NeuroImage 237 (2021) 118087

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

A three-time point longitudinal investigation of the arcuate fasciculus throughout reading acquisition in children developing dyslexia

Stijn Van Der Auwera [a], Maaike Vandermosten [b], Jan Wouters [b], Pol Ghesquière [a], Jolijn Vanderauwera [a,b,c,d,*]

[a] Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, KU Leuven, Belgium
[b] Research Group ExpORL, Department of Neurosciences, KU Leuven, Belgium
[c] Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[d] Institute of Neuroscience, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Keywords:
Developmental dyslexia
Reading
White matter
Diffusion Tensor Imaging
Automated Fiber Quantification
Longitudinal design

ABSTRACT

Although the neural basis of dyslexia has intensively been investigated, results are still unclear about the existence of a white matter deficit in the arcuate fasciculus (AF) throughout development. To unravel this ambiguity, we examined the difference in fractional anisotropy (FA) of the AF between children developing dyslexia and children developing typical reading skills in a longitudinal sample with three MRI time points throughout reading development: the pre-reading stage (5–6 years old), the early reading stage (7–8 years old) and the advanced reading stage (9–10 years old). Applying along-the-tract analyses of white matter organization, our results confirmed that a white matter deficit existed in the left AF prior to the onset of formal reading instruction in children who developed dyslexia later on. This deficit was consistently present throughout the course of reading development. Additionally, we evaluated the use of applying a continuous approach on the participants' reading skills rather than the arbitrary categorization in individuals with or without dyslexia. Our results confirmed the predictive relation between FA and word reading measurements later in development. This study supports the use of longitudinal approaches to investigate the neural basis of the developmental process of learning to read and the application of triangulation, i.e. using multiple research approaches to help gain more insight and aiding the interpretation of obtained results.

1. Introduction

Learning to read is an effortful long-term learning process, that requires formal reading instruction (Dehaene, 2009; Vogel et al., 2013). Nevertheless, most young children develop reading skills rather fluently. However, a proportion of children struggle to develop adequate reading and writing skills due to developmental dyslexia. Children with developmental dyslexia are defined as individuals who have persistent difficulties with accurate and/or fluent word recognition and spelling that cannot be explained by a lack of instruction, intelligence and sensory abilities (Peterson & Pennington, 2012). Difficulties have been estimated to occur in approximately 7% of the population (Peterson & Pennington, 2012).

Despite the enormous research interest in dyslexia, the underlying cause of the reading impairments remains unclear. The most commonly accepted underlying cognitive deficit in dyslexia is the phonological deficit (Pennington & Lefly, 2001; Snowling, 2000). This phonological impairment has even been shown to precede the start of formal reading instruction (Boets et al., 2010; Ziegler et al., 2008). However, it is important to note that poor phonological skills not always result in reading impairment and not all children with dyslexia have a phonological deficit (Boets et al., 2007; Law et al., 2014). It is currently accepted that dyslexia more likely results from a multiple deficit (Pennington, 2006), pointing towards interactive deficits at the cognitive, neural, genetic and/or environmental level (Pennington, 2006; van Bergen et al., 2014).

At the neural level, the ability to read requires the integration of pre-existing networks used for e.g. language and visual processing (Dehaene, 2009). Hence, when individuals learn to read, an entire network is being reorganized in a relatively short period of time. In advanced readers, reading relies on widespread regions predominantly located in the left hemisphere of the brain. These regions are connected to each other by means of white matter pathways (Wandell & Yeatman, 2013). When investigating neural correlates of reading, the importance of these white matter connections involved in reading has been acknowledged (Boets et al., 2013; Wandell & Yeatman, 2013). The white

* Corresponding author.
E-mail address: jolijn.vanderauwera@uclouvain.be (J. Vanderauwera).

https://doi.org/10.1016/j.neuroimage.2021.118087.
Received 18 October 2020; Received in revised form 6 April 2021; Accepted 8 April 2021
Available online 18 April 2021.

matter reading network consists of a broad set of, predominantly left hemispherical, pathways including the inferior longitudinal fasciculus, the inferior fronto-occipital fasciculus, the uncinate fasciculus, the superior longitudinal fasciculus, the arcuate fasciculus and the corpus callosum. While all of these white matter connections have been linked to reading processes, the left dorsal arcuate fasciculus (AF) has been most extensively investigated with respect to reading and dyslexia.

Despite the strong research interest in the AF, the obtained results remain divergent and inconclusive. On one hand there is a body of research pointing towards a key role of the AF for reading, connecting key regions of the reading network in the left hemisphere (Catani et al., 2005), i.e. the inferior frontal region, the temporo-parietal region and the occipito-temporal region (Pugh et al., 2000). These studies mostly investigated fractional anisotropy (FA), a quantitative measure for the structural organization of white matter pathways. Results of these studies indicated that FA in the fronto-temporal segment of the AF is related to phonological awareness in adult readers (Vandermosten, Boets, Poelmans, et al., 2012), and to grapheme-phoneme integration and word learning in school-aged children (Gullick & Booth, 2015) and adults (López-Barroso et al., 2013). Research in adults with dyslexia also indicated a different structural organization in the left AF compared to typical readers (Vandermosten, Boets, Poelmans, et al., 2012). With the aim to disentangle whether altered structural organization of the AF is a consequence of reading failure or rather precedes the onset of reading instruction, over the last years longitudinal studies were conducted that started prior to the onset of reading instruction and included children with an elevated risk for developing dyslexia. These studies revealed that structural organization of the AF can be related to phonological processing, an important predictor of reading outcome, already prior to the start of formal reading instruction (Kraft et al., 2016; Saygin et al., 2013; Vanderauwera et al., 2015, 2017; Vandermosten et al., 2015). Anomalies in the white matter organization of the AF were demonstrated in pre-reading children with a familial risk for dyslexia (Wang et al., 2017), a finding that was even confirmed in infants (Langer et al., 2017). FA differences in the AF have been related to having a cognitive risk for dyslexia (Saygin et al., 2013), or even to be specific to the group of children who developed poor reading (Kraft et al., 2016; Wang et al., 2017) or dyslexia (Vanderauwera et al., 2017). Although the majority of these studies have focused on the left hemispherical AF, studies in children have also pointed towards an involvement of the right hemispherical counterpart. Wang et al. (2017) reported rightwards lateralization of the AF in pre-readers with a familial risk for dyslexia. Vanderauwera et al. (2017) found that children with dyslexia exhibit pre-reading white matter anomalies in both left and right AF.

Despite the results of several studies pointing towards a deficit in the left long AF segment in individuals with (a risk for) dyslexia, a recent Activation Likelihood Estimation (ALE) meta-analysis (Moreau et al., 2018), a technique that quantifies convergence of activation across studies, argued there is no evidence for a systematic white matter difference between children developing dyslexia and children developing typical reading skills. However, in ALE meta-analyses only voxel-based analyses can be taken into account, while most recent studies use a tractography approach to investigate anomalies in the AF. In general, these tractography studies show consistent deficits in pre-readers linked to having an elevated risk for dyslexia, although it should be noted that anomalies related to the family history of dyslexia were not consistently retrieved among all studies (Vandermosten et al., 2015). While the study by Vanderauwera et al. (2017) suggested lower FA in the left and right AF for children developing dyslexia compared to children developing typical reading skills at the pre-reading stage, this difference could not be traced after two years of reading and writing instruction. We argue that further examination of these results is warranted, especially because a difference in white matter structure is also reported in the left AF in adults (Vandermosten, Boets, Poelmans, et al., 2012).

For this divergence in findings, multiple reasons can be put forward. First, when conducting research on the underlying neural factors of reading and dyslexia, different study designs can be applied, i.e. cross-sectional and longitudinal designs. Second, different approaches are used to investigate the relation between reading and white matter organization. To start with, researchers can opt to divide all participants into certain categories versus choosing a continuous approach. On top of these differing distributions, the large discrepancies with respect to criteria that are used to assign participants to a certain group might influence research findings. Third, one must acknowledge that there are multiple approaches and procedures to analyze neuroimaging data, supported by a large variety of software packages. Consequentially, all these differences in approaches can impede the interpretability of results across studies. In the next section, we will elaborate further on these three reasons and their implications for research on the neural correlates of reading and dyslexia.

1.1. Study designs

As previously stated, discrepancies might be explained by the differing designs between previous studies. Two frequently employed designs for data collection and analyses are the cross-sectional and longitudinal design. Although both designs can provide insights into the dynamics of various processes, the longitudinal design has some clear advantages in terms of capturing developmental trajectories. In dyslexia research, unclarity remains concerning the developmental pattern of neuroanatomical alterations. Some studies reported on different developmental patterns for the reading network in pre-readers (Vanderauwera et al., 2017) and beginning or advanced readers (Lebel et al., 2019). One study even reported a decline of FA in the left AF in a group of poor readers (Yeatman, Dougherty, Ben-shachar, et al., 2012). While the majority of longitudinal studies investigated the relation between cognitive and neural measures starting at kindergarten age and older, the study by Zuk et al. (2019) revealed that white matter organization in infancy can be predictive of phonological skills at the kindergarten age. Another study applied a longitudinal MRI design to investigate the white matter dynamics paralleling an intensive reading intervention in school-aged children (Huber et al., 2018). Despite some exceptions, the majority of studies in the field comprise a single MRI measurement. Therefore, investigations of the longitudinal development of these neural deficits are scarce. Hence, the stability or variability of the neural alterations in relation to the development of dyslexia remain to be investigated.

To be sensitive to different developmental trajectories, research needs to include at least three measuring points (King et al., 2018). Doing so will help to obtain a clear view on the development of the structural brain organization, and will also positively influence statistical power (Lu et al., 2013). Due to this gain in power, sample sizes required to detect an effect of a given size can be relatively smaller, which is of great importance in MRI studies (especially studies containing young children).

1.2. Continuous vs. Categorical approaches

The choice for a categorical or a continuous approach of studying reading skills can influence the interpretability and comparability of the wide array of neuroimaging studies conducted in children developing dyslexia and children developing typical reading skills. The idea of applying a strict cut-off (categorical approach) in a population sample based on one reading measure to evaluate whether a child is seen as dyslectic reader or typical reader is contradictory to the normal distribution of reading skills in the population and is inevitably arbitrary (Peterson & Pennington, 2012). A presumably more accurate way is to look at reading ability as a broad spectrum, of which children we classically assign to the group of dyslexic readers are merely a small proportion. This has been done before, as previous studies have identified regions where structural organization significantly correlates with performance on reading tasks. Positive correlations between structural organi-

zation in the left temporoparietal area of dyslexic or typical readers and reading ability have been reported in previous research (Moreau et al., 2018; Vandermosten, Boets, Wouters, et al., 2012).

1.3. Neuroimaging analysis approach

Furthermore, we emphasize that the use of different data analytic approaches might interfere with general interpretability. Choosing certain software packages or procedures for data analyses can have unintended consequences for the obtained results. Researchers usually select the approach that fits their research goals the best, however, the risk is that different approaches for examining white matter structure might generate slightly differing results.

For example, one could use Automated Fiber Quantification (AFQ; Yeatman, Dougherty, Myall et al., 2012) to automatically perform tractography, ROI-based fiber tract segmentation and quantification of diffusion characteristics at multiple nodes along the trajectory of the fiber to examine white matter structure. In contrast, other authors might choose to manually delineate white matter pathways using a region of interest (ROI) approach to define obligatory passages (Catani & de Schotten, 2008). To complete this manual process specific ROIs are added to detect and delete streamlines that are no part of the pathway of interest (NOT-ROIs; Catani & de Schotten, 2008; Wakana et al., 2007).

The quality of structural white matter connections can be quantified by means of different indices. Fractional anisotropy (FA), an indicator for the degree of diffusion anisotropy, has been most widely applied (Wang et al., 2017). High anisotropy reflects fast diffusivity along the fibers, high isotropy reflects slow diffusivity (Assaf & Pasternak, 2008). The quantitative value for FA is driven by microstructural properties such as myelinization and axon density but moreover, macrostructural properties such as fiber crossings also have an influence on FA (Alexander et al., 2010). In reading and dyslexia research, the FA index has been reported in many different studies. Note that in the aforementioned comparison of techniques, both Wang et al. (2017) and Vanderauwera et al. (2017) used FA as a measure for structural organization. However, as was briefly described above, their methods varied enormously. In the study by Vanderauwera et al. (2017) the authors made use of a manual tractography by which a mean FA value for a tract is calculated. In comparison, the automated method applied by Wang et al. (2017) makes it possible to examine FA values in more detail, because a Tract Profile is calculated for every tract (Yeatman, Dougherty, Myall, et al., 2012). Tract Profiles contain far more information than mean diffusion measures as they quantify FA at various nodes along each tract. Hence, study results might to differ depending on the applied approach.

1.4. The present study

The present study includes MRI assessments from three time points throughout reading development (pre-reading, early reading and advanced reading stage) and data from annual cognitive assessments, allowing for a better estimation of longitudinal trajectories (King et al., 2018). It is important to notice that the study by Vanderauwera et al. (2017) was based on the first two time points of data we analyze in the current study. In an attempt to clarify the aforementioned inconsistencies in results between studies, we aim to investigate three main research questions. (A) First, we examine the difference in FA of the AF long segment between children developing dyslexia and children developing typical reading skills throughout primary school. We hypothesize to find a difference in FA at the pre-reading stage. Expectations regarding school-aged children are more hampered due to the inconsistencies in previous studies. (B) Second, the potential added value of a continuous approach compared to a categorical classification will be investigated. (C) Finally, previous results reported on this sample (see Vanderauwera et al., 2017) were based on mean FA-values for each tract. We aim at supplementing these previous findings by applying a method that allows for investigating FA-properties along the course of a tract. Therefore, we examine whether comparable results are obtained when applying the AFQ procedure.

2. Method

2.1. Participants

This study is part of a longitudinal project in which 87 kindergartners participated. The longitudinal project included behavioral assessments, MRI examinations and EEG examinations prior to the start of formal reading instruction and during reading development. The present study focusses on the longitudinal behavioral and MRI measures. All 87 participants (except one) were children who had not received any formal reading and/or writing instruction at the start of the study in last year of kindergarten.

Specific inclusion criteria, measured in kindergarten, were applied for this longitudinal study (Vanvooren et al., 2014): (1) a non-verbal IQ $\geq$ 80, based on the Coloured Progressive Matrices (Raven et al., 1984) (2) normal hearing (i.e. a Fletcher index of less than 20 dB HL), (3) monolingual native Dutch speaking, (4) no history of brain damage, vision deficits, or articulatory problems, and (5) no high risk for developing ADHD. At the start of second grade non-verbal intelligence was tested again with the WISC-III-NL subtest Block Design (Wechsler, 2005). In the original sample (Vanvooren et al., 2014), 44 children were selected because of their elevated risk for developing dyslexia, based on having a first-degree relative with dyslexia (FRD^+). These children were matched individually to children who had no familial risk for dyslexia (FRD^-) based on sex, age, parent's socio-economic status (SES) assessed with the Family Affluence Scale (Boudreau & Poulin, 2009), non-verbal intelligence, and school environment (i.e. same class). Although this study does not comprise data of all 87 subjects, no group differences were found for these matching variables (see Results section).

During the summer prior to first grade, 75 pre-readers (5–6 years old) participated in the MRI examinations of the current longitudinal study. During the summer prior to third grade, when the children were 7–8 years old (i.e. early readers), MRI scans were conducted in 65 of the previously examined children. The third MRI scan was conducted in 63 children (9–10 years old, advanced readers) during spring of fifth grade. Children were included based on willingness to participate. The local ethical board of the University Hospital approved this study and informed consents were obtained for all participants according to the Declaration of Helsinki. Participants were not asked to give consent for their data to be made publicly available, so supporting data is not available.

