Investigating the impact of early literacy training on white matter structure in pre-readers at risk for dyslexia

Maria Economou[1,2,✉], Shauni Van Herck[1,2], Femke Vanden Bempt[1,2], Toivo Glatz[1,2,3], Jan Wouters[1], Pol Ghesquiere[2], Jolijn Vanderauwera[1,2,4,5], and Maaike Vandermosten[1]

[1] Research group ExpORL, Department of Neurosciences, KU Leuven, 3000 Leuven, Belgium

[2] Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, KU Leuven, 3000 Leuven, Belgium

[3] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Public Health, Charitéplatz 1, 10117 Berlin, Germany

[4] Psychological Sciences Research Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

[5] Institute of Neuroscience, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

Correspondence to: Maria Economou

E-mail: maria.economou@kuleuven.be

Full address: ExpORL, Dept. Neurosciences KU Leuven, Herestraat 49 bus 721, 3000 Belgium

Abstract

Recent pre-reading evidence demonstrates that white matter alterations are associated with dyslexia even before the onset of reading instruction. At the same time, remediation of reading difficulties is suggested to be most effective when provided as early as kindergarten, yet evidence is currently lacking on the early neuroanatomical changes associated with such preventive interventions. To address this open question, we investigated white matter changes following early literacy intervention in Dutch-speaking pre-readers (aged 5-6) with an increased cognitive risk for developing dyslexia. Diffusion-weighted images were acquired before and after a 12-week digital intervention in three groups: i) at-risk children receiving phonics-based training (n = 31), ii) at-risk children engaging with active control training (n = 25) and iii) typically developing children (n = 27) receiving no intervention. Following automated quantification of white matter tracts relevant for reading, we first examined baseline differences between at-risk and typically developing children, revealing bilateral dorsal and ventral differences. Longitudinal analyses showed that white matter properties changed within the course of the training, however the absence of intervention-specific results suggests that these changes rather reflect developmental effects. This study contributes important first insights on the neurocognitive mechanisms of intervention that precedes formal reading onset.

Short running title: White matter structure after early literacy training

Keywords: diffusion-weighted imaging; dyslexia; intervention; pre-readers; white matter

Learning to read is one of the most important educational achievements of early childhood, setting the foundation for long-term learning. Reading development relies on the reorganization of existing brain circuits including dorsal temporo-parietal, ventral occipito-temporal and inferior frontal regions, that facilitate the processing and integration of orthographic, phonological and semantic information (Pugh et al. 2001; Schlaggar and McCandliss 2007). Although the majority of children learn to read with relative ease during the first school years, around 7% are diagnosed with developmental dyslexia, which means they experience severe and persistent reading difficulties that occur despite adequate opportunity and instruction (Peterson and Pennington 2015). In most educational settings, dyslexia is typically diagnosed after years of demonstrated reading failure, which has a considerable impact on socio-emotional well-being (Valas 1999; Grills-Taquechel et al. 2012) and future educational attainment (Dougherty 2003). Importantly, a gap in reading ability between dyslexic and typical readers is shown to be present as early as first grade (Ferrer et al. 2015), and is likely to persist throughout the school years and potentially widen over time if not addressed properly (Stanovich 1986; Ferrer et al. 2015). Research has shown that extensive reading interventions (comprising at least 100 sessions) are shown to be more effective if provided during kindergarten and first grade, compared to later grades (Wanzek and Vaughn 2007). Yet, in current practice, struggling readers are likely to receive the necessary support only after years of reading failure and thus after the most sensitive period for intervention has passed. This has been described as the "dyslexia paradox" (Ozernov-Palchik and Gaab 2016). With the aim of reducing or closing the gap, preventive interventions (i.e. interventions that take place before the onset of reading acquisition) offer promising potential to overcome the consequences of this paradox and ensure that reducing, or even closing the gap between struggling and typical readers does not become unattainable. To date however, no studies have investigated the neuroanatomical changes

