selecting a picture from an array) are therefore identical to our reading comprehension measure.
- Reading Fluency: indexed using the WJ-III silent reading fluency subtest, from the reading and oral comprehension area subtests (WJ-III; Woodcock, McGrew, & Mather, 2001). This test measures the speed of reading sentences silently and answering yes/no questions about each. We also measured oral reading fluency through a subset of the Gray Oral Reading Test, fourth edition (GORT, passages 5, 7, and 9; Wiederholt & Bryant, 2001). Reading time for each passage was converted to a rate using the published tables; these were summed to yield a single score.
Decoding skill: assessed using both the Woodcock–Johnson-III Tests of Achievement, reading and oral comprehension area subtests (WJ-III; Woodcock et al., 2001), Word Attack (reading a list of pseudowords aloud) and Letter-Word Identification (naming words from a list); and the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), sight word efficiency subtest (indexes the number of words that can be named in 45 s) and phonemic decoding efficiency subtests (indexes the number of pronounceable nonwords that can be named in 45 s).

2.3. Data preparation

We inspected the distribution of the raw scores for each measure using density and quantile-quantile plots in order to assess univariate normality, and using the Shapiro–Wilk test, which assesses both skewness and kurtosis (Shapiro & Wilk, 1965; implemented in the stats R package, R Core Team, 2016). Some measures showed significantly skewed distributions, which can inflate the influence of non-normal data. In order to correct this, we applied the Box–Cox transformation where appropriate (Box & Cox, 1964). The transformation equation is \( y(\lambda) = \frac{y^{\lambda} - 1}{\lambda} \) given \( \lambda \neq 0 \), if \( \lambda = 0 \), then \( \log(y) \). Using the caret R package (Kuhn, 2016), we calculated optimal lambda values for each measure. Lambda values close to 1.00 indicate that no transformation is necessary, since the distribution of the transformed data will be identical to the original data.

After addressing the distributional characteristics of the measures, all predictor variables were standardised: each was first mean-centred (i.e. the mean of each measure is subtracted from its value, setting its mean to zero, but leaving the standard deviation unchanged) and then scaled (i.e. the resulting values were converted to z-scores). Standardisation has several well-established analytic benefits (as described in, e.g. McElreath, 2016), such as allowing straightforward comparison of the relative influence of predictor variables, (which might not be possible with unstandardised data due to, for example, differences in measurement and/or scale), and reducing potential problems related to multicollinearity among predictor variables (see below).

Some measures in our test battery target the same theoretical constructs. For these, we built composite variables to improve reliability and to more robustly represent the underlying constructs. We created two composite variables: Reading Fluency (comprised of the WJ-III Reading Fluency subtest and the GORT, \( r = .372, p = .028 \)) and Word Decoding (comprised of the WJ-III Word Attack and Letter-Word Identification subtests, and the TOWRE subtests, all \( r > .45, \) all \( p < .01 \)). Composites were derived by averaging component measures after they were first standardised, and then rescaling the resulting composite values. This approach is common in clinical and psycholinguistic studies of reading and reading-related skills (e.g. Braze et al., 2007; Guo, Roehrig, & Williams, 2011; Hua & Keenan, 2014; Kukona et al., 2016; Pugh et al., 2008; Sabatini, Sawaki, Shore, & Scarborough, 2010; Shankweiler et al., 2008; Van Dyke et al., 2014).

Finally, it is well known that performance on individual difference measures tends to be correlated, sometimes highly, making it difficult to uniquely relate specific constructs to dependent variables (for discussion see Freed, Hamilton, & Long, 2017). However, neither the number nor the magnitude of bivariate correlations is an unambiguous indicator of troublesome multicollinearity: strongly correlated measures may not induce problematic multicollinearity, while high multicollinearity can occur even when all bivariate correlations in a set of variables are quite low (Belsley, 1991a; Flom, 1999). Thus, although standardising the predictor variables is known to reduce such dependencies (for discussion see McElreath, 2016), we nonetheless assessed the potential for problematic multicollinearity among our individual differences assessments. Using the perturb R package (Hendrickx, 2012), we calculated each predictor’s condition number (\( \kappa \)), which provides information about how much the variance associated with an estimated regression coefficient is increased because of overlap with other predictors (Belsley, Kuh, & Welsch, 1980). Condition numbers are considered to be both more informative and precise than other estimates of multicollinearity, such as the more commonly used variance inflation factor (Belsley, 1991a). This is because condition numbers, unlike variance inflation factors, provide not only estimates of shared variance but information about the ensembles of variables which may be sharing variance. By contrast, variance inflation factors are relatively uninformative, because they cannot account for connections among variables (Harrell, 2001). For condition numbers that are “absolutely small, for example, 5 or 10 … collinearity is not really a major problem” (Belsley, 1991b, p. 42), whereas \( \kappa \geq 30 \) suggests problematic multicollinearity (Belsley et al., 1980; Belsley, 1991a, 1991b; Faraway, 2014). Similarly, Baayen notes that for \( \kappa \) values “between 0 and 6, there is no collinearity to speak of. Medium collinearity is indicated by condition numbers around 15, and conditions numbers of 30 or more indicate potentially harmful collinearity” (Baayen, 2008, p. 182).
2.4. Structural imaging