Of the children in whom scans were conducted, some were excluded from the study based on the following exclusion criteria. At the *pre-reading stage* nine participants were removed due to inadequate data collection and one participant was removed based on excessive motion. In addition, seven participants were not included because no left long AF segment could be tracked using AFQ (see section 2.4.2 on Automatic Fiber Quantification), which is essential for the aim of the present study. The absence of a left AF reconstruction has been reported in previous studies in children (Eluvathingal et al., 2007; Vanderauwera et al., 2017) and infants (Dubois et al., 2009; Langer et al., 2017) and can be presumed to be linked to the sharp curvature and low anisotropy in the region of the principal arc of the tract, as well as the unstable neural location of the tract (Duan et al., 2015). Among these seven children, there were three children who developed dyslexia (DR) and three FRD^+ children. In addition, six participants were excluded for co-occurring disorders, i.e., dyscalculia, autism spectrum disorder and attention-deficit (hyperactivity) disorder. In total, data of 52 participants from the pre-reading stage were included in the present study (27 FRD^+; Table 1). Because only children in which scans were conducted in the previous scanning period were suited to be analyzed, a maximum of 52 children

Table 1
Number of participants at each stage.

Time point		DR	TR	FRD$^+$	FRD$^-$
Pre-reading stage	N = 52	16	36	27	25
Early reading stage	N = 45	16	29	27	18
Advanced reading stage	N = 35	12	23	20	15

DR = Child developing dyslexia, TR = Child developing typical reading skills, FRD$^+$ = Child with familial risk for dyslexia, FRD$^-$ = Child without familial risk for dyslexia

were available for the early reading stage. Due to this criterion, and the loss of some data, the sample size decreases throughout the consecutive scanning periods. At the *early reading stage* four children did not participate, two children were excluded due to inadequate data acquisition and one additional child (FRD$^-$TR) was removed because no left long AF segment could be tracked in the second MRI scan (in line with Vanderauwera et al., 2017). Thus, data of 45 children were included at the early reading stage (27 FRD$^+$; Table 1). At the *advanced reading stage,* six children did not participate, three children were excluded due to inadequate data acquisition and one FRD$^+$TR child was removed because no left long AF segment could be tracked in the third MRI scan. Therefore, the advanced reading stage consisted of 35 children (20 FRD$^+$; Table 1). For all children that were included based on the preceding procedure, the right AF was also traced. However, in line with previous studies (Ozernov-Palchik et al., 2019; Vanderauwera et al., 2017; Vandermosten, Boets, Poelmans, et al., 2012; Yeatman et al., 2011), this tract could not be traced in a considerable amount of children, reducing data for the right AF to 34 pre-readers (9 DR; 15 FRD$^+$), 28 early readers (9 DR; 15 FRD$^+$) and 22 advanced readers (5 DR; 10 FRD$^+$). Due to the small sample size, especially the small number of DR, analysis for the right AF were limited (see Results section). It is important to note that one child, included in all MRI scans, doubled the first year of elementary school. Another child who participated in the pre- and early reading stage, changed from a Flemish to a French school after first grade. One additional child, who participated in the first two MRI scans, doubled the third year in kindergarten and subsequently was in a lower grade compared to children of the same age. Note that removing these participants from the analyses did not change the results.

2.2. Identification of children developing dyslexia

We examined which children met the criteria to be identified as DR, based on their reading ability. Children who scored below percentile ten on one of the reading tasks (either word reading or pseudo word reading; see 2.3. Cognitive assessment) at the three last time points, in case at least three time points were available, were classified as DR (n = 15). In case data from only two time points was available, children who scored below percentile ten on both time points (for either one of the reading tasks) were classified as DR (n = 1). This criterion thus classified sixteen children as DR out of the sample of 52 in the pre-reading stage, of whom twelve where FRD$^+$. At the early reading stage, data of all sixteen DR were available and at the advanced reading stage data of twelve DR were available (Table 1).

2.3. Cognitive assessment

A summary of all cognitive measurements that were conducted approximately each year as part of this longitudinal project can be found in Vanderauwera et al. (2017). Important for this study, an identical word reading test (Brus & Voeten, 1973) and pseudo-word reading test (Van den Bos et al., 1994) was assessed from second to fifth grade. Both tests consisted of a list of (pseudo-)words which the children were asked to read aloud as accurate and fast as possible. The score was defined as the number of correct read items in one minute (for the word reading test) or two minutes (for the pseudo-word reading test). A writing on dictation test by Dudal (1997), adjusted to the grade of the children, was used to administer spelling skills.

2.4. DW-MRI acquisition

MRI acquisition was identical over all three time points. Single shot EPI with SENSE (parallel) MRI scans were conducted with a 32-channel head coil on a 3T MRI scanner (Philips, Best, The Netherlands). The following parameters were used obtaining sagittal diffusion imaging slices: repetition time 7600 ms, echo time 65 ms, flip angle 90°, voxel size 2.5 × 2.5 × 2.5 mm, 60 non-collinear directions, b-value 1300 s/mm^2, 6 nondiffusion-weighted images. The scan acquisition time was 10:32 min.

2.4.1. Image processing

DWI and T_1-weighted structural images were converted to respectively 4D and 3D NifTI format using dcm2nii software (Li et al., 2016), part of MRIcron (www.nitrc.org/projects/mricron). The MCFLIRT tool (Jenkinson et al., 2002) from FSL software was used to perform motion correction on the DWI data. DTI preprocessing consisted of three steps, (1) correction for eddy current spatial distortions, (2) creating a binary mask of the brain and (3) calculating the tensors. Various functions of the FSL software package were used to execute these steps. FSL's Diffusion Toolkit was used for eddy current correction and outputs an eddy corrected dataset in NifTi format. A binary brain mask of the corrected image was created using BET Brain Extraction. The default fractional intensity threshold of .05 was retained. DTIFit Reconstruct diffusion tensors from FSL's Diffusion Toolkit was used to fit tensors to the diffusion weighted images, again by using the default parameters. To align the T_1-weighted anatomical volume into AC-PC space the mrAnatAverageAcPcNifti function from VISTASOFT (www.github.com/vistalab/vistasoft) was used. To execute the realignment, the anterior commissure (AC) and posterior commissure (PC) were manually specified as landmarks and a mid-sagittal reference point was specified to help the alignment algorithm. This procedure resliced each scan to 1 mm isotropic voxels and saves the resliced ACPC data to a NifTi file. The dtiMakeDt6FromFsl function was used to create a dt6 MATLAB file. When providing a T_1-weighted image on top of the DTI file this function attempts to align them and saves them out in a dt6 MATLAB file. Quality control of the data was executed through visual inspection using mrDiffusion, a toolbox from the VISTALab (Stanford Vision and Imaging Science and Technology) diffusion MRI software (http://vistalab.stanford.edu/). Control for head motion during the MRI scans was checked by calculating the root mean square (RMS) of all six rigid body translation/rotation parameters, i.e., the absolute displacement and rotation in all three directions (Theys et al., 2014). Data from children whose RMS indicated average movement greater than half the voxel size, was labeled as excessive motion and excluded from this study. As stated before, only one participant was removed due to this criterion. Furthermore, we detected no difference in subject motion between DR and TR at the pre- ($T_{(50)}$ = 1.439, p = .156), early ($T_{(43)}$ = -0.211, p = .834) and advanced (W = -1.700, p = .113) reading stage, or between FRD$^+$ and FRD$^-$ at the pre-reading stage ($T_{(50)}$ = 0.274, p = .786). All reported results were not affected by adding subject motion into the analysis (see results).

2.4.2. Automated Fiber Quantification

Automated Fiber Quantification (AFQ) software was used to identify the AF and thereafter calculate FA at 100 nodes along the trajectory of this fiber. Only a brief description of the AFQ analysis pipeline is provided here, a more detailed description can be found in Yeatman, Dougherty, Myall et al. (2012). First, the AFQ pipeline performs whole brain tractography where all fibers are tracked using deterministic tractography. Two criteria are used for streamline delineation, the tracking is halted if FA estimated at the current position is below .2 and the minimum angle between the last path segment and next step direction is greater than 30°. In the second phase the whole

brain fiber group is segmented into different white matter pathways defined in the white matter atlas (Wakana et al., 2007). Two ROIs are used to select candidate fibers for each fascicle in an automated approach. Eventually fiber tracts are refined by comparing all fibers in the fascicle to a probabilistic fiber tract atlas. The AFQ pipeline proceeds with deleting outliers, i.e. fibers that deviate substantially from the core of the fiber group are removed. In the last step tract properties are determined, diffusion measurements are computed along the trajectory of the fiber group calculating FA for 100 nodes along the tract. While conducting the AFQ pipeline, the default parameters described in AFQ_Create (https://github.com/yeatmanlab/AFQ/wiki) were used with only one exception: we chose to analyze the entire fiber tract in contrast to the prescribed setting where each fiber is clipped to the ROIs. Using this method allowed for analyzing cortical endpoints in the segmented fibers.

2.5. Statistical analysis

All statistical analyses were conducted using JASP (JASP Team, 2020). Because group differences detected in brain research are usually rather small, we used a specific procedure to detect nodes with a difference in FA between groups, that diminishes the chance on type I and type II errors. Simply using a Bonferroni correction to adjust for the 100 group comparisons, for each of the 100 nodes specified by AFQ, would greatly increase the chance of a type II error. Therefore, in the pre-reader categorical group comparisons, using an independent samples t-test, only when at least two consecutive nodes reached significance at the $p = .01$ level, they were taken into further analysis. All nodes adjacent to these nodes that reached significance at the $p = .05$ level, were also included into further analysis. The cluster of nodes that was revealed using this procedure were the only nodes further analyzed in both the longitudinal and continuous analysis. To provide additional evidence for the specific procedure described above, the mean absolute effect size (Cohen's *d*) was calculated for the detected cluster of nodes and compared to the mean absolute effect size for all 100 nodes. Furthermore, a comparison for mean diffusivity (MD), was added for the pre-reader categorical group comparison and the longitudinal comparison in the nodes where significant differences were found for FA. MD is used as a complementary measure to examine differences of structural integrity, as it describes the average mobility of water molecules. While FA is known to increase, MD decreases with age in early childhood, as a result of the decrease of water content and increase of myelin (Hüppi & Dubois, 2006). For the longitudinal, continuous and descriptive analysis, significance was set at $p = .05$ and False Discovery Rate (FDR) correction was applied per grade, according to the Benjamini and Hochberg's procedure, as described in Benjamini and Yekutieli (2001).

2.5.1. Pre-reader categorical group comparisons

For the initial categorical group comparison between DR and TR at the pre-reading stage, the Shapiro-Wilk Normality test and Levene's Test for Equality of Variances were used to determine whether an independent samples t-test or a Mann-Whitney U test should be used. To investigate to what extend white matter differences between TR and DR children were solely driven by reading ability, and not by the familial risk for dyslexia, additional independent samples t-tests were conducted. For these analyses, the TR group was restricted to $FRD^{-}TR$ children, and this group was compared to all DR children (75% FRD^{+}). To decide which nodes contained a valuable difference, the previously described rules for significance were applied.

To examine the effect of having a familial risk for dyslexia on the brain, group comparisons were conducted using a Mann-Whitney U test or independent samples t-test to compare FA of all 100 nodes between FRD^{+} and FRD^{-} children. Because the distribution of TR and DR was unequal in these groups, additional analysis were executed. For this inquiry the group of children was limited to the TR, Mann-Whitney U test or independent samples t-test were used to compare FRD^{+} children and FRD^{-} children. In this way the potential influence of TR and DR in these groups was avoided.

2.5.2. Longitudinal analysis

Among all participants, the left AF could be traced at all three reading stages for 35 children (23 TR and 12 DR). Only the cluster of nodes that was revealed in the pre-reader categorical group comparison was used in the longitudinal analysis. We examined the maturation, and potential group differences over time for these children in a longitudinal analysis by means of repeated measures ANOVA with group (TR and DR) as between subjects factor, time point (pre-reading, early reading and advanced reading) as within subject factor and FA or MD as dependent variable. To additionally investigate whether white matter differences between TR and DR children were driven by reading ability regardless of familial risk, similar repeated measures ANOVA were run with time point (pre-reading, early reading and advanced reading) as within subject factor, group ($FRD^{-}TR$ and DR) as between subjects factor and FA or MD as dependent variable in a subsample of 23 children (11 TR and 12 DR). To further examine potential interactions for group and time in the above repeated measures ANOVA's, i.e. different maturation patterns for TR versus DR and $FRD^{-}TR$ versus DR, we used Bayesian statistics. BF inclusion (BF_{incl}) was calculated for the interaction term in the model. This Bayes Factor is a quantification of the comparison between a model with all variables without the interaction term and a model including the interaction term, i.e. a measure for the additional value of this predictor to the model. We calculated BF_{incl} specifically for the interaction term, as we expected to find no differences for this predictor, since frequentist statistics are less suitable to test the hypothesis that no differences are present. The classification scheme from Andraszewicz et al. (2015) was used to interpret all BF_{incl} values.

2.5.3. Continuous analysis

To investigate the possibility of a more continuous approach concerning the difference between TR and DR, the cohesion between FA and the score on a word reading task was calculated. As previously described, only the cluster of nodes that was revealed in the pre-reader categorical group comparison was used in this analysis. At every reading stage data from all available children was used for this analysis. Therefore, data from 52, 45 and 35 children were available for respectively the pre-reading stage, the early reading stage and the advanced reading stage. By means of linear regression we examined to what extend FA was a predictor for a word reading task conducted in second to fifth grade. Motion was added into this analysis as a covariate, this predictor never reached significance.

3. Results

3.1. Descriptives of participants

Descriptive statistics and cognitive measurements are summarized for the pre-reader categorical group comparison and for the longitudinal comparison. All participant characteristics were compared between DR and TR groups on the one hand and between FRD^{+} and FRD^{-} groups on the other hand, see Table 2 and Table 3 respectively. For both pre-reader categorical group comparison and longitudinal data no group differences were found for all participant characteristics ($p_s > .11$).

3.1.1. DR vs TR

In the pre-reader categorical group comparison, all cognitive measures significantly differed between the DR and TR group except for letter knowledge assessed in first grade (Table 2). Similar differences between TR and DR were found in the longitudinal sample. Again, all cognitive measures significantly differed between the DR and TR group except for letter knowledge (first grade) and Spoonerisms (Grade 2; Table 2). These analyses indicate the effectiveness of our procedure to identify dyslectic readers and typical readers.

Table 2
Participant characteristics comparing children with dyslexia (DR) and typical reading (TR) children.