associated with preventive interventions in young children at risk for dyslexia. Hence, the goal of this study is to characterize white matter microstructural changes during such an early intervention. In the last two decades, an abundance of neuroimaging studies provided evidence of atypical brain structure in dyslexia, and especially in white matter. Long-range white matter pathways carry signals between cortical regions and therefore constitute a key part of the reading circuitry. Diffusion-weighted imaging (DWI) studies, which enable the identification and quantification of white matter connections, have identified both dorsal and ventral pathways linked to reading: the arcuate fasciculus (AF), the inferior longitudinal fasciculus (ILF), and the inferior fronto-occipital fasciculus (IFOF) (Ben-Shachar et al. 2007; Vandermosten, Boets, Wouters, et al. 2012; Wandell and Yeatman 2013; Zhang et al. 2014). The AF dorsally connects inferior frontal to superior temporal and inferior parietal cortex (Catani et al. 2005). The left AF is thought to constitute a key region for sensory-motor mapping of sound to articulation (Saur et al. 2008) as well as grapheme-phoneme mapping (Gullick and Booth 2014). Indeed, several studies support the role of the left AF sustaining phonological aspects of reading (Yeatman et al. 2011; Saygin et al. 2013; Vanderauwera et al. 2015; Vandermosten et al. 2015), while reduced fractional anisotropy (FA) has been reported in both children (Vanderauwera et al. 2017; Van Der Auwera et al. 2021) and adults (Vandermosten, Boets, Poelmans, et al. 2012) with dyslexia. The IFOF and ILF are ventral association tracts connecting the occipital to the inferior frontal and to the anterior temporal cortex respectively (Saur et al. 2008; Forkel et al. 2014). The left IFOF has been linked to orthographic aspects of reading in both adults and school-aged children (Vandermosten, Boets, Poelmans, et al. 2012; Vanderauwera et al. 2018; Banfi et al. 2019). A different pattern seems to emerge in pre-readers however, with evidence suggesting associations between bilateral IFOF and early phonological ability (Walton et al. 2018) as well as early reading-related cognitive measures (Vanderauwera et al. 2018). Properties of bilateral ILF have been shown to correlate with

phonological decoding in school-aged children (Lebel et al. 2013), while in pre-reading children, left ILF properties were related to emergent literacy skills including both phonological and orthographic aspects (Vanderauwera et al. 2018). In terms of anatomy, both the IFOF and ILF likely contain projections adjacent to the posterior occipito-temporal cortex around the visual word form area (VWFA) (Yeatman et al. 2013), a region specialized in visual letter recognition, and thus important for word recognition and consequently reading (Dehaene and Cohen 2011).

Most strikingly, recent studies in pre-readers suggest that these white matter alterations associated with dyslexia are evident even before the onset of reading acquisition. More specifically, properties of the left fronto-temporal and fronto-parietal AF segments and left IFOF were shown to differ between pre-readers with and without a familial risk for dyslexia (Vandermosten et al. 2015; Langer et al. 2017; Wang et al. 2017). In addition to a link with familial risk, FA differences in bilateral AF were already present in pre-reading children who later developed dyslexia compared to children who later developed typical reading skills (Vanderauwera et al. 2017). Longitudinal evidence of children throughout reading development also suggests that white matter properties of the left AF_{direct} in kindergarten were predictive of later reading outcomes (Vanderauwera et al. 2017; Borchers et al. 2019; Van Der Auwera et al. 2021), while the rate of white matter development was significantly associated with subsequent reading skill (Yeatman et al. 2012a; Wang et al. 2017). Moreover, recent evidence points out that early deficits in the left AF_{direct} of dyslexic compared to typical readers remain present throughout primary school (Van Der Auwera et al. 2021). Taken together, these findings provide converging evidence that early differences in white matter structure predate the formal onset of reading, suggesting a central role in determining (a)typical trajectories of reading development.

In light of these findings, the question arises whether preventive intervention can target these early white matter deviances and potentially have an impact on subsequent reading development.

Previous research provides convincing evidence that white matter is malleable and subject to learning- and experience-dependent changes in controlled designs. Studies in school-aged children demonstrated that explicit training in high-level cognitive domains such as spelling and math, resulted in behavioural improvements accompanied by training-induced white matter plasticity (Gebauer et al. 2012; Jolles et al. 2016). Such training-induced plasticity was also demonstrated in a younger age group, where Ekerdt et al. (2020) showed white matter changes in typically developing 4-year-old preliterate children following three weeks of oral word learning. More specifically, children in the oral word training group had increased FA in the left dorsal precentral gyrus, when compared to both active and passive control groups.