2.4.1. Image acquisition and processing

We collected volumetric data from high-resolution 3D MPRAGE anatomical images, acquired on a Siemens 1.5 T Sonata MR system (192 sagittal slices; TE = 4.66 ms, TR = 2530 ms; FOV = 256 × 256 voxel matrix; resolution = 1.33 × 1.33 × 1.30 mm³). One whole-head, high resolution T1-weighted MPRAGE anatomical volume was acquired per participant. We used the FreeSurfer image analysis suite to perform cortical reconstruction and volumetric segmentation (Dale, Fischl, & Sereno, 1999; Dale & Sereno, 1993; Fischl & Dale, 2000; Fischl et al., 2001: http://surfer.nmr.mgh.harvard.edu, version 5.3). Prior to segmentation and classification, all images were visually inspected to identify potentially problematic motion artifacts. Subsequently, each segmentation was visually inspected to ensure (1) accurate skull stripping, (2) correct classification of grey/white matter boundaries, (3) appropriate separation of brain/non-brain matter. Minor adjustments to the automated segmentation and parcellation routines were made when necessary (e.g. adding control points to facilitate grey/white matter classification), but no major alterations were necessary. Non-brain tissue was removed using a hybrid watershed/surface deformation procedure (Ségonne et al., 2004). The resulting skull-stripped brain was processed using an automated Talairach transformation, segmentation of the subcortical white matter and deep grey matter volumetric structures (Fischl et al., 2002, 2004), intensity normalisation (Sled, Zijdenbos, & Evans, 1998), tessellation of the grey/white matter boundary, automated topology correction (Fischl et al., 2001; Ségonne, Pacheco, & Fischl, 2007), and surface deformation following intensity gradients to optimally place the grey/cerebrospinal and grey/white fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class (Dale et al., 1999; Dale & Sereno, 1993; Fischl & Dale, 2000). Both intensity and continuity information from the entire three-dimensional MR volume are used to produce representations of cortical thickness, calculated as the closest distance from the grey/white boundary to the grey/CSF boundary at each vertex on the tessellated surface (Fischl & Dale, 2000). The maps are not restricted to the voxel resolution of the original data, and are created using spatial intensity gradients across tissue classes (i.e. they are not simply reliant on absolute signal intensity). These procedures for the measurement of cortical thickness have been validated against histological analysis (Rosas et al., 2002) and manual measurements (Kuperberg et al., 2003; Salat et al., 2004).

2.4.2. Spatial analysis

Differences in the measurements of grey matter volume and thickness were examined for both the left and right cerebral hemispheres with a vertex by vertex general linear model (GLM). Statistical analysis was performed at each vertex to test the significance of the correlation between the individual difference measurements and these structural measurements. The model included 11 regressors: 10 corresponding to the behavioural individual difference assessments, and an additional covariate of no interest – participant age (mean centred) – which was modelled to reduce error variance. Total intracranial volume (TIV) is often included as a covariate for between-group designs and regions-of-interest analyses, with the goal of normalising the data so that group-level effects are not confounded with individual differences in GMV. However, since we are specifically concerned with modelling correlations with individual differences in GMV, including TIV would explicitly remove variance relevant to this research question, making it an inappropriate covariate for our within-subjects design. In addition, there is broad agreement that GMT should not be normalised in any case (e.g. Feng et al., 2009; Westman, Aguilar, Muehlboeck, & Simmons, 2013), and previous individual difference studies of GMT have thus not done so (e.g. Goldman & Manis, 2013; He et al., 2013).

Separate GLMs were constructed for GMV and GMT to estimate parameters for the covariates. These parameter estimates were then submitted to a group-level analysis and converted to $p$-value maps. Given the exploratory nature of this study, all vertex-wise results were thresholded at an individual vertex level of $p < 0.05$, and cluster extent thresholds corrected for multiple comparisons ($p < 0.05$) were calculated through Monte Carlo simulations of white noise on the cortical surface (Hagler, Saygin, & Sereno, 2006). These analyses permit the evaluation of the unique contributions of the regressors, in that they assess the proportion of variance in the dependent
variables (GMT and GMV) that is associated with one predictor but not any of the other predictors (i.e. their squared semi-partial correlations; see, e.g. Keith, 2014; Tabachnick & Fidell, 2012). Finally, it is worth noting that the appropriate parameters for the analysis of neuroimaging data is a topic of active, ongoing discussion. For example, although Eklund, Nichols, and Knutsson (2016) advocate more stringent cluster-forming thresholds for functional data, their results were not subsequently replicated (Cox, Chen, Glen, Reynolds, & Taylor, 2017a, 2017b), and may not account for elevated Type II error (Lohmann et al., 2017). At the time of this writing, the implications of this debate for the analysis of structural MRI data have not been tested. Therefore, our cluster-forming threshold was selected to be consistent with both the exploratory nature of our study and the current “state-of-the-science” threshold used in other recent studies examining structural imaging data (e.g. Bizzo et al., 2017; Gardumi, Ivanov, Havlicek, Formisano, & Uludağ, 2017; Jasińska et al., 2017).