DR vs. TR	Pre-reader categorical group comparison			Longitudinal categorical group comparison		
	N = 52			N = 35		
	DR (n = 16)	TR (n = 36)	Test statistics	DR (n = 12)	TR (n = 23)	Test statistics
Personal characteristics						
Sex (male/female)	12/4	22/14	$\chi 2_{(1)} = .944$, $p = .331$	9/3	16/7	$\chi 2_{(1)} = .114$, $p = .735$
SES	5.31 (1.70)	5.47	$\chi 2_{(5)} = 4.240$, $p = .515$	5.42 (1.73)	5.22 (1.24)	$\chi 2_{(5)} = 5.854$, $p = .321$
Non-verbal intelligence	65.00 (27.44)	69.89 (20.58)	W = 307.5, $p = .706$	68.42 (26.47)	71.30 (20.89)	W = 139.5, $p = .972$
Age at pre-reading MRI (months)	73.33 (3.02)	74.19 (2.95)	$T_{(44)} = .921$, $p = .362$[1]	73.91 (2.84)	74.23 (2.78)	$T_{(31)} = .308$, $p = .760$[7]
Age at early reading MRI (months)	N/A	N/A	N/A	95.36 (2.69)	96.09 (2.51)	$T_{(31)} = .767$, $p = .449$[7]
Age at fluent reading MRI (months)	N/A	N/A	N/A	127.00 (3.033)	127.80 (2.749)	$T_{(27)} = .762$, $p = .453$[8]
Pre-reading cognitive skills						
Phonological awareness (CS)	-.42 (.78)	.20 (.73)	$t_{(50)} = 2.767$, $p = .008$	-.50 (.87)	.15 (.73)	$T_{(33)} = 2.346$, $p = .025$
Grade 1 cognitive skills						
Letter knowledge (CS)	-.43 (1.15)	.18 (.86)	W = 376, $p = .083$	-.30 (1.00)	.26 (.83)	W = 175.5, $p = .198$
Grade 2 cognitive skills						
Phoneme deletion	10.6 (7.5)	16.6 (5.4)	$T_{(49)} = 3.196$, $p = .002$[2]	11.0 (7.1)	16.0 (5.5)	$T_{(33)} = 2.277$, $p = .029$
Spoonerism	16.9 (8.0)	26.6 (10.8)	$T_{(49)} = 3.123$, $p = .003$[2]	19.1 (7.0)	25.9 (12.0)	W = 174.0, $p = .217$
Word reading	11.7 (5.0)	29.3 (13.1)	W = 504.5, $p < .001$[2]	12.7 (5.2)	28.5 (12.0)	W = 259.5, $p < .001$
Pseudo-word reading	8.7 (2.9)	22.3 (9.7)	W = 514.0, $p < .001$[2]	9.3 (3.1)	21.6 (9.6)	W = 260.0, $p < .001$
Spelling	45.3 (9.6)	53.6 (6.1)	W = 451.5, $p < .001$[2]	47.5 (7.9)	53.3 (6.5)	W = 221.0, $p = .004$
Grade 3 cognitive skills						
Phoneme deletion	14.2 (6.3)	20.4 (4.9)	W = 459.0, $p < .001$	16.1 (4.6)	20.3 (4.9)	W = 211.5, $p = .011$
Spoonerism	24.3 (10.2)	35.9 (8.9)	W = 470.5, $p < .001$	27.6 (7.8)	35.3 (9.0)	W = 208.5, $p = .015$
Word reading	20.8 (8.1)	47.0 (11.5)	$T_{(50)} = 8.224$, $p < .001$	24.3 (5.9)	47.2 (9.0)	$T_{(33)} = 7.983$, $p < .001$
Pseudo-word reading	15.7 (6.7)	34.9 (10.5)	$T_{(50)} = 6.713$, $p < .001$	17.7 (6.2)	34.7 (8.3)	$T_{(33)} = 6.225$, $p < .001$
Spelling	34.3 (8.9)	48.0 (7.9)	$T_{(50)} = 5.563$, $p < .001$	36.8 (7.5)	49.2 (6.0)	$T_{(33)} = 5.350$, $p < .001$
Grade 4 cognitive skills						
Phoneme deletion	14.4 (6.6)	22.7 (3.7)	W = 437.0, $p < .001$[3]	16.2 (6.1)	22.6 (3.4)	W = 227.5, $p = .002$
Spoonerism	30.7 (9.7)	40.7 (8.5)	W = 414.0, $p < .001$[3]	34.5 (5.8)	40.7 (8.3)	W = 213.0, $p = .010$
Word reading	30.7 (8.4)	56.3 (12.3)	$T_{(47)} = 7.314$, $p < .001$[3]	33.5 (6.3)	54.3 (10.3)	$T_{(33)} = 6.391$, $p < .001$
Pseudo-word reading	19.0 (7.3)	45.6 (14.1)	W = 484.0, $p < .001$[3]	21.3 (6.2)	44.2 (12.8)	W = 258.0, $p < .001$
Spelling	26.9 (11.7)	44.9 (8.9)	$T_{(46)} = 5.900$, $p < .001$[4]	30.8 (7.9)	44.9 (9.0)	$T_{(33)} = 4.545$, $p < .001$
Grade 5 cognitive skills						
Phoneme deletion	20.0 (5.6)	25.0 (2.4)	W = 424.0, $p < .001$[5]	22.0 (3.0)	24.8 (2.6)	$T_{(33)} = 2.785$, $p = .009$
Spoonerism	37.7 (7.0)	45.3 (4.0)	W = 430.0, $p < .001$[5]	39.9 (5.6)	45.1 (4.5)	W = 220.0, $p = .005$
Word reading	43.5 (9.8)	70.9 (11.9)	$T_{(47)} = 7.848$, $p < .001$[5]	47.3 (6.6)	68.7 (10.1)	$T_{(33)} = 6.617$, $p < .001$
Pseudo-word reading	31.7 (12.6)	64.6 (11.7)	$T_{(41)} = 8.114$, $p < .001$[6]	33.6 (12.9)	61.9 (9.9)	$T_{(29)} = 6.768$, $p < .001$[9]
Spelling	27.7 (12.5)	47.4 (8.7)	W = 465.5, $p < .001$[5]	32.4 (8.6)	47.9 (7.5)	$T_{(33)} = 5.504$, $p < .001$

This table contains the means (and standard deviations) per subgroup. DR = Child developing dyslexia, TR = Child developing typical reading skills.
Missing data pre-reader categorical comparison: [1] N = 46, [2] N = 51, [3] N = 49, [4] N = 48, [5] N = 49 and [6] N = 43. Missing data longitudinal: [7] N = 33, [8] N = 29 and [9] N = 31.
Significant effects are marked in bold. Significance levels of $p > .05$ are seen as non-significant. All effects remain significant after False Discovery Rate Correction.

3.1.2. Familial risk for Dyslexia

Children with FRD$^+$ and FRD$^-$ showed no difference in composite score for phonological awareness in the pre-reader categorical group comparison. In first grade FRD$^-$ children scored significantly higher than FRD$^+$ children on letter knowledge ($T_{(50)} = 2.282$, $p = .027$). In the second grade, children with FRD$^-$ scored significantly better on word reading (W = 462.5, $p = .010$) compared to children with FRD$^+$. Cognitive measures for phonological awareness, (pseudo-)word reading and spelling showed significant higher scores for children with FRD$^-$ in grade three ($p_s < .010$) and grade four ($p_s < .042$) compared to FRD$^+$ children. In fifth grade (advanced reading stage) FRD$^-$ children scored higher on all measures except for phoneme deletion. A summary of these results can be found in Table 3. The longitudinal sample seems not to show differences on cognitive measurements when comparing FRD$^+$ and FRD$^-$ children. The sudden loss of significance for this effect might be due to the smaller sample size and subsequent decreased power compared to the categorical group comparisons sample.

3.2. Pre-reader categorical group comparisons

3.2.1. DR vs TR

First, differences for FA in the left AF between all TR and DR in the pre-reading stage were examined. A cluster of nodes was revealed that significantly differed between the TR and DR group, including node 19 to 22, driven by lower FA in the DR group (Table 4; Fig. 1). When looking into the adjacent nodes ($p = .05$ criterion), three additional nodes showed significant higher FA for TR compared to DR. Anterior to the described nodes, FA for node 18 differed significant between TR and DR. Posterior to node 22, FA for node 23 and node 24 showed a significant difference. The mean absolute effect size was calculated for the detected cluster of nodes (node 18–24) resulting in $d = .80$, which is a large effect size (Cohen, 1988) and over 2.5 SD larger than the mean absolute effect size for all 100 nodes ($d = .30$). Additional analyses showed no significant group differences in MD in nodes 18 to 24 ($p_s > .05$; Table 4).

Table 3
Participant characteristics comparing children with familial risk (FRD$^+$) and children without familial risk (FRD$^-$).

FRD$^+$ vs. FRD$^-$	Pre-reader categorical group comparison			Longitudinal categorical group comparison		
	N = 52			N = 35		
	FRD$^+$ (n = 27)	FRD$^-$ (n = 25)	Test statistics	FRD$^+$ (n = 20)	FRD$^-$ (n = 15)	Test statistics
Personal characteristics						
Sex (male/female)	19/8	15/10	$\chi2_{(1)} = .617$, $p = .432$	15/5	10/5	$\chi2_{(1)} = .292$, $p = .589$
SES	5.52 (1.45)	5.32 (1.60)	$\chi2_{(5)} = 3.147$, $p = .677$	5.55 (1.40)	4.93 (1.39)	$\chi2_{(5)} = 4.919$, $p = .426$
Non-verbal intelligence	64.74 (25.13)	72.32 (19.59)	$T_{(50)} = 1.206$, $p = .233$	65.00 (24.94)	77.40 (17.39)	$T_{(33)} = 1.646$, $p = .109$
Age at pre-reading MRI (months)	73.68 (3.09)	74.19 (2.86)	$T_{(44)} = .577$, $p = .567$[1]	74.11 (2.74)	74.13 (2.88)	$T_{(31)} = .023$, $p = .982$[7]
Age at early reading MRI (months)	/	/	/	95.94 (2.60)	95.73 (2.58)	$T_{(31)} = -.233$, $p = .817$[7]
Age at fluent reading MRI (months)	/	/	/	128.10 (3.13)	126.80 (2.26)	$T_{(27)} = -1.235$, $p = .227$[8]
Pre-reading cognitive skills						
Phonological awareness (CS)	-.097 (.900)	.117 (.653)	$T_{(50)} = .977$, $p = .333$	-.061 (.998)	-.081 (.551)	$T_{(33)} = -.068$, $p = .947$
Grade 1 cognitive skills						
Letter knowledge (CS)	-.299 (1.087)	.301 (.768)	$T_{(50)} = 2.282$, $p = .\mathbf{027}$	-.115 (.999)	.315 (.758)	$W = 186.5$, $p = .229$
Grade 2 cognitive skills						
Phoneme deletion	13.8 (6.7)	15.9 (6.4)	$T_{(49)} = 1.106$, $p = .274$[2]	14.6 (6.6)	13.8 (6.4)	$T_{(33)} = -.345$, $p = .732$
Spoonerism	21.4 (12.0)	26.3 (9.3)	$T_{(49)} = 1.634$, $p = .109$[2]	23.8 (11.9)	23.3 (10.0)	$T_{(33)} = -.110$, $p = .913$
Word reading	19.8 (13.0)	28.6 (13.6)	$W = 462.5$, $p = .\mathbf{010}$[2]	22.3 (13.7)	24.1 (11.6)	$W = 170.0$, $p = .515$
Pseudo-word reading	16.1 (10.6)	20.5 (9.8)	$W = 426.5$, $p = .057$[2]	17.8 (11.4)	16.8 (7.8)	$W = 154.0$, $p = .907$
Spelling	48.3 (10.1)	54.1 (3.8)	$W = 436.5$, $p = .\mathbf{036}$[2] *	45.0 (9.1)	53.1 (3.9)	$W = 169.5$, $p = .524$
Grade 3 cognitive skills						
Phoneme deletion	16.3 (6.6)	20.8 (4.4)	$W = 479.0$, $p = .\mathbf{010}$	18.1 (5.5)	19.8 (4.6)	$W = 259.5$, $p = .395$
Spoonerism	28.2 (11.6)	36.8 (7.7)	$W = 481.0$, $p = .\mathbf{009}$	31.0 (9.8)	34.9 (8.4)	$T_{(33)} = 1.216$, $p = .232$
Word reading	32.5 (16.2)	45.8 (13.0)	$T_{(50)} = 3.254$, $p = .\mathbf{002}$	37.4 (15.1)	42.0 (11.3)	$T_{(33)} = .999$, $p = .325$
Pseudo-word reading	24.1 (12.8)	34.4 (11.2)	$T_{(50)} = 3.069$, $p = .\mathbf{003}$	27.8 (12.5)	30.3 (9.3)	$T_{(33)} = .657$, $p = .515$
Spelling	39.4 (11.4)	48.5 (6.6)	$W = 506.5$, $p = .\mathbf{002}$	42.9 (10.2)	47.6 (5.7)	$W = 193.5$, $p = .150$
Grade 4 cognitive skills						
Phoneme deletion	17.7 (7.0)	22.6 (3.6)	$W = 426.0$, $p = .\mathbf{012}$[3]	19.5 (6.2)	21.6 (3.8)	$W = 176.0$, $p = .394$
Spoonerism	34.6 (10.9)	40.8 (7.8)	$W = 402.0$, $p = .\mathbf{042}$[3]	37.3 (8.9)	40.3 (6.6)	$W = 172.5$, $p = .463$
Word reading	43.2 (17.0)	54.0 (14.0)	$T_{(47)} = 2.425$, $p = .\mathbf{019}$[3]	46.8 (15.7)	47.7 (10.3)	$T_{(33)} = .186$, $p = .854$
Pseudo-word reading	31.6 (15.9)	43.6 (15.9)	$T_{(47)} = 2.539$, $p = .\mathbf{014}$[3]	35.4 (16.7)	37.6 (14.3)	$T_{(33)} = .410$, $p = .684$
Spelling	34.0 (14.2)	45.0 (8.3)	$W = 412.0$, $p = .\mathbf{010}$[4]	37.4 (12.0)	43.6 (8.4)	$T_{(33)} = 1.713$, $p = .096$
Grade 5 cognitive skills						
Phoneme deletion	22.1 (5.4)	24.8 (2.3)	$W = 387.5$, $p = .081$[5]	23.6 (3.3)	24.1 (2.6)	$W = 157.5$, $p = .815$
Spoonerism	40.6 (6.9)	45.3 (4.4)	$W = 441.0$, $p = .\mathbf{005}$[5]	42.2 (5.5)	44.9 (5.2)	$W = 205.0$, $p = .069$
Word reading	57.1 (18.2)	67.7 (14.2)	$T_{(47)} = 2.275$, $p = .\mathbf{027}$[5]	61.4 (15.9)	61.3 (10.7)	$T_{(33)} = -.004$, $p = .997$
Pseudo-word reading	48.9 (20.0)	61.6 (16.1)	$T_{(41)} = 2.287$, $p = .\mathbf{027}$[6]	51.8 (19.4)	54.2 (14.2)	$T_{(29)} = .374$, $p = .711$[9]
Spelling	36.0 (15.8)	46.5 (8.3)	$W = 413.0$, $p = .\mathbf{024}$[5]	40.8 (12.2)	45.0 (8.3)	$W = 175.0$, $p = .413$

This table contains the means (and standard deviations) per subgroup. FRD$^+$ = Child with familial risk for dyslexia, FRD$^-$ = Child without familial risk for dyslexia.
Missing data pre-reader categorical comparison: [1] N = 46, [2] N = 51, [3] N = 49, [4] N = 48, [5] N = 49 and [6] N = 43. Missing data longitudinal: [7] N = 33, [8] N = 29 and [9] N = 31.
Significant effects are marked in bold. Significance levels of $p > .05$ are seen as non-significant. * Significant effects not surviving False Discovery Rate correction.

We also examined differences between DR and TR in the right AF. It was shown that none of the 100 nodes reached significance at the $p = 0.01$ level (see Appendix A). Therefore, and due to the limited number of children for whom a right AF was traced, this comparison was not further examined.