Only a few studies have directly investigated structural plasticity following targeted reading intervention. Following a reading intervention in school-aged children between 6 and 9 years old, increased cortical thickness was reported in treatment responders, when compared to treatment nonresponders and a waiting control group (Romeo et al. 2018). With respect to white matter studies, Keller and Just (2009) demonstrated for the first time that poor readers (aged 8-12) who received intensive reading instruction over the course of 6 months, showed increased FA in the left centrum semiovale in contrast to both poor readers who received no training and good readers. Using a longitudinal design with several imaging assessments, Huber et al. (2018) provide additional evidence for reading intervention-related white matter plasticity in children with reading difficulties. The authors reported widespread changes in diffusion properties together with significant growth in reading skill, after a training period of eight weeks. On the other hand, Partanen et al. (2020) did not find any evidence of intervention-related white matter changes following a three-month intervention in dyslexic children. Overall, these studies provide important insights on how behavioral interventions targeting reading skills can induce measurable changes in white matter properties of school-aged children, although evidence is lacking on the impact of

interventions during the sensitive period for learning to read. Moreover, the absence of adequate inclusion of control groups in many training studies makes it difficult to separate maturation from learning or experience effects. As a result, very little is known about the mechanisms of early intervention, particularly when it takes place prior to formal reading instruction.

The present work aimed to address these remaining gaps in knowledge by investigating the impact of a preventive intervention in 5-year-old pre-readers with an increased cognitive risk for developing dyslexia. Our objectives were twofold. First, we aimed to examine cognitive risk-related white matter differences prior to intervention by comparing at-risk with typically developing pre-readers. Second, by means of a controlled intervention design, we sought to characterize intervention-related white matter changes following 12 weeks of digital at-home training. In order to comprehensively address these goals, we assessed white matter microstructure longitudinally in i) at-risk children who received phonics-based training adapted for kindergarten, ii) at-risk children engaging with another immersive game that did not train literacy skills, and iii) typically developing children receiving no intervention. We hypothesized phonics-based training-related FA changes in white matter connections sustaining early phonological and orthographic processing, therefore focusing on four bilateral pathways: the fronto-temporal and fronto-parietal segments of the dorsal AF (AF_{direct} and $AF_{anterior}$ respectively) and the ventral ILF and IFOF.

Materials and Methods

Study design

Diffusion-weighted images were collected before and after a 12-week training period to evaluate changes in white matter structure. These data were collected as part of a larger longitudinal

intervention project (N = 149) in which additional behavioral and neurophysiological (Vanden Bempt et al. 2021; Van Herck et al. 2021) assessments were carried out. The current study focuses on the direct effects of phonics-based training on white matter properties.

Participants

The participants of this study were 90 monolingual Dutch-speaking children in their third year of kindergarten, which is the last year before starting primary school. Participants had no prior known hearing or neurological problems, no formal diagnosis or risk for ADHD (Smidts and Oosterlaan 2007) and no history of speech and language therapy. All children had received comparable kindergarten education up until the beginning of the study and had not received any formal reading or writing instruction, in accordance with Flemish government guidelines (https://onderwijsdoelen.be/). The data reported in this study were collected between December 2018 and June 2019, thus before the start of the Covid-19 pandemic.

Children with and without an increased cognitive risk for developing dyslexia were selected based on a screening conducted at school at the beginning of their last year of kindergarten. Figure 1 provides a detailed description of participant flow in the study. Over 1900 families responded to a first questionnaire, 1225 of whom met the inclusion criteria described above and were selected to participate in a risk screening. Non-verbal reasoning and subsequent intelligence estimates were calculated using a tablet-based version of the Raven's Colour Progressive Matrices (Raven et al. 1984). Three early literacy predictors were assessed to define risk criteria: letter knowledge, phonological awareness and rapid automatized naming (Puolakanaho et al. 2007; Caravolas et al. 2012). After excluding incomplete or unreliable scores (n = 21) and children with the lowest 10% of non-verbal intelligence estimates (n = 113), percentile scores for the early literacy measures were calculated in a sample of 1091 Flemish kindergarteners. Children were

classified as at-risk if they performed below the 30th percentile on two out of the three measures and below the 40th percentile on letter knowledge. They were classified as typically developing if they performed above the 40th percentile on all measures. See Verwimp et al. (2020) for more details on the screening procedure and behavioral assessments.

According to the criteria described above, 60 children with and 30 children without a cognitive risk were selected to take part in the intervention study. Of the 60 at-risk children, 31 received a tablet-based intervention that aimed at training early literacy skills, while the other 29 children engaged with a comparably immersive control intervention. Group assignment for the at-risk children was done using a block randomization procedure (R package *randomizr*; Coppock 2019), to balance the covariates of sex and birth trimester across groups. The group of typically developing children (n=30) was matched for sex, age and non-verbal intelligence and participated in the behavioral and MRI assessments but not in any intervention.