### Table 1. Descriptive statistics for all battery measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Range</th>
<th>M</th>
<th>SD</th>
<th>Max. possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Magazine recognition test</td>
<td>0–23</td>
<td>10.31</td>
<td>5.66</td>
<td>40</td>
</tr>
<tr>
<td>2. Vocabulary knowledge</td>
<td>132–192</td>
<td>169.80</td>
<td>16.84</td>
<td>204</td>
</tr>
<tr>
<td>Age equivalent</td>
<td>10.4–23.1</td>
<td>18.93</td>
<td>4.51</td>
<td>&gt;23.1</td>
</tr>
<tr>
<td>3. Working memory capacity</td>
<td>26–60</td>
<td>46.37</td>
<td>7.26</td>
<td>60</td>
</tr>
<tr>
<td>5. Phonological awareness</td>
<td>61–118</td>
<td>93.23</td>
<td>15.66</td>
<td>150</td>
</tr>
<tr>
<td>Percentile equivalent</td>
<td>1–89</td>
<td>38.86</td>
<td>29.42</td>
<td>99</td>
</tr>
<tr>
<td>6. Rapid reading naming (in seconds)</td>
<td>16–34</td>
<td>23.91</td>
<td>4.18</td>
<td>n/a</td>
</tr>
<tr>
<td>Percentile equivalent</td>
<td>5–99</td>
<td>52.34</td>
<td>28.45</td>
<td>99</td>
</tr>
<tr>
<td>7. Listening comprehension</td>
<td>14–40</td>
<td>33.97</td>
<td>4.79</td>
<td>41</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>3–13</td>
<td>9.80</td>
<td>2.31</td>
<td>13</td>
</tr>
<tr>
<td>8. Reading comprehension</td>
<td>20–40</td>
<td>32.74</td>
<td>5.60</td>
<td>41</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>3.8–13</td>
<td>9.26</td>
<td>2.9</td>
<td>13</td>
</tr>
<tr>
<td>10. Silent reading fluency</td>
<td>51–98</td>
<td>74.69</td>
<td>15.04</td>
<td>98</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>5.8–19</td>
<td>12.65</td>
<td>6.60</td>
<td>&gt;18</td>
</tr>
<tr>
<td>Rate (items/min., max. 180 s)</td>
<td>17–43</td>
<td>25.68</td>
<td>6.68</td>
<td>n/a</td>
</tr>
<tr>
<td>11. Word identification</td>
<td>59–76</td>
<td>68.49</td>
<td>4.37</td>
<td>76</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>6.7–19</td>
<td>13.65</td>
<td>4.16</td>
<td>&gt;18</td>
</tr>
<tr>
<td>12. Letter-word identification</td>
<td>20–31</td>
<td>27.23</td>
<td>2.91</td>
<td>32</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>4.3–19</td>
<td>10.67</td>
<td>4.26</td>
<td>&gt;18</td>
</tr>
<tr>
<td>13. Sight word efficiency</td>
<td>75–104</td>
<td>91.71</td>
<td>8.92</td>
<td>104</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>5.8–12.6</td>
<td>10.68</td>
<td>2.12</td>
<td>&gt;12.6</td>
</tr>
<tr>
<td>Rate (items/min., max. 45 s)</td>
<td>100–156</td>
<td>123.90</td>
<td>14.66</td>
<td>n/a</td>
</tr>
<tr>
<td>14. Phonemic decoding efficiency</td>
<td>22–62</td>
<td>52.23</td>
<td>8.73</td>
<td>63</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>2.8–12.6</td>
<td>10.63</td>
<td>2.76</td>
<td>&gt;12.6</td>
</tr>
<tr>
<td>Rate (items/min., max. 45 s)</td>
<td>29–116</td>
<td>72.34</td>
<td>15.50</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Note: 1: Print Exposure (adapted from Cunningham & Stanovich, 1990); 2: Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1997); 3: Listening span (Daneman & Carpenter, 1980); 4: Weschler Abbreviated Scales of Intelligence (Psychological Corp., 1999); 5–6: Comprehensive Test of Phonological Awareness (Wagner et al., 1999); 7–8: Peabody Individual Achievement Test-Revised (Markwardt, 1998); 9: Gray Oral Reading Test (Wiederhold & Bryant, 2001); 10: Woodcock-Johnson-III Tests of Achievement (Woodcock et al., 2001) silent reading fluency; 11–12: Woodcock-Johnson-III Tests of Achievement (Woodcock et al., 2001); 13: Test of Word Reading Efficiency (Torgesen et al., 1999).

### 3. Results

#### 3.1. Literacy-related cognitive assessments

All battery measures were analysed with the R statistical software, version 3.3.1 (R Core Team, 2016). Descriptive information, including range, mean, standard deviation, are shown in Table 1. To aid interpretability and ease comparison with other studies, we include grade or age equivalents where possible. CTOPP scores do not have age or grade equivalents, but do include age-leveled percentile ranks. All analyses are based on raw scores except for the two TOWRE subtests and the WJ-III reading fluency subtest: because some participants completed these timed tests in less than the maximum time allotted, we converted scores on these measures into rates that index items-per-minute.

We next examined the distribution of the raw scores for each measure for normality and potential outliers. Density and quantile-quantile plots suggested that some battery data were non-normal. The results of the Shapiro-Wilk normality test indicate that the data from seven battery measures deviated from normality. After applying the Box–Cox transformation to these measures (as described above), all distributions but one no longer deviated from normality. The remaining measure, WJ-III (silent) reading fluency, had a lambda of 1.04, indicating no advantage to transformation. The details of the tests of normality and of the Box–Cox data transformations appear in Table 2.

Correlations among the 10 battery regressors are shown in Table 3. We observed a range of correlation strength, which we characterise according to the finer gradation proposed in Evans (1996). The correlations appear in Table 2.

### Table 2. Results of the Shapiro-Wilk (W) test of normality before and after data transformation (and associated lambda (λ) values, where appropriate) for all battery and composite measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>W</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Magazine recognition test</td>
<td>0.972</td>
<td></td>
</tr>
<tr>
<td>2. Vocabulary knowledge</td>
<td>0.930*</td>
<td>4.36</td>
</tr>
<tr>
<td>3. Working memory capacity</td>
<td>0.978</td>
<td></td>
</tr>
<tr>
<td>4. Matrix Reasoning</td>
<td>0.869***</td>
<td>2.93</td>
</tr>
<tr>
<td>5. Phonological awareness</td>
<td>0.965</td>
<td></td>
</tr>
<tr>
<td>6. Rapid letter naming</td>
<td>0.978</td>
<td></td>
</tr>
<tr>
<td>7. Listening comprehension</td>
<td>0.767****</td>
<td>4.81</td>
</tr>
<tr>
<td>8. Reading comprehension</td>
<td>0.890**</td>
<td>3.49</td>
</tr>
<tr>
<td>9. Oral reading fluency</td>
<td>0.902**</td>
<td>2.94</td>
</tr>
<tr>
<td>10. Silent reading fluency</td>
<td>0.900**</td>
<td>1.04</td>
</tr>
<tr>
<td>11. Word identification</td>
<td>0.976</td>
<td></td>
</tr>
<tr>
<td>12. Letter-word identification</td>
<td>0.916*</td>
<td>4.80</td>
</tr>
<tr>
<td>13. Sight word efficiency</td>
<td>0.965</td>
<td></td>
</tr>
<tr>
<td>14. Phonemic decoding efficiency</td>
<td>0.967</td>
<td></td>
</tr>
<tr>
<td>Composite measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9–10. Reading Fluency Composite</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td>11–14. Decoding Composite</td>
<td>0.975</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01; ***p < .001; ****p < .0001; *Transform unnecessary.
Table 3. Condition numbers (κ) for and correlations (r) among the individual differences regressors.