3.2.2. FRD$^-$ TR vs DR

We investigated whether the observed differences in FA between TR and DR in the left AF were driven by reading ability, regardless of the large amount of FRD$^+$ children. To control for this effect of familial risk, tests were conducted comparing FRD$^-$ TR (n = 21) and all DR (n = 16). The FRD$^-$ TR group forms a better representation of the general population. Significant differences in FA between these two groups were found in the same nodes in the left long AF, see Table 4. Again, three adjacent nodes reached significance for the $p = .05$ criterion: FA in node 19, node 23 and node 24 showed a significant higher estimate for FRD$^-$ TR compared to DR. These results indicate that a difference in FA at these particular nodes between TR and DR is not driven by the large amount of FRD$^+$ children in our study. Additional comparisons on MD showed no significant group differences were found ($p_s > .05$; Table 4).

3.2.3. Familial risk for dyslexia

No differences for FA were found between FRD$^+$ and FRD$^-$ children in the left AF at the pre-reading stage (see Appendix B). Additional analyzes in the group of TR children (FRD$^+$TR versus FRD$^-$TR) confirm these results ($p_s > .05$). Analysis in the right AF also found no significant cluster of nodes differing in AF between FRD$^+$ and FRD$^-$. We were unable to restrict this comparison to the TR group, due to the limited number of children for whom a right AF was traced. These findings are in line with results on the cognitive measures for phonological awareness, (pseudo-)word reading and spelling, where equally no difference was found between family risk groups in the longitudinal sample (Table 3). Due to this conformity, the potential difference between these groups was no further examined in the longitudinal or continuous analyses.

3.3. Longitudinal analysis

3.3.1. DR vs TR

The categorical group comparisons in the pre-reading stage showed a cluster of nodes in which FA is significantly higher for TR compared to DR. Additional figures showing an along the tract comparison of FA

Table 4
Categorical group comparison of FA and MD in the left AF in the pre-reading stage.

		DR vs. TR			DR vs. FRD⁻ TR		
		N = 52			N = 37		
	Node on AF	DR (n = 16)	TR (n = 36)	Test statistics	DR (n = 16)	FRD⁻ TR (n = 21)	Test statistics
FA	Node 18	.359 (.066)	.400 (.058)	$T_{(50)} = 2.264$, $p = .028$	.359 (.066)	.402 (.062)	$T_{(35)} = 2.003$, $p = .053$
	Node 19	.356 (.063)	.401 (.054)	$T_{(50)} = 2.668$, $p = .010$	.356 (.063)	.404 (.060)	$T_{(35)} = 2.402$, $p = .022$
	Node 20	.353 (.061)	.401 (.050)	$T_{(50)} = 2.997$, $p = .004$	.353 (.061)	.407 (.055)	$T_{(35)} = 2.810$, $p = .008$
	Node 21	.352 (.058)	.400 (.047)	$T_{(50)} = 3.122$, $p = .003$	.352 (.058)	.407 (.051)	$T_{(35)} = 2.049$, $p = .004$
	Node 22	.354 (.057)	.398 (.047)	$T_{(50)} = 2.915$, $p = .005$	.354 (.057)	.406 (.050)	$T_{(35)} = 2.934$, $p = .006$
	Node 23	.359 (.057)	.397 (.049)	$T_{(50)} = 2.479$, $p = .017$	.359 (.057)	.404 (.053)	$T_{(35)} = 2.511$, $p = .017$
	Node 24	.364 (.057)	.398 (.052)	$T_{(50)} = 2.128$, $p = .038$	.364 (.057)	.404 (.057)	$T_{(35)} = 2.131$, $p = .040$
MD	Node 18	.718 (.049)	.705 (.044)	$T_{(50)} = -0.906$, $p = .369$	.718 (.049)	.707 (.033)	$T_{(35)} = -0.779$, $p = .441$
	Node 19	.715 (.049)	.702 (.043)	$T_{(50)} = -0.923$, $p = .361$	.715 (.049)	.705 (.033)	$T_{(35)} = -0.761$, $p = .452$
	Node 20	.712 (.048)	.700 (.043)	$T_{(50)} = -0.903$, $p = .371$	.712 (.048)	.702 (.034)	$T_{(35)} = -0.715$, $p = .479$
	Node 21	.709 (.047)	.698 (.044)	$T_{(50)} = -0.840$, $p = .405$	.709 (.047)	.700 (.038)	$T_{(35)} = -0.637$, $p = .529$
	Node 22	.706 (.046)	.696 (.045)	$T_{(50)} = -0.763$, $p = .449$	.706 (.046)	.698 (.043)	$T_{(35)} = -0.570$, $p = .572$
	Node 23	.703 (.044)	.693 (.047)	$T_{(50)} = -0.708$, $p = .482$	.703 (.044)	.695 (.047)	$T_{(35)} = -0.552$, $p = .584$
	Node 24	.700 (.042)	.691 (.048)	$T_{(50)} = -0.657$, $p = .514$	.700 (.042)	.692 (.050)	$T_{(35)} = -0.547$, $p = .588$

This table contains the means (and standard deviations) per subgroup. FA = Fractional anisotropy, MD = Mean diffusivity, DR = Child developing dyslexia, TR = Child developing typical reading skills, FRD⁻ TR = Child developing typical reading skills without familial risk for dyslexia. Significant effects at $p \leq .01$ are underlined and marked in bold, significant effects at $p \leq .05$ in adjacent nodes are underlined.

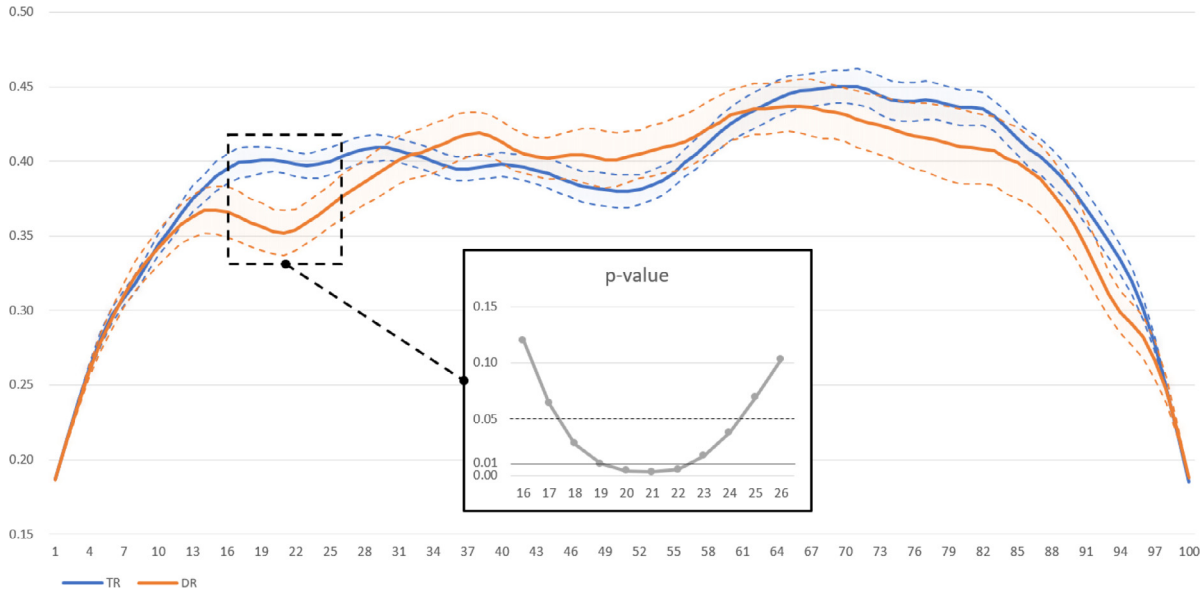

Fig. 1. Mean FA and SE (dotted line) for the DR and TR groups at all 100 nodes of the left AF at the pre-reading stage (N = 52).

between DR and TR at the early and advanced reading stage show similar results (see Appendix C and D). We examined the maturation of the cluster of nodes that was found at the pre-reading stage, using a longitudinal analysis (N = 35), see Table 5. FA increased significantly throughout time in all discussed nodes ($p_s < .028$). Further results showed no group by time point interaction effect, suggesting FA maturation did not differ throughout time between TR and DR. Additional Bayesian analyses on this interaction effect, i.e. the calculation of BF inclusion (BF_{incl}), provided 'moderate evidence' (Andraszewicz et al., 2015) for the hypotheses that there is no difference in FA maturation (Table 5). When examining the group differences between TR and DR in this longitudinal sample, the selected cluster of nodes (node 19 to node 22) contained significant differences (Table 5), although it should be noted that none of these effects survived FDR correction. The surrounding nodes, selected based of the $p = 0.05$ criterion, were also examined in this longitudinal sample. No significant group differences between TR and DR children were retrieved in these nodes. Additional analyses showed significant decrease of MD over time ($p_s < .001$). No significant group by time point interaction effects were found (effect in node 21 did not survive FRD correction) as well as no significant group differences.

3.3.2. FRD⁻ TR vs DR

To approach a better representation of the general population, tests were conducted comparing FRD⁻ TR (n = 11) and all DR (n = 12). Similar results were found in this analysis (Table 6), as FA increased significantly throughout time in all discussed nodes ($p_s = .011$). Again, no interaction effect was found for group and time, strengthened by the Bayesian analyses that provides moderate evidence that there is no difference in FA maturation between both groups. Significant group differences between FRD⁻ TR and DR were found in five adjacent nodes: node 20 to 24 (Fig. 2). The FA-values of node 18 and node 19 were also compared between groups, revealing no significant group difference. By narrowing the group of TR children down to the FRD⁻ TR children, our results revealed evidence for a local difference in FA of the left AF between children developing dyslexia and the general population.

Additional analyses for the MD data revealed similar results. MD decreased significantly throughout time in all discussed nodes ($p_s = .016$)

Table 5
Longitudinal analysis by means of repeated measures ANOVA for FA and MD in left AF (DR vs. TR)

DR vs. TR		Descriptives (DR = 12, TR = 23)						Test statistics		
		Pre-reading stage		Early reading stage		Advanced reading stage		Within subject effect		Between subject effects
	Nodes	*DR*	*TR*	*DR*	*TR*	*DR*	*TR*	*Effect of time*	*Interaction effect*	*Effect of group*
FA	**Node 18**	.373 (.067)	.405 (.054)	.393 (.042)	.417 (.043)	.404 (.053)	.429 (.056)	$F_{(2,66)} = 4.556$, $p = .\mathbf{014}$	$F_{(2,66)} = .110$, $p = .896$, $BF_{incl} = 0.150$	$F_{(1,33)} = 2.986$, $p = .093$
	Node 19	.369 (.064)	.406 (.049)	.390 (.037)	.416 (.040)	.398 (.052)	.426 (.052)	$F_{(2,66)} = 3.929$, $p = .\mathbf{024}$	$F_{(2,66)} = .205$, $p = .815$, $BF_{incl} = 0.176$	$F_{(1,33)} = 4.588$, $p = .\mathbf{040}$*
	Node 20	.366 (.059)	.405 (.045)	.386 (.032)	.414 (.039)	.395 (.051)	.423 (.049)	$F_{(2,66)} = 3.771$, $p = .\mathbf{028}$	$F_{(2,66)} = .243$, $p = .785$, $BF_{incl} = 0.176$	$F_{(1,33)} = 6.174$, $p = .\mathbf{018}$*
	Node 21	.365 (.056)	.403 (.045)	.385 (.030)	.413 (.038)	.393 (.049)	.422 (.049)	$F_{(2,66)} = 4.046$, $p = .\mathbf{022}$	$F_{(2,66)} = .231$, $p = .794$, $BF_{incl} = 0.171$	$F_{(1,33)} = 6.438$, $p = .\mathbf{016}$*
	Node 22	.368 (.053)	.402 (.046)	.384 (.030)	.412 (.040)	.395 (.046)	.423 (.052)	$F_{(2,66)} = 4.273$, $p = .\mathbf{018}$	$F_{(2,66)} = .110$, $p = .896$, $BF_{incl} = 0.169$	$F_{(1,33)} = 5.202$, $p = .\mathbf{029}$*
	Node 23	.374 (.051)	.403 (.048)	.386 (.032)	.413 (.044)	.399 (.046)	.426 (.055)	$F_{(2,66)} = 4.324$, $p = .\mathbf{017}$	$F_{(2,66)} = .007$, $p = .993$, $BF_{incl} = 0.154$	$F_{(1,33)} = 3.801$, $p = .060$
	Node 24	.381 (.050)	.405 (.049)	.389 (.037)	.415 (.048)	.404 (.050)	.430 (.057)	$F_{(2,66)} = 4.447$, $p = .\mathbf{015}$	$F_{(2,66)} = .011$, $p = .989$, $BF_{incl} = 0.163$	$F_{(1,33)} = 2.876$, $p = .099$
MD	**Node 18**	.701 (.042)	.708 (.043)	.701 (.038)	.682 (.053)	.681 (.043)	.678 (.037)	$F_{(2,66)} = 10.103$, $p < .\mathbf{001}$	$F_{(2,66)} = 2.586$, $p = .083$, $BF_{incl} = 1.041$	$F_{(1,33)} = 0.122$, $p = .730$
	Node 19	.698 (.041)	.705 (.042)	.697 (.038)	.679 (.053)	.677 (.042)	.674 (.036)	$F_{(2,66)} = 11.818$, $p < .\mathbf{001}$	$F_{(2,66)} = 2.709$, $p = .074$, $BF_{incl} = 1.311$	$F_{(1,33)} = 0.106$, $p = .746$
	Node 20	.695 (.039)	.701 (.041)	.694 (.039)	.676 (.052)	.674 (.040)	.672 (.036)	$F_{(2,66)} = 12.591$, $p < .\mathbf{001}$	$F_{(2,66)} = 2.931$, $p = .060$, $BF_{incl} = 1.250$	$F_{(1,33)} = 0.102$, $p = .751$
	Node 21	.692 (.037)	.698 (.040)	.691 (.040)	.674 (.050)	.671 (.037)	.670 (.036)	$F_{(2,66)} = 12.438$, $p < .\mathbf{001}$	$F_{(2,66)} = 3.150$, $p = .\mathbf{049}$*, $BF_{incl} = 1.537$	$F_{(1,33)} = 0.091$, $p = .765$
	Node 22	.689 (.035)	.695 (.040)	.688 (.040)	.672 (.048)	.668 (.035)	.670 (.036)	$F_{(2,66)} = 11.403$, $p < .\mathbf{001}$	$F_{(2,66)} = 3.047$, $p = .054$, $BF_{incl} = 1.285$	$F_{(1,33)} = 0.056$, $p = .814$
	Node 23	.686 (.033)	.692 (.040)	.685 (.039)	.671 (.047)	.665 (.034)	.670 (.036)	$F_{(2,66)} = 10.158$, $p < .\mathbf{001}$	$F_{(2,66)} = 2.635$, $p = .079$, $BF_{incl} = 1.014$	$F_{(1,33)} = 0.011$, $p = .916$
	Node 24	.683 (.031)	.689 (.039)	.680 (.037)	.670 (.046)	.662 (.033)	.670 (.036)	$F_{(2,66)} = 8.726$, $p < .\mathbf{001}$	$F_{(2,66)} = 2.060$, $p = .136$, $BF_{incl} = 0.710$	$F_{(1,33)} = 0.005$, $p = .942$

FA = Fractional anisotropy, MD = Mean diffusivity, DR = Child developing dyslexia, TR = Child developing typical reading skills, BF_{incl} = BF inclusion for interaction effect. The mean (and standard deviations) for every group are presented under descriptives. Significant effects are marked in bold. Significance levels of p > .05 are seen as non-significant.

* Significant effects not surviving False Discovery Rate correction.

Table 6
Longitudinal analysis by means of repeated measures ANOVA for FA and MD in left AF (DR vs. FRD⁻ TR)

DR vs. FRD⁻ TR		Descriptives (*DR = 12,* FRD⁻ TR = 11)						Test statistics		
		Pre-reading stage		Early reading stage		Advanced reading stage		Within subject effect		Between subject effects
	Nodes	*DR*	*FRD⁻ TR*	*DR*	*FRD⁻ TR*	*DR*	*FRD⁻ TR*	*Effect of time*	*Interaction effect*	*Effect of group*
FA	**Node 18**	.373 (.067)	.400 (.057)	.393 (.042)	.415 (.045)	.404 (.053)	.433 (.037)	$F_{(2,42)} = 6.343$, $p = .\mathbf{004}$	$F_{(2,42)} = .112$, $p = .894$, $BF_{incl} = 0.197$	$F_{(1,21)} = 1.882$, $p = .185$
	Node 19	.369 (.064)	.403 (.056)	.390 (.037)	.416 (.045)	.398 (.052)	.434 (.035)	$F_{(2,42)} = 5.550$, $p = .\mathbf{007}$	$F_{(2,42)} = .150$, $p = .861$, $BF_{incl} = 0.203$	$F_{(1,21)} = 3.236$, $p = .086$
	Node 20	.366 (.059)	.408 (.053)	.386 (.032)	.418 (.045)	.395 (.051)	.436 (.034)	$F_{(2,42)} = 5.086$, $p = .\mathbf{011}$	$F_{(2,42)} = .210$, $p = .812$, $BF_{incl} = 0.226$	$F_{(1,21)} = 5.302$, $p = .\mathbf{032}$
	Node 21	.365 (.056)	.411 (.051)	.385 (.030)	.421 (.042)	.393 (.049)	.440 (.035)	$F_{(2,42)} = 5.390$, $p = .\mathbf{008}$	$F_{(2,42)} = .231$, $p = .794$, $BF_{incl} = 0.217$	$F_{(1,21)} = 7.620$, $p = .\mathbf{012}$
	Node 22	.368 (.053)	.413 (.052)	.384 (.030)	.422 (.042)	.395 (.046)	.446 (.037)	$F_{(2,42)} = 5.779$, $p = .\mathbf{006}$	$F_{(2,42)} = .292$, $p = .749$, $BF_{incl} = 0.229$	$F_{(1,21)} = 8.544$, $p = .\mathbf{008}$
	Node 23	.374 (.051)	.415 (.053)	.386 (.032)	.422 (.049)	.399 (.046)	.452 (.038)	$F_{(2,42)} = 5.982$, $p = .\mathbf{005}$	$F_{(2,42)} = .459$, $p = .635$, $BF_{incl} = 0.283$	$F_{(1,21)} = 7.455$, $p = .\mathbf{013}$
	Node 24	.381 (.050)	.417 (.055)	.389 (.037)	.423 (.056)	.404 (.050)	.457 (.039)	$F_{(2,42)} = 6.283$, $p = .\mathbf{004}$	$F_{(2,42)} = .607$, $p = .550$, $BF_{incl} = 0.291$	$F_{(1,21)} = 6.012$, $p = .\mathbf{023}$
MD	**Node 18**	.701 (.042)	.698 (.028)	.701 (.038)	.663 (.023)	.681 (.043)	.667 (.029)	$F_{(2,42)} = 10.766$, $p < .\mathbf{001}$	$F_{(2,42)} = 5.038$, $p = .\mathbf{011}$, $BF_{incl} = 4.758$	$F_{(1,21)} = 1.923$, $p = .180$
	Node 19	.698 (.041)	.695 (.027)	.697 (.038)	.658 (.022)	.677 (.042)	.664 (.028)	$F_{(2,42)} = 12.313$, $p < .\mathbf{001}$	$F_{(2,42)} = 5.780$, $p = .\mathbf{006}$, $BF_{incl} = 7.224$	$F_{(1,21)} = 2.051$, $p = .167$
	Node 20	.695 (.039)	.692 (.026)	.694 (.039)	.655 (.021)	.674 (.040)	.661 (.027)	$F_{(2,42)} = 12.354$, $p < .\mathbf{001}$	$F_{(2,42)} = 6.098$, $p = .\mathbf{005}$, $BF_{incl} = 8.190$	$F_{(1,21)} = 2.139$, $p = .158$
	Node 21	.692 (.037)	.688 (.026)	.691 (.040)	.653 (.021)	.671 (.037)	.661 (.027)	$F_{(2,42)} = 10.648$, $p < .\mathbf{001}$	$F_{(2,42)} = 5.879$, $p = .\mathbf{006}$, $BF_{incl} = 7.773$	$F_{(1,21)} = 2.069$, $p = .165$
	Node 22	.689 (.035)	.684 (.027)	.688 (.040)	.652 (.022)	.668 (.035)	.662 (.027)	$F_{(2,42)} = 8.082$, $p = .\mathbf{001}$	$F_{(2,42)} = 5.065$, $p = .\mathbf{011}$, $BF_{incl} = 4.873$	$F_{(1,21)} = 1.776$, $p = .197$
	Node 23	.686 (.033)	.681 (.028)	.685 (.039)	.653 (.023)	.665 (.034)	.664 (.028)	$F_{(2,42)} = 6.002$, $p = .\mathbf{005}$	$F_{(2,42)} = 4.195$, $p = .\mathbf{022}$, $BF_{incl} = 2.915$	$F_{(1,21)} = 1.319$, $p = .264$
	Node 24	.683 (.031)	.677 (.028)	.680 (.037)	.652 (.024)	.662 (.033)	.665 (.028)	$F_{(2,42)} = 4.564$, $p = .\mathbf{016}$	$F_{(2,42)} = 3.628$, $p = .\mathbf{035}$, $BF_{incl} = 2.032$	$F_{(1,21)} = 0.888$, $p = .357$

FA = Fractional anisotropy, MD = Mean diffusivity, DR = Child developing dyslexia, FRD⁻ TR = Child developing typical reading skills without familial risk for dyslexia. BF_{incl} = BF inclusion for interaction effect. The mean (and standard deviations) for every group are presented under descriptives. Significant effects are marked in bold. Significance levels of p > .05 are seen as non-significant. All effects remain significant after False Discovery Rate Correction.

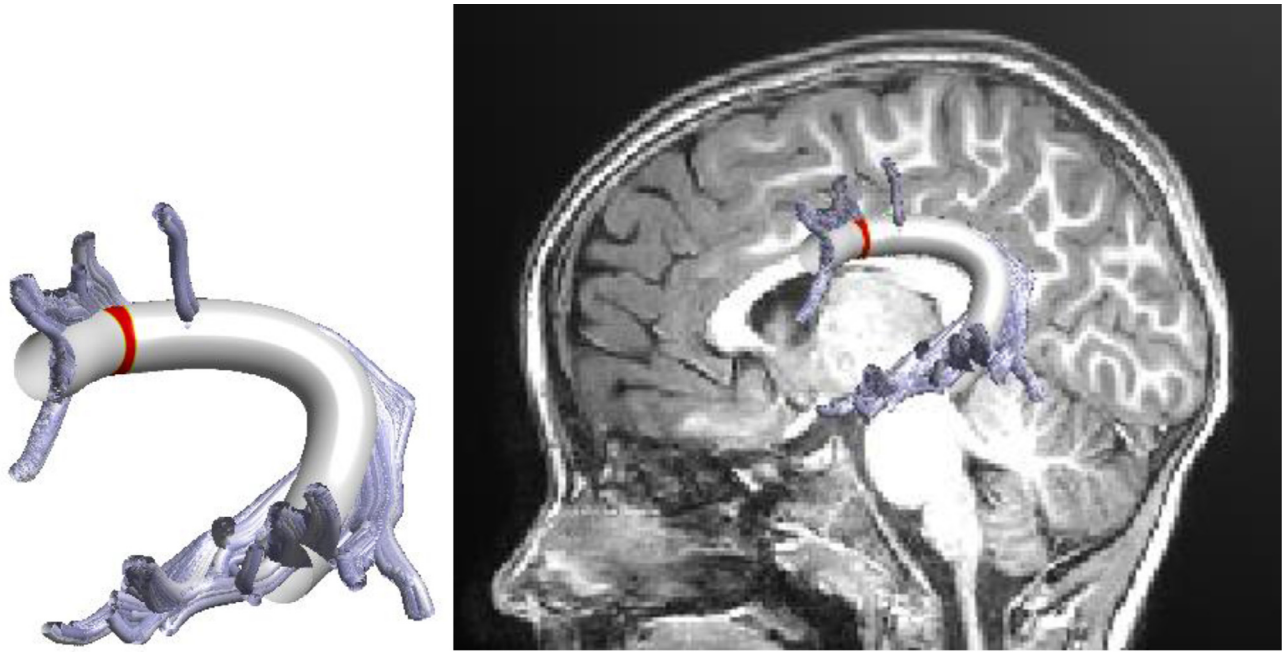

Fig. 2. Node 20 to 24 (marked red) on the left AF at the pre-reading stage

in the absence of a significant group difference. A significant group by time point interaction effect was found for all nodes (p_s = .035), strengthened by the Bayesian analyses: Group differences for MD were largest at the early reading stage, driven by a higher MD in the DR group. However, post hoc analyses showed no significant differences at the early reading stage (p_s = .119), in line with the missing main effect for group differences.

3.4. Continuous approach

Previous categorical group comparisons and longitudinal sample analyses provided evidence for a difference in white matter structure between FRD⁻TR and DR in a specific cluster. We argue that children were assigned to the group of TR or DR based on a rather stringent threshold. To assess the possibility of a more continuous approach, the predictive and concurrent relation between FA for the previously mentioned cluster (FA for nodes 18 to 24 of the left long AF from all three MRI time point) and all reading measurements was examined. Analogous with previous analyses, various nodes of this cluster prove to be significantly correlated to scores on word reading tests administered in second to fifth grade, see Table 7. Please note that for these continuous analyses, restricting the sample to the individuals with typical reading skills, resulted in similar effect sizes (average difference in effect size of .051). Due to the absence of group differences for the MD data in previous analyses, MD was not examined in this continuous approach.

3.4.1. Pre-reading stage

In first instance we focus on nodes 19 to 22 because they appear to contain the most variance between TR and DR. FA in node 19 at the pre-reading stage appears to be a significant predictor for scores on word reading tests conducted in third ($F_{(2,49)}$ = 4.288, p = .007) and fourth ($F_{(2,45)}$ = 2.224, p = .041) grade, in second grade the effect was non-significant ($F_{(2,48)}$ = 1.225, p = .125) and in fifth grade the effect did not survive FDR correction ($F_{(2,46)}$ = 2.175, p = .048). Additional results were found for FA in node 20. Linear regression with FA at node 20 in the pre-reading stage as the variable to predict scores for word reading in third, fourth and fifth grade turns out to be significant (Table 7), in second grade significance was not reached ($F_{(2,48)}$ = 1.972, p = .053). FA in node 21 (p_s < .008) and node 22 (p_s < .007) in from the pre-reading stage appears as significant coefficients for word reading in third to fifth grade, in second grade the effect did not survive FDR correction.

Additionally, we investigated the other nodes in this previously determined cluster. FA in node 18 from the pre-reading stage only had a significant predictive value for word reading in third grade ($F_{(2,49)}$ = 3.008, p = .025). FA in node 23 (p_s < .011) and node 24 (p_s < .018) from the pre-reading stage both were significant coefficients for word reading in third to fifth grade. These results provide evidence for the hypothesis that white matter FA of the left long AF from the pre-reading stage has a predictive value for further reading development.

3.4.2. Early and advanced reading stage

Inspection of the linear regression conducted on data of the MRI scan in the early reading stage provides additional evidence for the previous findings. FA values for node 19, node 20, node 21 and node 22 are significant coefficients for word reading measured in the second and third grade (Table 7). In the fifth grade, FA for node 20 ($F_{(2,39)}$ = 2.816, p = .031) and node 21 ($F_{(2,39)}$ = 2.637, p = .038) appear as significant coefficients, however they did not survive FDR correction. The strong dependency of FA in the left long AF and the score on the word reading task seem to fade. Table 7 shows no significant effect for FA as a coefficient in any of the nodes of the previously described cluster from the MRI scans conducted in the advanced reading stage.

4. Discussion

Although in the past decade important advancements have been made with respect to understanding the neural basis of dyslexia, apart from the consistent deficit in the left AF found in pre-readers, results remain unclear. Some studies point towards the presence of altered white matter organization in the left AF of school-aged children, while other studies did not report such evidence. Moreover, the comparison of findings across studies is often hampered by methodological differences between studies and the lack of longitudinal studies with 3 data points. Although past research has applied a multiple MRI-time point approach to investigate the longitudinal development of core reading tracts in typically developing children (Reynolds et al., 2019), children at risk for dyslexia and children developing below average read-