| Measure                           | κ  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|----------------------------------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1. Print Exposure                | 2.29 | 2.29 | 3.14 | 2.72 | 0.047 | 0.574 | 0.066 | 0.099 | 0.009 | 0.015 |
| 2. Vocabulary                    | 4.64 | 0.001 | 0.002 | 0.0057 | 0.000 | 0.084 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3. Working Memory                | 3.98 | 0.175 | 0.506 | 0.089 | 0.006 | 0.933 | 0.002 | 0.022 | 0.046 | 0.006 |
| 4. Reasoning                     | 3.16 | 0.061 | 0.324 | 0.291 | 0.002 | 0.049 | 0.004 | 0.000 | 0.257 | 0.117 |
| 5. Phonological Awareness        | 2.85 | 0.339 | 0.640 | 0.456 | 0.508 | 0.178 | 0.000 | 0.000 | 0.012 | 0.000 |
| 6. Rapid Naming                  | 3.68 | -0.098 | -0.026 | -0.135 | 0.323 | 0.481 | 0.123 | 0.173 | 0.028 | - |
| 7. Reading Comprehension         | 6.87 | 0.315 | 0.813 | 0.506 | 0.470 | 0.630 | 0.123 | 0.000 | 0.006 | 0.003 |
| 8. Listening Comprehension       | 5.04 | 0.283 | 0.747 | 0.385 | 0.569 | 0.605 | 0.266 | 0.778 | 0.000 | 0.000 |
| 9. Reading Fluency Composite     | 8.26 | 0.435 | 0.581 | 0.339 | 0.197 | 0.420 | 0.236 | 0.560 | 0.454 | 0.000 |
| 10. Decoding Composite           | 10.18 | 0.408 | 0.684 | 0.452 | 0.270 | 0.624 | -0.372 | -0.669 | 0.483 | 0.654 |

<table>
<thead>
<tr>
<th>Measure</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Fluency Composite</td>
<td>0.552</td>
</tr>
<tr>
<td>Print Exposure</td>
<td>0.684</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.552</td>
</tr>
<tr>
<td>Working Memory</td>
<td>0.408</td>
</tr>
<tr>
<td>Reasoning</td>
<td>-0.408</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>0.684</td>
</tr>
<tr>
<td>Rapid Naming</td>
<td>0.552</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>0.408</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>0.684</td>
</tr>
<tr>
<td>Decoding Composite</td>
<td>0.552</td>
</tr>
</tbody>
</table>

Note: Correlations appear below the diagonal; their associated p-values appear above the diagonal.

r ≥ .335, p < .05, |r| ≥ .435, p < .01, |r| ≥ .552, p < .001, |r| ≥ .624, p < .0001.

Table 4. Grey matter thickness: individual difference measures, peak t-values, and centroid coordinates.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cluster</th>
<th>Area (mm²)</th>
<th>t</th>
<th>p</th>
<th>MNI Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Exposure</td>
<td>LH ITG</td>
<td>1617</td>
<td>-4.01</td>
<td>.0001</td>
<td>-40.8</td>
</tr>
<tr>
<td></td>
<td>LH IFG (pars opercularis)</td>
<td>1202</td>
<td>-5.75</td>
<td>.0037</td>
<td>54.3</td>
</tr>
<tr>
<td></td>
<td>LH SMG</td>
<td>1150</td>
<td>-2.53</td>
<td>.0051</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>LH caudal middle frontal</td>
<td>901</td>
<td>-2.20</td>
<td>.0251</td>
<td>38.3</td>
</tr>
<tr>
<td></td>
<td>RH rostral middle frontal</td>
<td>3280</td>
<td>-3.83</td>
<td>.0001</td>
<td>42.3</td>
</tr>
<tr>
<td></td>
<td>RH MTG</td>
<td>1229</td>
<td>-3.66</td>
<td>.0040</td>
<td>47.8</td>
</tr>
<tr>
<td>Workng Memory</td>
<td>RH precentral</td>
<td>1581</td>
<td>-3.15</td>
<td>.0004</td>
<td>34.8</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>RH MTG</td>
<td>1129</td>
<td>-3.07</td>
<td>.0069</td>
<td>60.8</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td>RH lingual</td>
<td>1420</td>
<td>3.58</td>
<td>.0008</td>
<td>32.8</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>LH rostral middle frontal</td>
<td>2325</td>
<td>-2.92</td>
<td>.0001</td>
<td>-19.3</td>
</tr>
<tr>
<td></td>
<td>LH ITG</td>
<td>1401</td>
<td>-3.39</td>
<td>.0007</td>
<td>40.8</td>
</tr>
<tr>
<td></td>
<td>LH superior frontal</td>
<td>860</td>
<td>-3.27</td>
<td>.0133</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>RH transverse temporal</td>
<td>3093</td>
<td>-5.97</td>
<td>.0001</td>
<td>39.2</td>
</tr>
<tr>
<td></td>
<td>RH rostral middle frontal</td>
<td>1380</td>
<td>-4.52</td>
<td>.0010</td>
<td>41.9</td>
</tr>
<tr>
<td></td>
<td>RH inferior parietal</td>
<td>849</td>
<td>-3.54</td>
<td>.0470</td>
<td>52.0</td>
</tr>
<tr>
<td></td>
<td>RH caudal middle frontal</td>
<td>841</td>
<td>-2.79</td>
<td>.0492</td>
<td>39.3</td>
</tr>
<tr>
<td>Decoding Composite</td>
<td>RH STG</td>
<td>1720</td>
<td>4.23</td>
<td>.0001</td>
<td>43.8</td>
</tr>
<tr>
<td></td>
<td>RH precentral</td>
<td>1357</td>
<td>3.76</td>
<td>.0012</td>
<td>23.1</td>
</tr>
<tr>
<td></td>
<td>RH lateral occipital</td>
<td>1191</td>
<td>4.60</td>
<td>.0047</td>
<td>29.4</td>
</tr>
</tbody>
</table>

4. Discussion

Our discussion focuses on the subset of the overall findings that converge with previous studies. This encompasses SMG, left IFG, left STS, and areas in right frontal and parietal areas. Given SMG’s prominence in previous studies, we highlight significant correlations between...
the grey matter structure of this region and literacy skills in Figure 3.