Table 7
Linear regression

FA	Word reading test			
	Second grade F-value *p-value* R^2	Third grade F-value *p-value* R^2	Fourth grade F-value *p-value* R^2	Fifth grade F-value *p-value* R^2
Pre-reading stage (N = 52)	*(N = 51)*	*(N = 52)*	*(N = 49)*	*(N = 49)*
Node18	$F_{(2,48)} = .621$, $p = .272$, $R^2 = .025$	$F_{(2,49)} = 3.008$, $p = .\underline{\mathbf{025}}$, $R^2 = .109$	$F_{(2,46)} = 1.313$, $p = .114$, $R^2 = .054$	$F_{(2,46)} = 1.338$, $p = .125$, $R^2 = .055$
Motion	$p = .856$	$p = .523$	$p = .897$	$p = .687$
Node19	$F_{(2,48)} = 1.225$, $p = .125$, $R^2 = .049$	$F_{(2,49)} = 4.288$, $p = .\underline{\mathbf{007}}$, $R^2 = .149$	$F_{(2,46)} = 2.224$, $p = .\underline{\mathbf{041}}$, $R^2 = .088$	$F_{(2,46)} = 2.175$, $p = .\underline{\mathbf{048}}^*$, $R^2 = .086$
Motion	$p = .828$	$p = .532$	$p = .916$	$p = .686$
Node20	$F_{(2,48)} = 1.972$, $p = .053$, $R^2 = .076$	$F_{(2,49)} = 5.676$, $p = .\underline{\mathbf{002}}$, $R^2 = .188$	$F_{(2,46)} = 3.243$, $p = .\underline{\mathbf{014}}$, $R^2 = .124$	$F_{(2,46)} = 3.163$, $p = .\underline{\mathbf{017}}$, $R^2 = .121$
Motion	$p = .789$	$p = .556$	$p = .920$	$p = .692$
Node21	$F_{(2,48)} = 2.482$, $p = .\underline{\mathbf{031}}^*$, $R^2 = .094$	$F_{(2,49)} = 6.540$, $p < .\underline{\mathbf{001}}$, $R^2 = .211$	$F_{(2,46)} = 3.912$, $p = .\underline{\mathbf{008}}$, $R^2 = .145$	$F_{(2,46)} = 3.967$, $p = .\underline{\mathbf{008}}$, $R^2 = .147$
Motion	$p = .763$	$p = .570$	$p = .918$	$p = .693$
Node22	$F_{(2,48)} = 2.560$, $p = .\underline{\mathbf{028}}^*$, $R^2 = .096$	$F_{(2,49)} = 6.379$, $p = .\underline{\mathbf{001}}$, $R^2 = .207$	$F_{(2,46)} = 3.935$, $p = .\underline{\mathbf{007}}$, $R^2 = .146$	$F_{(2,46)} = 4.334$, $p = .\underline{\mathbf{006}}$, $R^2 = .159$
Motion	$p = .741$	$p = .597$	$p = .883$	$p = .712$
Node23	$F_{(2,48)} = 2.441$, $p = .\underline{\mathbf{032}}^*$, $R^2 = .092$	$F_{(2,49)} = 5.624$, $p = .\underline{\mathbf{002}}$, $R^2 = .187$	$F_{(2,46)} = 3.495$, $p = .\underline{\mathbf{011}}$, $R^2 = .132$	$F_{(2,46)} = 4.170$, $p = .\underline{\mathbf{007}}$, $R^2 = .153$
Motion	$p = .739$	$p = .559$	$p = .856$	$p = .727$
Node24	$F_{(2,48)} = 2.263$, $p = .\underline{\mathbf{039}}^*$, $R^2 = .086$	$F_{(2,49)} = 4.798$, $p = .\underline{\mathbf{005}}$, $R^2 = .164$	$F_{(2,46)} = 3.046$, $p = .\underline{\mathbf{018}}$, $R^2 = .117$	$F_{(2,46)} = 3.880$, $p = .\underline{\mathbf{009}}$, $R^2 = .144$
Motion	$p = .741$	$p = .597$	$p = .826$	$p = .745$
Early reading stage (N = 45)	*(N = 44)*	*(N = 45)*	*(N = 43)*	*(N = 42)*
Node18	$F_{(2,41)} = 2.385$, $p = .067$, $R^2 = .104$	$F_{(2,42)} = 2.041$, $p = .\underline{\mathbf{050}}$, $R^2 = .089$	$F_{(2,40)} = 1.991$, $p = .125$, $R^2 = .091$	$F_{(2,39)} = 1.672$, $p = .107$, $R^2 = .079$
Motion	$p = .247$	$p = .853$	$p = .227$	$p = .466$
Node19	$F_{(2,41)} = 3.484$, $p = .\underline{\mathbf{022}}$, $R^2 = .145$	$F_{(2,42)} = 2.841$, $p = .\underline{\mathbf{022}}$, $R^2 = .119$	$F_{(2,40)} = 2.502$, $p = .071$, $R^2 = .111$	$F_{(2,39)} = 2.305$, $p = .053$, $R^2 = .106$
Motion	$p = .250$	$p = .897$	$p = .247$	$p = .520$
Node20	$F_{(2,41)} = 4.761$, $p = .\underline{\mathbf{007}}$, $R^2 = .188$	$F_{(2,42)} = 3.510$, $p = .\underline{\mathbf{011}}$, $R^2 = .143$	$F_{(2,40)} = 2.819$, $p = .051$, $R^2 = .124$	$F_{(2,39)} = 2.816$, $p = .\underline{\mathbf{031}}^*$, $R^2 = .126$
Motion	$p = .247$	$p = .940$	$p = .262$	$p = .566$
Node21	$F_{(2,41)} = 5.235$, $p = .\underline{\mathbf{004}}$, $R^2 = .203$	$F_{(2,42)} = 3.443$, $p = .\underline{\mathbf{012}}$, $R^2 = .141$	$F_{(2,40)} = 2.540$, $p = .068$, $R^2 = .113$	$F_{(2,39)} = 2.637$, $p = .\underline{\mathbf{038}}^*$, $R^2 = .119$
Motion	$p = .230$	$p = .937$	$p = .259$	$p = .569$
Node22	$F_{(2,41)} = 4.694$, $p = .\underline{\mathbf{007}}$, $R^2 = .186$	$F_{(2,42)} = 3.028$, $p = .\underline{\mathbf{018}}$, $R^2 = .126$	$F_{(2,40)} = 2.212$, $p = .098$, $R^2 = .100$	$F_{(2,39)} = 2.227$, $p = .058$, $R^2 = .102$
Motion	$p = .204$	$p = .881$	$p = .240$	$p = .526$
Node23	$F_{(2,41)} = 3.673$, $p = .\underline{\mathbf{018}}$, $R^2 = .152$	$F_{(2,42)} = 2.687$, $p = .\underline{\mathbf{026}}$, $R^2 = .113$	$F_{(2,40)} = 2.274$, $p = .091$, $R^2 = .102$	$F_{(2,39)} = 2.176$, $p = .061$, $R^2 = .100$
Motion	$p = .193$	$p = .812$	$p = .213$	$p = .467$
Node24	$F_{(2,41)} = 2.949$, $p = .\underline{\mathbf{037}}$, $R^2 = .126$	$F_{(2,42)} = 2.456$, $p = .\underline{\mathbf{032}}$, $R^2 = .105$	$F_{(2,40)} = 2.425$, $p = .077$, $R^2 = .108$	$F_{(2,39)} = 2.159$, $p = .062$, $R^2 = .100$
Motion	$p = .194$	$p = .767$	$p = .190$	$p = .417$
Advanced reading stage (N = 35)	*(N = 35)*	*(N = 35)*	*(N = 35)*	*(N = 35)*
Node18	$F_{(2,32)} = 1.003$, $p = .679$, $R^2 = .059$	$F_{(2,32)} = 1.303$, $p = .900$, $R^2 = .075$	$F_{(2,32)} = 1.178$, $p = .912$, $R^2 = .069$	$F_{(2,32)} = 1.340$, $p = .967$, $R^2 = .077$
Motion	$p = .167$	$p = .150$	$p = .169$	$p = .136$
Node19	$F_{(2,32)} = .944$, $p = .804$, $R^2 = .056$	$F_{(2,32)} = 1.338$, $p = .778$, $R^2 = .077$	$F_{(2,32)} = 1.212$, $p = .784$, $R^2 = .070$	$F_{(2,32)} = 1.367$, $p = .823$, $R^2 = .079$
Motion	$p = .187$	$p = .173$	$p = .195$	$p = .159$
Node20	$F_{(2,32)} = .920$, $p = .898$, $R^2 = .054$	$F_{(2,32)} = 1.374$, $p = .704$, $R^2 = .079$	$F_{(2,32)} = 1.263$, $p = .682$, $R^2 = .073$	$F_{(2,32)} = 1.404$, $p = .732$, $R^2 = .081$
Motion	$p = .204$	$p = .188$	$p = .217$	$p = .176$
Node21	$F_{(2,32)} = .913$, $p = .957$, $R^2 = .054$	$F_{(2,32)} = 1.386$, $p = .683$, $R^2 = .080$	$F_{(2,32)} = 1.296$, $p = .633$, $R^2 = .075$	$F_{(2,32)} = 1.424$, $p = .694$, $R^2 = .082$
Motion	$p = .211$	$p = .186$	$p = .219$	$p = .177$
Node22	$F_{(2,32)} = .912$, $p = .969$, $R^2 = .054$	$F_{(2,32)} = 1.375$, $p = .702$, $R^2 = .079$	$F_{(2,32)} = 1.268$, $p = .674$, $R^2 = .073$	$F_{(2,32)} = 1.402$, $p = .735$, $R^2 = .081$
Motion	$p = .207$	$p = .171$	$p = .199$	$p = .160$
Node23	$F_{(2,32)} = .916$, $p = .923$, $R^2 = .054$	$F_{(2,32)} = 1.342$, $p = .769$, $R^2 = .077$	$F_{(2,32)} = 1.208$, $p = .796$, $R^2 = .070$	$F_{(2,32)} = 1.358$, $p = .851$, $R^2 = .078$
Motion	$p = .197$	$p = .157$	$p = .175$	$p = .142$
Node24	$F_{(2,32)} = .922$, $p = .889$, $R^2 = .054$	$F_{(2,32)} = 1.315$, $p = .848$, $R^2 = .076$	$F_{(2,32)} = 1.175$, $p = .932$, $R^2 = .068$	$F_{(2,32)} = 1.339$, $p = .980$, $R^2 = .077$
Motion	$p = .194$	$p = .151$	$p = .162$	$p = .131$

Significant effects are underlined and marked in bold. Significance levels of $p > .05$ are seen as non-significant.
* Significant effects not surviving False Discovery Rate correction.

ing skills (Wang et al., 2017; Yeatman, Dougherty, Ben-shachar, et al., 2012), this study further contributes to our understanding of the development of the left AF in children meeting the stringent criteria of a dyslexia diagnosis. We attempted to enhance the interpretation of neural findings across the developmental course of learning to read. Our results showed that, at the pre-reading stage, a cluster of nodes was found in which FA for children developing dyslexia was significantly lower compared to FA for children developing typical reading skills. Interestingly, the results of our three time point longitudinal analyses throughout the course of reading development, i.e. at the pre-reading stage (5–6 years old), early reading stage (7–8 years old) and the advanced reading stage (9–10 years old), revealed that FA values for this specific cluster of nodes remained significantly lower for children developing dyslexia relative to children developing typical reading skills, and that the development of this cluster of nodes in the left AF was not different across groups. Significant results obtained for the FA index were further explored with respect to mean diffusivity (MD), these analyses did not reveal similar group differences. Overall, the longitudinal increase in FA was paralleled by a longitudinal decrease in MD, in line with the expected developmental changes (Hüppi & Dubois, 2006). However, group differences in FA were not related to group differences in MD. Further analyses are required to understand the microstructural properties underlying the results found in the FA index. Extending these findings by a continuous approach on the participant's reading skills confirmed that FA values for the same cluster of nodes in the pre-reading stage and (partly) the early reading stage were correlated to the performance on a word reading test at the concurrent age and later on in reading development. Hence, combining different neuroimaging approaches and combining a categorical and continuous approach revealed that altered white matter organization was present in relation

to poor reading ability. This deviance was present prior to the onset of formal reading instruction as well as throughout the course of reading development.

Inconsistent findings between studies with respect to the neural deficit in the left AF can be considered to be related to different factors. A first factor is related to the study design. The present study extends previous findings by comparing different study design approaches. We detected a cluster of nodes differing in FA between dyslexic readers and typical readers in a cross-sectional comparison in the pre-reader stage, i.e., nodes 18 to 24 showed significantly higher FA for typically developing children. The typical reading group in this study contained a larger proportion of children with a familial risk for dyslexia than there is present in the general population. Therefore, we conducted a similar comparison where the children developing typical reading skills were restricted to children containing no familial risk for dyslexia. Results indicated that the same cluster of nodes showed higher FA in the children developing typical reading skills respectively to the children with dyslexia, indicating that the differences in FA were not fully driven by familial risk.

Many studies investigating white mater structural organization applied a cross-sectional design (Klingberg et al., 2000; Niogi & McCandliss, 2006; Schmithorst et al., 2002). However, the number of longitudinal studies has risen sharply over the years. Most of these studies only contain two (or one) MRI examination supplemented by one or more reading and spelling tests to assess the coherence between white matter structural organization and developmental dyslexia. We argue that a longitudinal design should contain at least three time points (King et al., 2018), which allows a much more thorough investigation of developmental changes. When using two time points, the inferences are bound by the exact MRI time points that were assessed, while effects might change across time intervals. In comparison, a three time point investigation is more reliable as subjects' individual trajectory is less influenced by measurement errors at each time point.

The stability of our findings detected at the pre-reading stage were evaluated throughout the course of reading development (i.e. at the early reading and advanced reading stage) using a longitudinal analysis. The differences in FA found between typical readers and dyslexic readers for the cluster examined did not survive FDR correction. We compared FA for children with typical reading skills and no familial risk for dyslexia to children with dyslexia in the same cluster. This analysis showed an advantage in FA for children with typical reading skills through development. In line with Vanderauwera et al. (2017) and Wang et al. (2017), our results show a significant maturation of FA over time. Interestingly, the maturation rate did not differ between typical readers and dyslexic readers, indicating that the differences that arose early on remained stable over time. Our results indicate that a deficit in the left AF is present in dyslexic readers at the pre-reading stage and thereafter development is similar for both typical readers and dyslexic readers.

These results contradict the conclusions from the ALE meta-analysis by Moreau et al. (2018) and the missing distinction between typical readers and dyslexic readers at a later age as reported in a longitudinal study by Vanderauwera et al. (2017) that examined the same sample of individuals as was used for this study. Our findings indicate that the use of an automated along-the-tract analyses fine-grained approach confirms there is a difference in FA between individuals with and without dyslexia. The structural organization of the left arcuate fasciculus thus seems to be related to (later) reading performance, already prior to the start of formal reading instruction. While having a family risk for dyslexia has been considered to impact reading and brain development (Bishop, 2013) by means of a combined genetic and environmental effect (van Bergen et al., 2014), no evidence of such impact was found on the left AF. Hence, the present results seem to favor the hypothesis that structural organization of the left AF is related to the development of cognitive skills that are predictive of later reading development, such as phonological skills. Such relationship was indeed evidenced in pre-readers (Vanderauwera et al., 2015). Interestingly, recent research pointed towards a very early predictive relationship between FA in the left AF in infancy and phonological skills in kindergarten (Zuk et al., 2019). These studies thus point towards an important, potentially dynamic, relation between the left AF and one of the most important cognitive predictors of later reading ability.

A second factor related to the inconsistent findings between studies with respect to the neural deficit in the left AF could be the choice between a categorical or continuous approach. Recently a plea is made to step away from the traditional categorical group comparison and instead apply a continuous approach (for an opinion see Peters & Ansari, 2019). Adapting the use of a continuous approach corresponds to the continuous distribution of reading skills in the general population, i.e. ranging from individuals with severe reading impairments to individuals with excellent reading skills. The main drawbacks of a categorical approach are (1) the strict cut-off that has to decide to which group each participant belongs, possibly erroneously classifying participants as dyslexic or as typical reader and (2) the absence of generally applicable criteria to assign children to a group, most likely leading to problems when interpreting and comparing different studies. To enhance the probability of a correct classification, ascertaining that reading and spelling problems are both severe and persistent, in line with the standard process for diagnosing dyslexia (American Psychiatric Association, 2013), before classifying a participant as dyslectic reader is advisable. Still, children scoring only slightly above the cut-off will be classified as typical readers when adopting a categorical approach, while their reading skills are far from 'typical'. These classification errors can be prevented by adapting a continuous approach. To investigate the potential additional value of a continuous approach compared to a categorical classification, we investigated the coherence and predictive value of FA in the pre-reading, early reading and advanced reading stage for the previously mentioned cluster of nodes with reading measurements from second to fifth grade, i.e. the same reading measures that were used to classify children in the previous categorical comparisons. White matter organization in the left AF was significantly correlated to reading ability across several grades (R^2 in .150 to .200 range), in line with previous research (Klingberg et al., 2000; Niogi & McCandliss, 2006; R^2 in .380 range), indicating the beneficial employability of this continuous approach above and beyond the categorical approach. The predictive value of these continuous data seemed not driven by the group of individuals meeting the dyslexia criteria, evidenced by similar effect sizes and supported by a normal distribution of the reading skills in the whole sample. FA values from the pre-reading stage were predictive for the scores on reading measures conducted in third, fourth and fifth grade, providing similar evidence to the categorical analysis that a relation between structural organization of the left AF and reading skills later on are present even before the start of reading acquisition. Hence, we hypothesize that the organization of this tract might have an effect later on in development. The relation with the reading scores at the start of grade 2 did not survive FDR correction. We hypothesize that the absence of this predictive relation might be related to the instability of reading scores after one year of reading instruction, relative to the later grades where scores should be stabilized. FA values from the early reading stage were correlated to reading measures from second and third grade, i.e. the reading measurements closest in time to the MRI scan for the early reading stage. No predictive value was found for reading measurements later on in reading development. No correlations were found between the FA values from the cluster of nodes in the advanced reading stage and any of the reading measurements. This result is somewhat surprising, as the longitudinal analysis indicated similar maturation patters for children developing typical reading skills and children developing dyslexia and thus provides evidence for a permanent deficiency. The latter analysis even showed group differences between children developing dyslexia and children developing typical reading skills at this stage. A potential explanation for the absence of a correlation in the advanced reading

stage is that at this time point reading scores might be more influenced by environmental factors, i.e. the home literacy environment, remedial teaching etc., thereby reducing the influence of neurobiology. Also, the decreased sample size might explain the absence of a correlation. In a smaller sample results depend on the specific individuals in that sample, possibly not being representative for the general population. In addition, this smaller sample size leads to loss of statistical power compared to the pre-reading and early reading stage, N = 35 compared to N = 52 and N = 45, respectively. Here the benefit of a longitudinal approach in terms of statistical power is reflected, as our longitudinal analysis did show group differences between children developing dyslexia and children developing typical reading skills at the advanced reading stage.

A final factor explaining inconsistent findings between studies could be the application of different neuroimaging approaches. We conclude that the use of different approaches might lead to differing results. Multiple approaches for examining white matter structure are available, each with their benefits and potential drawbacks (e.g., Kraft et al., 2016; Vanderauwera et al., 2017; Wang et al., 2017). This abundance of neuroimaging approaches and accompanying software packages leads to challenges related to the overall interpretability. We advocate the use of triangulation, i.e. using multiple approaches to analyze data as a means to improve the interpretability of research findings. To assess the advantages of triangulation, we used data from the same sample in which Vanderauwera et al. (2017) previously examined structural organization in the left AF in children developing dyslexia and children developing typical reading skills. The latter study reported a significant difference between children developing dyslexia and children developing typical reading skills for FA at the pre-reading stage, applying an approach in which a mean FA value for a tract is calculated. After two years of reading and writing instruction this difference seemed to have disappeared, which is contradictory given the frequent reports of FA differences in the left AF in adults (Vandermosten et al., 2013; Vandermosten, Boets, Poelmans, et al., 2012; Vandermosten, Boets, Wouters, et al., 2012). In this study, applying a more fine grained neuroimaging analysis approach, a difference in FA between children developing typical reading skills and children developing dyslexia was found to be present throughout reading development, revising the missing distinction at a later age in Vanderauwera et al. (2017; based on the same data) and thereby highlighting the benefits of the triangulation principle. This difference was specific to a cluster of nodes located on the anterior portion of the tract. Although white matter alterations have been reported in the same portion of the tract in children with a family risk for dyslexia (Wang et al., 2017), the specific role of the location of the obtained results remains to be further investigated.

To conclude, by combining different approaches, i.e. applying a fine-grained neuroimaging approach and the comparison of a categorical and continuous approach, the present study revealed that a white matter deficit was present in relation to poor reading ability. This deviance was present prior to the onset of formal reading instruction and throughout the course of reading development. In short, (1) FA for children developing typical reading skills was significantly higher compared to FA for children developing dyslexia in a cluster of nodes in the left AF throughout the three time points assessed in this study. (2) The present study confirms the added value of adopting a longitudinal approach to investigate the neural basis of the developmental process of learning to read, compared to the limited results obtained when restricting studies to a cross-sectional design. (3) The use of a continuous approach opposed to a categorical classification was confirmed as the results for both approaches were largely coherent. This supports the plea for increasingly using a continuous approach when a longitudinal sample of participants from the wide spectrum of reading skill is selected, as this method overcomes problems concerning strict cut-offs and ambiguous classification schemes. In a sample for which a categorical recruitment procedure was applied, i.e. selecting a group with participants with dyslexia and a group of participants with typical reading skills, a categorical classification might be the preferred approach. (4) Following research should strive for triangulation. The application of multiple research approaches aids the interpretation of obtained results and contributes to the exclusion of erroneous findings, e.g. outliers, analysis with insufficient statistical power, wrongful application of research approaches and many other issues.

Declarations of Competing Interest

None.

Acknowledgements

We are most grateful to all children, parents, teachers and schools for their cooperation in this study. A special thanks goes to the Master students that assisted in data collection and to Ron Peeters for technical assistance with MRI data acquisition. This study was funded by the Research Foundation - Flanders (FWO; G.0920.12) and by the Research Council of KU Leuven (OT/12/044). Jolijn Vanderauwera was a postdoctoral fellow of the Research Foundation - Flanders (FWO; 12T4818N).

Credit Author Statement

Stijn Van Der Auwera: Conceptualization, Methodology, Formal analysis, Writing

Maaike Vandermosten: Conceptualization, Data collection, Supervision

Jan Wouters: Conceptualization, Data collection, Supervision

Pol Ghesquière: Conceptualization, Data collection, Supervision

Jolijn Vanderauwera: Conceptualization, Methodology, Formal analysis, Data collection, Writing

Data availability

Participants were not asked to give consent for their data to be made publicly available, so supporting data is not available.

Appendix A

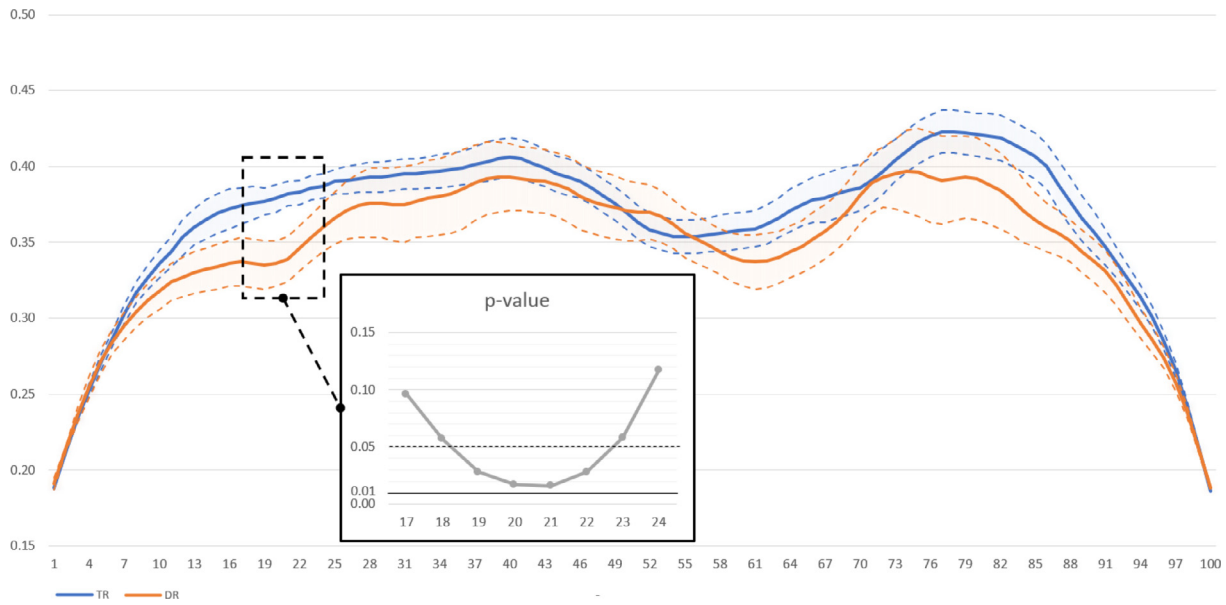

Fig. 3. Mean FA and SE (dotted line) for the DR and TR groups at all 100 nodes of the right AF at the pre-reading stage (N = 34).

Appendix B

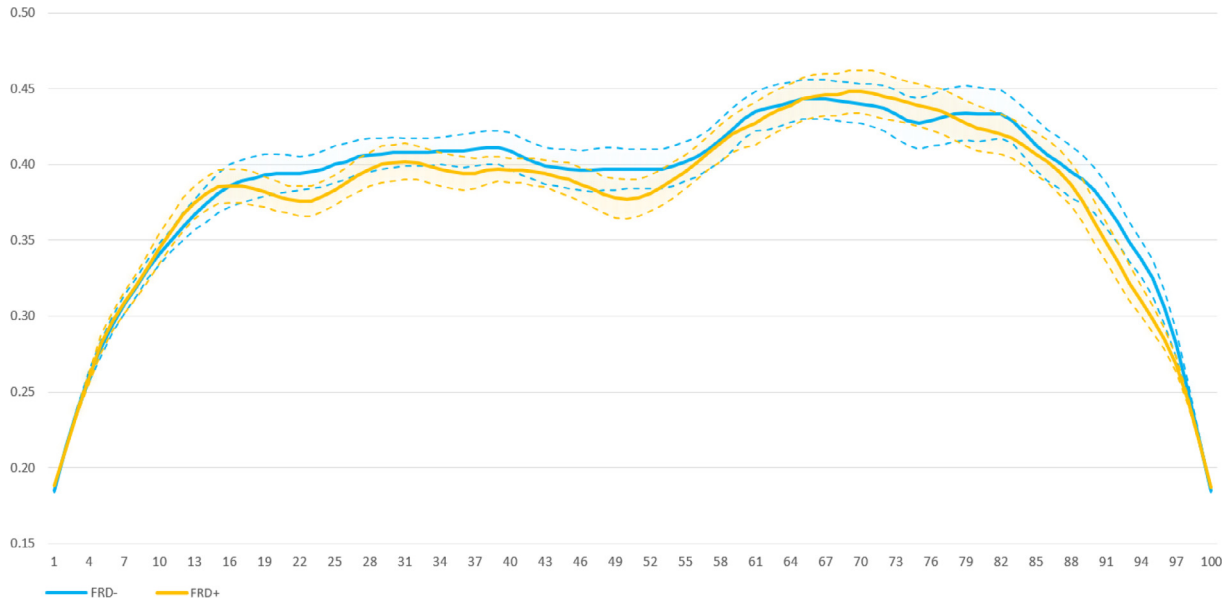

Fig. 4. Mean FA and SE (dotted line) for the FRD^{+} and FRD^{-} groups at all 100 nodes of the left AF at the pre-reading stage (N = 52).

Appendix C

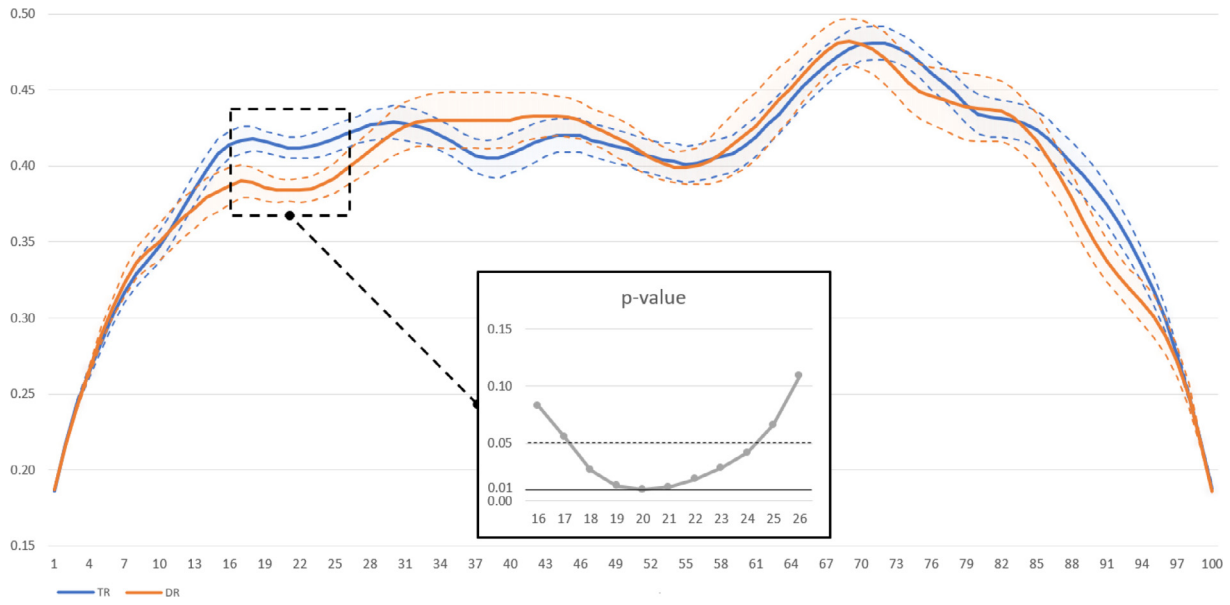

Fig. 5. Mean FA and SE (dotted line) for the DR and TR groups at all 100 nodes of the left AF at the early reading stage (N = 45).

The mean absolute effect size for the detected cluster of nodes (node 18–24) is d = .76, which is close to a large effect size (Cohen, 1988) and over 2.5 SD larger than the mean absolute effect size for all 100 nodes (d = .24).

Appendix D

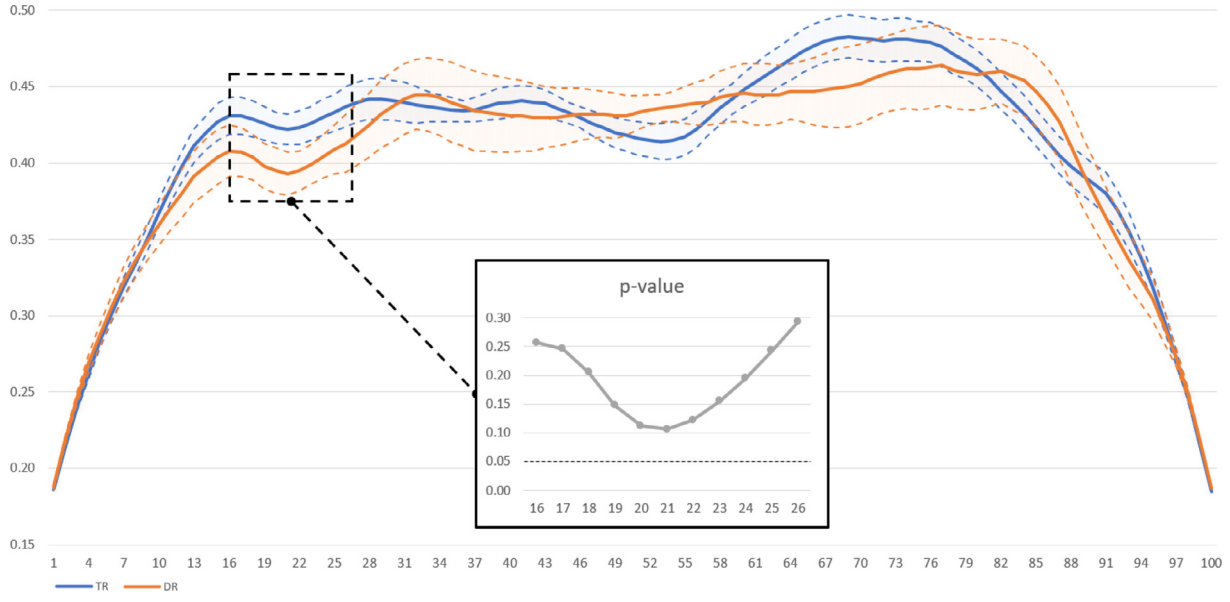

Fig. 6. Mean FA and SE (dotted line) for the DR and TR groups at all 100 nodes of the left AF at the advanced reading stage (N = 35).