4.1. Supramarginal gyrus

Our analysis confirms and extends several previous findings in this region, as well as producing significant novel results. In our study, decoding ability was positively correlated with GMV in SMG. This confirms a previous finding in children (Jednoróg et al., 2015), and is also in line with our interpretation of the group differences observed between good and both poor and resilient readers (Welcome et al., 2011). In addition, we observed a significant negative correlation between print exposure

Table 5. Grey matter volume: individual difference measures, peak t-values, and centroid coordinates.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cluster</th>
<th>Area (mm$^2$)</th>
<th>t</th>
<th>p</th>
<th>MNI Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Exposure</td>
<td>RH superior frontal</td>
<td>1133</td>
<td>−8.4</td>
<td>.0130</td>
<td>−22.0 1.3 56.1</td>
</tr>
<tr>
<td>Working Memory</td>
<td>LH superior frontal</td>
<td>905</td>
<td>−3.03</td>
<td>.0432</td>
<td>−17.2 16.1 57.7</td>
</tr>
<tr>
<td>Reasoning</td>
<td>LH postcentral</td>
<td>895</td>
<td>−2.05</td>
<td>.0470</td>
<td>−53.3 −20.7 33.1</td>
</tr>
<tr>
<td>Reading Fluency Composite</td>
<td>LH lateral occipital</td>
<td>1261</td>
<td>−2.29</td>
<td>.0048</td>
<td>−22.8 −94.0 7.0</td>
</tr>
<tr>
<td>Decoding Composite</td>
<td>LH lateral occipital</td>
<td>1308</td>
<td>2.54</td>
<td>.0036</td>
<td>−21.5 −94.7 5.3</td>
</tr>
<tr>
<td></td>
<td>LH SMG</td>
<td>1257</td>
<td>3.03</td>
<td>.0048</td>
<td>−52.3 −26.4 20.2</td>
</tr>
<tr>
<td></td>
<td>LH lateral orbitofrontal</td>
<td>1244</td>
<td>2.89</td>
<td>.0058</td>
<td>−27.2 25.6 0.2</td>
</tr>
<tr>
<td></td>
<td>LH medial orbitofrontal</td>
<td>1223</td>
<td>3.27</td>
<td>.0061</td>
<td>−7.7 41.7 −16.6</td>
</tr>
<tr>
<td></td>
<td>LH superior frontal</td>
<td>1103</td>
<td>2.75</td>
<td>.0118</td>
<td>−8.7 11.5 54.4</td>
</tr>
<tr>
<td></td>
<td>LH STS</td>
<td>920</td>
<td>3.23</td>
<td>.0394</td>
<td>−50.2 −45.4 8.2</td>
</tr>
<tr>
<td></td>
<td>RH postcentral</td>
<td>3448</td>
<td>3.90</td>
<td>.0001</td>
<td>37.7 −7.1 17.7</td>
</tr>
<tr>
<td></td>
<td>RH superior frontal</td>
<td>2889</td>
<td>5.05</td>
<td>.0001</td>
<td>7.9 50.5 17.6</td>
</tr>
<tr>
<td></td>
<td>RH MTG</td>
<td>1148</td>
<td>3.01</td>
<td>.0121</td>
<td>57.6 −53.4 7.1</td>
</tr>
<tr>
<td></td>
<td>RH lingual</td>
<td>1030</td>
<td>3.46</td>
<td>.0246</td>
<td>22.4 −53.6 6.6</td>
</tr>
<tr>
<td></td>
<td>RH superior parietal</td>
<td>1002</td>
<td>2.81</td>
<td>.0291</td>
<td>22.0 −61.3 37.8</td>
</tr>
</tbody>
</table>
and GMT in left SMG, extending posteriorly to AG, indicating that greater experience with printed material was associated with thinner grey matter in our sample. This confirms the previous report of a relation between print exposure and GMT in SMG, although the correlation in that study was positive (Goldman & Manis, 2013). This discrepancy may stem from methodological differences, e.g. our use of a broad, community-based sample of participants, rather than a relatively skilled subgroup (i.e. college students) of the population. Furthermore, our finding is also consistent with a hypothesis initially proposed by Goldman and Manis: that a negative correlation between cortical thickness and print exposure is compatible with longitudinal evidence that cortical thinning is a byproduct of maturation, possibly related to skill consolidation (as proposed in, e.g. Lu et al., 2007; see also Sowell, Thompson, Tessner, & Toga, 2001; Sowell et al., 2004). Finally, we found no relation between vocabulary

Figure 2. Cluster-corrected results for individual difference measures correlated with GMV projected onto the fsaverage template in FreeSurfer. All cluster-corrected results depicted at $p < 0.05$. (A) Results projected onto the left hemisphere; (B) results projected onto the right hemisphere. From top to bottom: (1) lateral view, (2) medial view.

Note: purple indicates overlap between matrix reasoning ability and the decoding composite; yellow indicates overlap between the reading fluency and decoding composite measures.

Figure 3. Left supramarginal gyrus: the bounded area is the Desikan-Killiany parcellation of SMG in MNI space. Positive correlations with GMV are depicted on a red-yellow scale, and negative correlations with GMT are depicted on a blue-white scale. Results depicted at an uncorrected vertex-wise threshold of $p < 0.05$. 
knowledge and the structure of SMG in our young adult readers. This is consistent with previous findings indicating that such a relation was only present in adolescents between 12 and 17 years of age (Lee et al., 2007; Richardson et al., 2010), a range with which our sample barely overlaps.

Our results also dovetail with previous studies indicating that SMG can be bisected into distinct anterior and posterior regions. Here, the correlation with decoding efficiency was centred in anterior SMG, and the centroid of the correlation with print exposure was in the posterior region of SMG. Although ours is the first study to directly relate decoding to the structure of aSMG in monolingual young adult readers, this result is analogous to previous findings in children (Jednoróg et al., 2015) and bilingual young adults (L2 English; He et al., 2013). In the latter case, decoding was positively correlated with GMV in aSMG in a large sample (N = 253) of native Chinese speakers. Our study is also the first to directly relate readers’ experience with printed material to structural aspects of pSMG. Furthermore, print exposure is known to be correlated with vocabulary knowledge (Acheson et al., 2008; Long et al., 2008; Stanovich, 1993; Stanovich & West, 1989; Stanovich, West, & Harrison, 1995), as indeed it is in our data (r = .552, p < .001). Reading facilitates acquisition of novel word forms and the development of skilled spelling (Mol & Bus, 2011). Indeed, readers with greater print experience are more likely to encounter rare words in print than in spoken language (e.g. Carroll, Davies, & Richman, 1971), and potentially see ten times as many words as readers with more limited exposure (Nagy & Anderson, 1984). It is therefore no surprise that experience with printed matter is a significant vehicle by which new words are acquired (Cunningham, Stanovich, & West, 1994; Stanovich, 1986). Thus, the relation between grey matter in pSMG and print exposure that we report here is clearly not inconsistent with previously reported structural correlations between pSMG and vocabulary knowledge (Lee et al., 2007; Richardson et al., 2010).

4.2. Left inferior frontal gyrus

Our whole-brain analysis also found a relation between literacy skills and pars opercularis in left IFG. Pars opercularis – as part of Broca’s Area – has long been associated with many language-related processes (for review, see Friederici, 2011). As in SMG, print exposure was negatively correlated with GMT in this area, indicating that greater experience with printed matter was associated with thinner cortex (see Figure 1). This confirms the previously reported link between print exposure and left IFG but, again, the direction of the correlation in our study differs from the first report (Goldman & Manis, 2013). As previously noted, we suspect that the results diverge in this way due to methodological and analytic differences between the two studies.

Our findings align with those of Goldman and Manis (2013) in one regard: neither study obtained evidence for a correlation between decoding ability and GMT in IFG. Such a relation might have been expected because of two prior findings. First, longitudinal research suggests that the cortex thickens in IFG (and bilateral STG) over time in young children (Sowell et al., 2004). Second, thickening in pars opercularis during normal development in childhood has been related to increasingly proficient phonological awareness (Lu et al., 2007). Phonemic awareness is a necessary precursor to skilled decoding ability, and corresponds to the capacity to exploit knowledge about individual phonemes in a language (Scarborough & Brady, 2002). However, as with decoding ability, we found no evidence that phonological awareness correlated with either GMT or GMV in IFG, suggesting that the correlation observed in children reflects active neural development that may not be present in our more developmentally mature sample. Instead, in our participants, phonological awareness was negatively correlated with GMT in right middle temporal gyrus (see Figure 1), a region that is functionally important for discriminating sublexical speech sounds (Boets et al., 2013).

The absence of a relation between decoding skill and cortical structure in left IFG in our participants may also be related to the fact that decoding skill – which is strongly correlated with phonological awareness (in this study, r = .624, p < .0001) – decreases in importance as children grow into competent comprehenders, explaining progressively less variance related to comprehension (Goldman & Manis, 2013; Wagner et al., 1997). In contrast, the importance of experience with printed matter increases throughout maturation, from explaining 12% of the variance in oral language comprehension in kindergarten-age children to 34% in college-aged young adults (Mol & Bus, 2011). The pattern of relations that we observed for both decoding and print exposure in this region is consistent with these developmental shifts.

4.3. Left superior temporal sulcus

We observed a positive correlation between decoding ability and GMV in left STS (Figure 2). There is no analog for this finding in the previous literature investigating cortical structure in non-dyslexic readers (but see our discussion of clinical similarities below). Rather, this region was previously related to vocabulary knowledge in this population (Richardson et al., 2010), such
that greater vocabulary knowledge was associated with greater grey matter density. In addition, because the correlation was consistent across age groups, Richardson and colleagues suggested that this region might be related to processes of vocabulary acquisition that are independent of formal instruction (as opposed to their proposed relation between pSMG and vocabulary, to which we return below). We obtained no evidence that vocabulary was related to grey matter structure in this region. However, vocabulary knowledge in our sample was strongly correlated with decoding ($r = .684$, $p < .0001$), and this may be a case in which variance that might otherwise have been assigned to vocabulary knowledge may instead have been attributed to other measures that were not modelled in Richardson et al.

4.4. Right frontal/parietal regions

Several correlations emerged between cortical structure in right frontal and parietal areas and measures in our individual differences battery. As shown in Table 4, these include relations between GMT and print exposure, working memory capacity, phonological awareness, comprehension, and decoding ability; and per Table 5, between GMV and both print exposure and decoding ability. We know of no previous reports of structural relations between these brain regions and behavioural assessments of these literacy-related skills. Nevertheless, Welcome et al. (2011) observed group differences in radial expansion, a measure related to cortical surface area, in broadly-defined right frontal/parietal regions. Specifically, poor comprehenders showed smaller radial expansion in these areas than both proficient and resilient readers. To the extent that either GMV or GMT are related to radial expansion, some of our results loosely correspond with those of Welcome and colleagues. However, whereas Welcome and colleagues reported that radial expansion was relatively small in poor readers, some of our findings are in the opposite direction, with thinner GMT and/or smaller GMV associated with better performance on individual difference measures. In light of the unclear relation between radial expansion and more traditional indices of cortical grey matter structure (discussed in greater detail by Welcome et al., 2011, p. 1203), and the associated difficulty in interpreting the group differences in Welcome and colleagues’ study relative to our continuous approach, we refrain from offering any interpretation of this difference.

5. General discussion

The goal of this study was to assess potential relations between a broad array of the components of literacy and cortical grey matter structure in young adult readers. Although there are few previous research reports addressing this question, many aspects of our study make contact with earlier work, while also offering some methodological or analytic extensions. This study clearly demonstrates that mastery of fundamental skills related to proficient reading comprehension can be reflected in neurostructural characteristics of an array of language-relevant brain regions. The breadth of our cognitive assessments allowed us to both identify correlations between cortical structure and literacy skills that are consistent across previous studies in spite of methodological differences, and to assess whether individual literacy skills contributed uniquely to any structural variance observed in our imaging results. In addition, by measuring both GMV and GMT, we were able to make a detailed assessment of neuroanatomical relations to literacy skills – one which potentially accounts for differing patterns of results in these two measures. Finally, our use of a community-based sample confers several advantages to our study. This sample is more representative of the population at large than a convenience sample of undergraduates, and consequently the range of literacy and literacy-related abilities is greater than is typical in most neuroimaging studies of adult readers, which increases our ability to detect meaningful relationships between our assessments and grey matter structure.

On balance, our findings indicate that two measures – print exposure and decoding ability – are associated with grey matter structure in regions known to be related to language comprehension in general and, as we will describe below, to reading-specific behaviours in particular. This latter point is not trivial, since both print exposure and decoding skill reflect a reader’s interaction with orthographic input. Both are particularly important to the development of orthographic processing skill (i.e. the ability to construct, encode, and retrieve orthographic representations), which is thought to explain unique variance in word recognition (Cunningham & Stanovich, 1990, 1997; Stanovich, 1986; Stanovich & West, 1989). Skilled orthographic processing strengthens readers’ awareness of the phonological parameters of their language (Ehri, 1984; Nation & Hulme, 2011), enabling a tight link between words’ orthographic and phonological representations (Barron, 1986; Ehri, 1980, 1987) and, ultimately, facilitating efficient lexical access by means of these representations.