The above figure, presenting FA for TR and DR at the advanced reading stage, shows a similar pattern to the figures for the pre- and early reading stage. This cross-sectional comparison for the 35 children for who MRI data was available at the advanced reading stage failed to reach significance at the p = .05 level. However, the mean absolute effect size for the detected cluster of nodes (node 18–24) is d = .53, which is a medium effect size (Cohen, 1988) and over 1.7 SD larger than the mean absolute effect size for all 100 nodes (d = .26). The absence of a significant difference is caused by the reduced sample size and the adherent loss of statistical power, highlighting the benefit of a longitudinal approach. A significant group difference at the advanced reading stage was found when comparing DR and FRD$^-$ TR in our longitudinal analysis.

References

Alexander, D.C., Hubbard, P.L., Hall, M.G., Moore, E.A., Ptito, M., Parker, G.J., Dyrby, T.B., 2010. Orientationally invariant indices of axon diameter and density from diffusion MRI. Neuroimage 52 (4), 1374–1389.

American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders, Fifth edition American Psychiatric Association Publishing *(DSM-5)*.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., Wagenmakers, E.J., 2015. An introduction to Bayesian hypothesis testing for management research. J. Manag. 41 (2), 521–543. doi:10.1177/0149206314560412.

Assaf, Y., Pasternak, O., 2008. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review. J. Mol. Neurosci. 34 (1), 51–61. doi:10.1007/s12031-007-0029-0.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29 (4), 1165–1188. doi:10.1016/j.mcm.2008.03.008.

Bishop, D., 2013. Cerebral asymmetry and language development: cause, correlate, or consequence? Science 340 (6138). doi:10.1126/science.1230531.

Boets, B., De Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., Ghesquiére, P., 2010. Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. Br. J. Dev. Psychol. 28 (1), 5–31. doi:10.1348/026151010X485223.

Boets, B., Op De Beeck, H.P., Vandermosten, M., Scott, S.K., Gillebert, C.R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., Ghesquière, P., 2013. Intact but less accessible phonetic representations in adults with dyslexia. Science 342 (6163), 1251–1254. doi:10.1126/science.1244333.

Boets, B., Wouters, J., van Wieringen, A., Ghesquière, P., 2007. Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory. Neuropsychologia 45 (8), 1608–1620. doi:10.1016/j.neuropsychologia.2007.01.009.

Boudreau, B., Poulin, C., 2009. An examination of the validity of the Family Affluence Scale II (FAS II) in a general adolescent population of Canada. Social Indic. Res. 94 (1), 29–42. doi:10.1007/s11205-008-9334-4.

Brus, B.T., Voeten, M.J.M, 1973. Een minuut test, vorm A en B. Berkhout.

Catani, M., de Schotten, M.T., 2008. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44 (8), 1105–1132. doi:10.1016/j.cortex.2008.05.004.

Catani, M., Jones, D.K., Ffytche, D.H., 2005. Perisylvian language networks of the human brain. Ann. Neurol. 57 (1), 8–16. doi:10.1002/ana.20319.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences. Routledge Acad..

Dehaene, S., 2009. Reading in the brain. Penguin.

Duan, F., Zhao, T., He, Y., Shu, N., 2015. Test-retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI study. J. Magn. Reson. Imaging 42 (4), 1106–1116. doi:10.1002/jmri.24859.

Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J.F., Le Bihan, D., Dehaene-Lambertz, G., 2009. Structural asymmetries in the infant language and sensori-motor networks. Cereb. Cortex 19 (2), 414–423. doi:10.1093/cercor/bhn097.

Dudal, P. (1997). Leerlingvolgsysteem VCLB (CSBO), Spelling: Toetsen 1-2-3. In *Basisboek en kopieerbundel [Student trajectory system. Spelling Grade 1-2-3. Manual]*. Garant.

Eluvathingal, T.J., Hasan, K.M., Kramer, L., Fletcher, J.M., Ewing-Cobbs, L., 2007. Quantitative diffusion tensor tractography of association and projection fibers in normally developing children and adolescents. Cereb. Cortex 17 (12), 2760–2768. doi:10.1093/cercor/bhm003.

Gullick, M.M., Booth, J.R., 2015. The direct segment of the arcuate fasciculus is predictive of longitudinal reading change. Dev. Cogn. Neurosci. 13, 68–74. doi:10.1016/j.dcn.2015.05.002.

Huber, E., Donnelly, P.M., Rokem, A., Yeatman, J.D., 2018. Rapid and widespread white matter plasticity during an intensive reading intervention. Nat. Commun. 9 (1), 1–13. doi:10.1038/s41467-018-04627-5.

Hüppi, P.S., Dubois, J., 2006. Diffusion tensor imaging of brain development. Semin. Fetal Neonatal Med. 11 (6), 489–497. doi:10.1016/j.siny.2006.07.006.

JASP Team. (2020). *JASP (Version 0.11.1)[Computer software]*.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841. doi:10.1006/nimg.2002.1132.

King, K.M., Littlefield, A.K., McCabe, C.J., Mills, K.L., Flournoy, J., Chassin, L., 2018. Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. Dev. Cogn. Neurosci. 33 (November), 54–72. doi:10.1016/j.dcn.2017.11.009.

Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J.D.E., Moseley, M.E., Poldrack, R.A, 2000. Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging. Neuron 25 (2), 493–500. doi:10.1016/S0896-6273(00)80911-3.

Kraft, I., Schreiber, J., Cafiero, R., Metere, R., Schaadt, G., Brauer, J., Neef, N.E., Müller, B., Kirsten, H., Wilcke, A., Boltze, J., Friederici, A.D., Skeide, M.A., 2016. Predicting early signs of dyslexia at a preliterate age by combining behavioral assessment with structural MRI. Neuroimage 143, 378–386. doi:10.1016/j.neuroimage.2016.09.004.

Langer, N., Peysakhovich, B., Zuk, J., Drottar, M., Sliva, D.D., Smith, S., Becker, B.L.C., Grant, P.E., Gaab, N, 2017. White matter alterations in infants at risk for developmental dyslexia. Cereb. Cortex 27 (2), 1027–1036. doi:10.1093/cercor/bhv281.

Law, J.M., Vandermosten, M., Ghesquiere, P., Wouters, J., 2014. The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. Front. Human Neurosci. 8 (July), 1–12. doi:10.3389/fnhum.2014.00482.

Lebel, C., Benischek, A., Geeraert, B., Holahan, J., Shaywitz, S., Bakhshi, K., Shaywitz, B., 2019. Developmental cognitive neuroscience developmental trajectories of white matter structure in children with and without reading impairments. Dev. Cogn. Neurosci. 36 (March), 100633. doi:10.1016/j.dcn.2019.100633.

Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J. Neurosci. Methods 264, 47–56. doi:10.1016/j.jneumeth.2016.03.001.

López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A., de Diego-Balaguer, R., 2013. Word learning is mediated by the left arcuate fasciculus. PNAS 110 (32), 13168–13173. doi:10.1073/pnas.1301696110.

Lu, N., Han, Y., Chen, T., Gunzler, D.D., Xia, Y., Lin, J.Y., Tu, X.M., 2013. Power analysis for cross-sectional and longitudinal study designs. Shanghai Arch. Psychiatry 25 (4), 259–262. doi:10.3969/j.issn.1002-0829.2013.04.009.

Moreau, D., Stonyer, J.E., McKay, N.S., Waldie, K.E., 2018. No evidence for systematic white matter correlates of dyslexia: an activation likelihood estimation meta-analysis. Brain Res. 1683, 36–47. doi:10.1016/j.brainres.2018.01.014.

Niogi, S.N., McCandliss, B.D., 2006. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. Neuropsychologia 44 (11), 2178–2188. doi:10.1016/j.neuropsychologia.2006.01.011.

Ozernov-Palchik, O., Norton, E.S., Wang, Y., Beach, S.D., Zuk, J., Wolf, M., Gabrieli, J.D.E., Gaab, N, 2019. The relationship between socioeconomic status and white matter microstructure in pre-reading children: a longitudinal investigation. Hum. Brain Mapp. 40 (3), 741–754. doi:10.1002/hbm.24407.

Pennington, B.F., 2006. From single to multiple deficit models of developmental disorders. Cognition 101 (2), 385–413. doi:10.1016/j.cognition.2006.04.008.

Pennington, B.F., Lefly, D.L., 2001. Early reading development in children at family risk for dyslexia. Child Dev. 72 (3), 816–833. doi:10.1111/1467-8624.00317.

Peters, L., Ansari, D., 2019. Are specific learning disorders truly specific, and are they disorders? Trends Neurosci. Educ. 17 (July 2018). doi:10.1016/j.tine.2019.100115.

Peterson, R.L., Pennington, B.F., 2012. Developmental dyslexia. Lancet North Am. Ed. 379 (9830), 1997–2007. doi:10.1016/S0140-6736(12)60198-6.

Pugh, K.R., Mencl, W.E., Jenner, A.R., Katz, L., Frost, S.J., Lee, J.R., Shaywitz, S.E., Shaywitz, B.A., 2000. Functional neuroimaging studies of reading and reading disability (developmental dyslexia). Ment. Retard. Dev. Disabil. Res. Rev. 6 (3), 207–213.

Raven, J.C., Court, J.H., Raven, J., 1984. Manual for Raven's progressive matrices and vocabulary scales. Lewis.

Reynolds, J.E., Long, X., Grohs, M.N., Dewey, D., Lebel, C., 2019. Structural and functional asymmetry of the language network emerge in early childhood. Dev. Cogn. Neurosci. 39 (March), 100682. doi:10.1016/j.dcn.2019.100682.

Saygin, Z.M., Norton, E.S., Osher, D.E., Beach, S.D., Cyr, A.B., Ozernov-Palchik, O., Yendiki, A., Fischl, B., Gaab, N., Gabrieli, J.D.E, 2013. Tracking the roots of reading ability: White matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. J. Neurosci. 33 (33), 13251–13258. doi:10.1523/JNEUROSCI.4383-12.2013.

Schmithorst, V.J., Wilke, M., Dardzinski, B.J., Holland, S.K., 2002. Correlation of white matter diffusivity and anisotropy with age during childhood and adolescence: A cross–sectional diffusion-tensor MR imaging study. Radiology 222 (1), 212–218.

Snowling, M.J., 2000. Dyslexia. Blackwell (Vol. 2nd).

Theys, C., Wouters, J., Ghesquière, P., 2014. Diffusion tensor imaging and resting-state functional MRI-scanning in 5- and 6-year-old children: Training protocol and motion assessment. PLoS One 9 (4), 1–7. doi:10.1371/journal.pone.0094019.

van Bergen, E., van der Leij, A., de Jong,, P.F., 2014. The intergenerational multiple deficit model and the case of dyslexia. Front. Human Neurosci. 8 (June), 1–13. doi:10.3389/fnhum.2014.00346.

Van den Bos, K.P., Spelberg, H., Scheepsma, A., De Vries, J., 1994. De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling. Berkhout.

Vanderauwera, J., Vandermosten, M., Dell'Acqua, F., Wouters, J., Ghesquière, P., 2015. Disentangling the relation between left temporoparietal white matter and reading: a spherical deconvolution tractography study. Hum. Brain Mapp. 36 (8), 3273–3287. doi:10.1002/hbm.22848.

Vanderauwera, J., Wouters, J., Vandermosten, M., Ghesquière, P., 2017. Early dynamics of white matter deficits in children developing dyslexia. Dev. Cogn. Neurosci. 27, 69–77. doi:10.1016/j.dcn.2017.08.003.

Vandermosten, M., Boets, B., Poelmans, H., Sunaert, S., Wouters, J., Ghesquière, P., 2012. A tractography study in dyslexia: Neuroanatomic correlates of orthographic, phonological and speech processing. Brain 135 (3), 935–948. doi:10.1093/brain/awr363.

Vandermosten, M., Boets, B., Wouters, J., Ghesquière, P., 2012. A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. Neurosci. Biobehav. Rev. 36 (6), 1532–1552. doi:10.1016/j.neubiorev.2012.04.002.

Vandermosten, M., Poelmans, H., Sunaert, S., Ghesquière, P., Wouters, J., 2013. White matter lateralization and interhemispheric coherence to auditory modulations in normal reading and dyslexic adults. Neuropsychologia 51 (11), 2087–2099. doi:10.1016/j.neuropsychologia.2013.07.008.

Vandermosten, M., Vanderauwera, J., Theys, C., De Vos, A., Vanvooren, S., Sunaert, S., Wouters, J., Ghesquière, P., 2015. A DTI tractography study in pre-readers at risk for dyslexia. Dev. Cogn. Neurosci. 14, 8–15. doi:10.1016/j.dcn.2015.05.006.

Vanvooren, S., Poelmans, H., Hofmann, M., Ghesquière, P., Wouters, J., 2014. Hemispheric asymmetry in auditory processing of speech envelope modulations in prereading children. J. Neurosci. 34 (4), 1523–1529. doi:10.1523/JNEUROSCI.3209-13.2014.

Vogel, A.C., Church, J.A., Power, J.D., Miezin, F.M., Petersen, S.E., Schlaggar, B.L., 2013. Functional network architecture of reading-related regions across development. Brain Lang. 125 (2), 231–243. doi:10.1016/j.bandl.2012.12.016.

Wakana, S., Caprihan, A., Panzenboeck, M.M., Fallon, J.H., Perry, M., Gollub, R.L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., Mori, S., 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage 36 (3), 630–644. doi:10.1016/j.neuroimage.2007.02.049.

Wandell, B.A., Yeatman, J.D., 2013. Biological development of reading circuits Brian. Curr. Opin. Neurobiol. 23 (2), 261–268. doi:10.1038/jid.2014.371.

Wang, Y., Mauer, M.V., Raney, T., Peysakhovich, B., Becker, B.L.C., Sliva, D.D., Gaab, N, 2017. Development of tract-specific white matter pathways during early reading development in at-risk children and typical controls. Cereb. Cortex 27 (4), 2469–2485. doi:10.1093/cercor/bhw095.

Wechsler, D., 2005. Wechsler Intelligence Scale for Children–Third Edition. Dutch Adaptation (WISC-III-NL).

Yeatman, J.D., Dougherty, R.F., Ben-shachar, M., Wandell, B.A., 2012. Development of white matter and reading skills. Natl. Acad. Sci. 109 (4), E3045–E3053. doi:10.1073/pnas.1206792109.

Yeatman, J.D., Dougherty, R.F., Myall, N.J., Wandell, B.A., Feldman, H.M., 2012. Tract profiles of white matter properties: automating fiber-tract quantification. PLoS One 7 (11), e49790. doi:10.1371/journal.pone.0049790.

Yeatman, J.D., Dougherty, R.F., Rykhlevskaia, E., Sherbondy, A.J., Deutsch, G.K., Wandell, B.A., Ben-Shachar, M., 2011. Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. J. Cogn. Neurosci. 23 (11), 3304–3317. doi:10.1162/jocn_a_00061.

Ziegler, J.C., Castel, C., Pech-Georgel, C., George, F., Alario, F.X., Perry, C., 2008. Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. Cognition 107 (1), 151–178. doi:10.1016/j.cognition.2007.09.004.

Zuk, J., Yu, X., Sanfilippo, J., Figuccio, M.J., Dunstan, J., Carruthers, C., Sideridis, G., Gagoski, B., Grant, P.E., Gaab, N., 2019. White matter in infancy is prospectively associated with language outcome in kindergarten. BioRxiv doi:10.1101/781914.