It is therefore notable that many of the cortical regions that we have discussed thus far have been functionally linked to establishing relations between orthographic and phonological information. For example, aSMG (but not pSMG) is active during the conversion of
orthographic input to licit phonological information (e.g. Booth et al., 2002; Hartwigsen et al., 2010; Richlan, Kron- 
bichler, & Wimmer, 2009; Tan et al., 2005; see also Heim et al., 2010, discussed below). In IFG, indices of print 
 exposure have been functionally related to this area’s role in print-speech convergence, such that greater 
 print exposure was related to greater functional overlap during the processing of spoken and written sen-
tences (Shankweiler et al., 2008). Left STS has been implicated in processes of orthography-to-phonology 
conversion (van Atteveldt, Formisano, Goebel, & Blomert, 2004), with dysfunction in this region indicating 
impairment of novel graphemes to licit phonemes. (Perhaps unsurprisingly, left STS is also proximate 
to heteromodal language processing regions associated with print-speech convergence; Braze et al., 2011; Frost 
et al., 2009; Shankweiler et al., 2008.) And the importance of left fusiform gyrus (and the VWFA) to word reading is 
of course well known (McCandliss, Cohen, & Dehaene, 2003).

Further, many of the right hemisphere regions in 
which we observed structural correlations with literacy 
skills have been functionally implicated in studies of 
semantic and syntactic relations, sentence comprehen-
sion, and discourse processing (e.g. Kuperberg, Lakshma-
nan, Caplan, & Holcomb, 2006; Petersson, Folia, & 
Hagoort, 2012; Prat & Just, 2011; Prat, Mason, & Just, 
2011; Robertson et al., 2000; St. George, Kutas, Martinez, 
& Sereno, 1999; Snijders et al., 2009). Right hemisphere 
homologs of “classical” left hemisphere regions associ-
ated with language processing are often concurrently 
activated during processing, and as such they might 
best be considered as part of an extended cortical 
network related to language comprehension in general 
(Hagoort, 2009).

In addition to the correspondence between the struc-
tural relations to literacy-related skills in this study and 
previous functional research, our results also support 
previous reports of a unique structural relation 
between pSMG and reading-related skills that has no 
functional analog. In previous research, grey matter 
density in this region was correlated with vocabulary, 
with greater density indicative of a more extensive voca-
ulary knowledge (Lee et al., 2007; Richardson et al., 
2010; see also Mechelli et al., 2004). Given its direct 
connections to both aSMG and AG, these studies concluded that pSMG is likely a binding site for orthographic, pho-
nological, and semantic information. In addition, they 
proposed that because this relation was obtained only 
in adolescents, it might be a cortical reflection of 
formal instruction in the service of vocabulary acquisition 
(see also Richardson & Price, 2009).

We concur with the broader conclusion of these 
studies regarding SMG’s potential relevance to the 
assembly of high-dimensional lexical representations. 
However, our observation of correlations between 
decoding ability and aSMG, and print exposure and 
pSMG, suggest a revised account of the link between 
pSMG and vocabulary. We agree that aSMG may be 
chiefly concerned with phonological aspects of gra-
pheme-to-phoneme translation. We suggest that 
pSMG, in contrast, may be primarily concerned with 
orthographic aspects of this process; that is, the structure 
of pSMG may not reflect the number of lexical forms 
added to a reader’s vocabulary, or the way in which 
these forms were acquired (Lee et al., 2007; Richardson 
& Price, 2009; Richardson et al., 2010), but the amount 
of reading that a person actually does. This explanation is 
more parsimonious than the proposal that pSMG is 
related to vocabulary instruction – an explanation that 
was always somewhat problematic given that the 
relation with vocabulary knowledge was not present in 
children, in whom increasing vocabulary knowledge 
has instead been related to cortical thinning in left parie-
tal regions (Sowell et al., 2004; see also Linkersdörfer 
et al., 2015). This explanation is also consistent with 
SMG’s well-established functional role in linking gra-
phemes to their phonemic equivalents during reading.

It is worth noting that our study reports different cor-
tical grey matter measurements than those studies 
examining the pSMG and vocabulary knowledge. We 
report correlations with GMT and GMV, whereas Lee 
et al. (2007) and Richardson et al. (2010) assessed 
relations to grey matter density, an index of the ratio 
of grey matter voxels to those of other types of brain 
tissue in a particular cortical region (for a detailed dis-
cussion, see Mechelli, Price, Friston, & Ashburner, 2005). 
Because grey matter density, GMT, and GMV are partially 
independent measures, it is reasonable to expect them 
to show different patterns of results. Further, as we dis-
cussed in section 4.1 above, experience with printed 
material is an established precursor to vocabulary knowl-
edge (see also Cunningham & Stanovich, 1991; Stanovich 
& Cunningham, 1992; Stanovich & West, 1989). By asses-
sing print exposure and vocabulary together, our study 
examined whether these measures explained unique 
variance in our indices of cortical structure. Our findings 
in this region (and others, such as left STS) suggests that 
it is important to assess skills linked to multiple compo-
ments of reading ability, rather than a single measure 
(or a small set of measures), in order to determine 
whether effects are related to common variance shared 
among the measures, or specific contributions of particu-
lar cognitive abilities.
Finally, there is also substantial agreement between our findings and the results of neuropsychological studies of group differences between dyslexic and typical readers. Although there is considerable heterogeneity among the results of such studies, two recent meta-analyses clearly identify several points of convergence. For example, Linkersdörfer, Lonnenmann, Lindberg, Hasselhorn, and Fiebach (2012, N = 277) identified four cortical areas in which dyslexic readers consistently exhibit smaller GMV: bilateral SMG, left fusiform gyrus (including VWFA) and the right superior temporal gyrus (STG). In contrast, Richlan, Kronbichler, and Wimmer (2013, N = 266) found only two: right STG and left STS. In our study, we observed significant relations between literacy skills and all of these regions that are consistent with this literature. Decoding ability was positively correlated with GMV in left SMG and left STS, and with GMT in right STG in our community-based sample, such that poor decoders also had smaller GMV/thinner cortex than those with more efficient decoding ability. Given the obvious relation between dyslexia and decoding ability, the parallel between our results and the clinical pattern is clear, with some functional imaging research explicitly suggesting that these group differences are related to impaired grapheme-to-phoneme conversion (e.g. Heim et al., 2010). Furthermore, we also observed a relation between GMT and both print exposure and reading comprehension in an area of left inferior temporal gyrus (ITG) corresponding to the fusiform gyrus. Here the correlation was negative, with greater experience with printed matter and greater reading comprehension ability coinciding with thinner grey matter. However, we do not believe this difference suggests any meaningful conflict between ours and the neuropsychological studies. Rather, the relation of GMV to group differences related to dyslexia and, in our study, of GMT and the development of skilled reading comprehension are parallel, rather than contradictory, findings.

6. Conclusions and future directions

Skilled reading is the result of the fluent orchestration of numerous component processes, some of which are thought to be linguistically-based, whereas others are domain-general. The results of this exploratory study confirm, extend, and refine many of the diverse findings of previously published work. In addition, this study clearly demonstrates the importance of taking a multifaceted approach to investigating the neurostructural correlates of skilled reading comprehension. This is true not just for future investigations of cortical grey matter, but for investigations of other aspects of brain morphology as well. For example, there are a small number of studies in which literacy skills have been correlated with white matter structure. We have mentioned one such study, in which vocabulary knowledge informed our understanding of the structural significance of pSMG (Lee et al., 2007). There are others exploring relations between the connectivity of regions in the left-hemisphere reading network and, for example, decoding ability (Welcome & Joanisse, 2014; Zhang et al., 2014; in dyslexic readers, see Pugh et al., 2000; Steinbrink et al., 2008). Taking white matter structure into account could be particularly important since increased white matter myelination decreases the space between the cortex and the skull, potentially contributing to grey matter thinning (Lu et al., 2007; Sowell et al., 2004). However, as in previous investigations of cortical grey matter, most such studies have been narrowly construed, suggesting that a broad-based approach to assessing participants’ literacy skills could be similarly applied, with the same potential benefits to analysis and interpretation (e.g. Van Dyke et al., 2015).

Indeed, we view this approach to assessing individual cognitive differences as essential to future research into the neural architecture supporting reading comprehension. This is consistent with recent behavioural research on the relation between individual cognitive variation and skilled reading. Although our own battery of assessments is more extensive than previous work, it is hardly exhaustive; large-scale behavioural studies typically administer many more assessments than were included in our project. For example, studies such as Van Dyke et al. (2014) and Freed et al. (2017) both assessed performance on more than two dozen measures. The benefits offered by such an approach are twofold. First, expanding the battery of assessments allows the testing of theoretically important constructs for which previous research (including our own) does not account. For example, as noted by Freed and colleagues, individual differences in both perceptual speed and suppression/inhibition ability, neither of which have been assessed in studies such as ours, are both likely contributors to comprehension performance. Second, more extensive test batteries allow theoretical constructs to be measured with multiple assessments, which would permit the construction of more stable composite predictors (cf. Braze et al., 2016). This is particularly important for non-standardised instruments, such as the complex span tasks that are in common use to assess working memory capacity: the test-retest reliability of these measures is known to be relatively low, and a composite based on two or three assessments (e.g. sentence span, operation span, symmetry span, etc.) has much greater predictive stability (Waters & Caplan, 2003).
Finally, as noted in our introduction, most of this research has been conducted with adult readers. However, studies such as Lee et al. (2007) and Richardson et al. (2010) make it clear that assessing developmentally distinct groups of readers is important for understanding neuro-structural correlates of literacy-related skills. For example, as discussed above, experience with printed material increases in importance as readers age, while decoding ability experiences a corresponding decrease in importance (Mol & Bus, 2011); this yields testable predictions about both when and where these abilities should have correlates in the brain. In light of evidence that both grey and white matter are sensitive to efforts at remediation in poor readers (e.g. Keller & Just, 2009; Krafnick, Flowers, Napoliello, & Eden, 2011), a comprehensive assessment of both cortical structure and the precise cognitive capacities that support the achievement of literacy is essential for our understanding of the neural architecture supporting skilled reading comprehension.

Notes
1. Although, as noted in the Method, we have reservations about using variance inflation factors (VIFs) to index multicollinearity, we nonetheless calculated them for our predictors. The VIFs converged with the results of the condition numbers: their range (1.43–6.08) was not only below the commonly recommended threshold of 10 (e.g. Cohen, Cohen, West, & Aiken, 2013; James, Witten, Hastie, & Tibshirani, 2013; Kutner, Nachtsheim, Neter, & Li, 2005), but also below a more conservative recent recommendation of 7 (Keith, 2014).
2. In line with recent recommendations (e.g. Eklund, Nichols, & Knutsson, 2017), all structural MRI data and analysis scripts are available at OpenNeuro.org, a free online repository for neuroimaging data: https://openneuro.org/datasets/ds001365/versions/00001.
3. Similarly, Goldman and Manis (2013) observed a relation between SMG and the amount of pleasure reading their participants engaged in outside of instructional settings. Although the absence of MNI coordinates in this study does not permit assessment of the locus of their SMG correlations, we mention it here as weak converging evidence for our account of pSMG’s structural relation to reader characteristics.

Acknowledgements
We are grateful to Jessica Grittner for assistance with data collection, to Morgan L. Bontrager, Alexis R. Johns, Leonard Katz, Victor Kuperman, and Ashley G. Lewis for insightful comments and advice on earlier drafts of this paper, and to P. R. Nelson for support and inspiration.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This research was supported by the following grants to Haskins Laboratories from the Eunice Kennedy Shriver National Institute of Child Health and Human Development: P01 HD-01994 (Jay G. Rueckl, PI), R01 HD-040353 (Donald P. Shankweiler, PI), R01 HD-071988 (David Braze, PI), R01 HD-073288 (Julie A. Van Dyke, PI), and NRSA HD-080331 (Dave Kush, PI).

ORCID
Clinton L. Johns http://orcid.org/0000-0003-0593-7508

References


hemisphere during discourse processing. *Brain*, 122(7), 1317–1325.