For all cognitive and neuroimaging assessments, written informed consent and verbal assent were obtained from all primary caregivers and children respectively. This study was approved by the Medical Ethics Committee of the Leuven University Hospital.

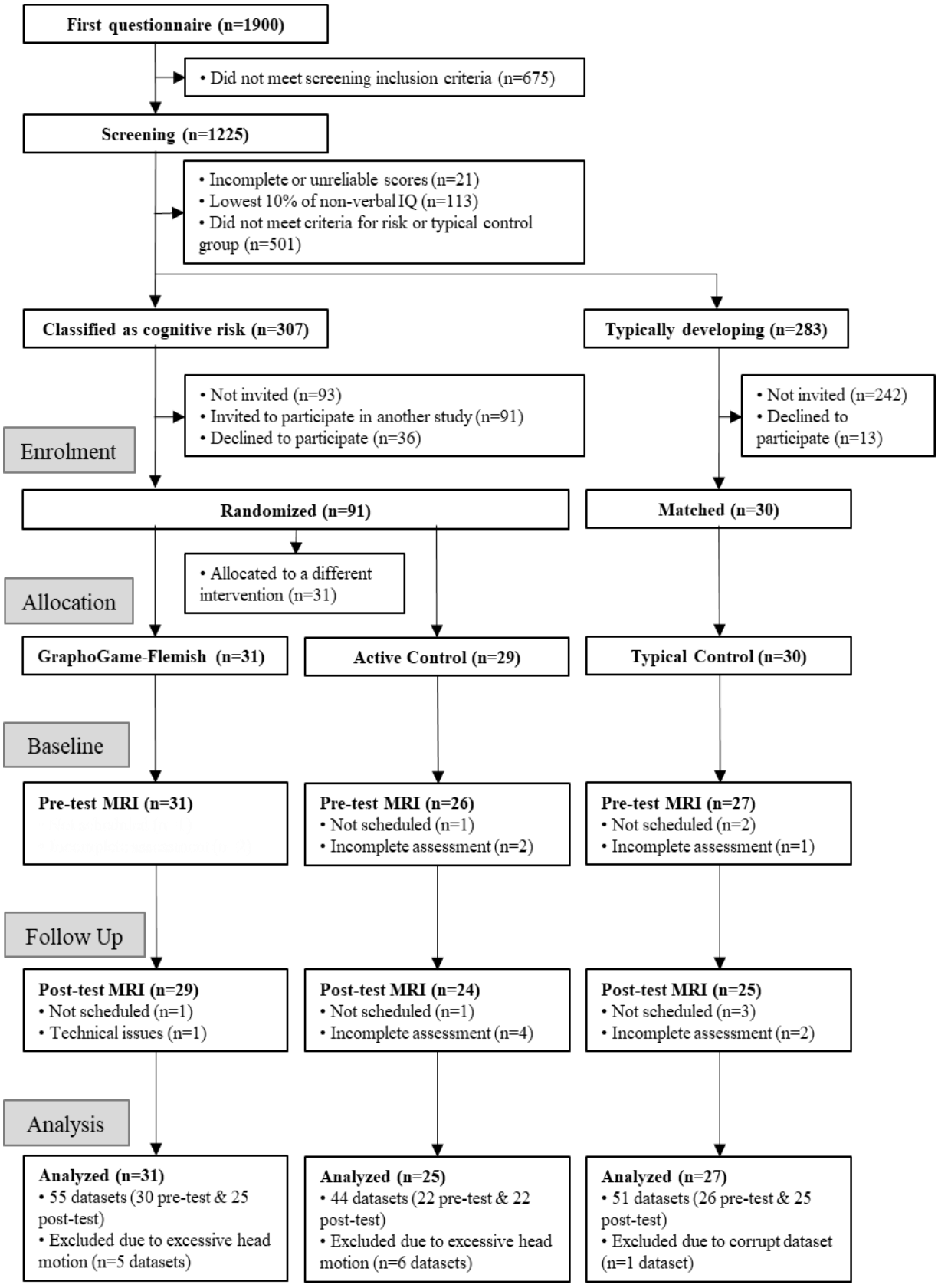


Figure 1 Adapted CONSORT flow diagram

Tablet-based intervention

In the present study, early literacy intervention focused on the systematic training of grapheme-phoneme coupling (how letters correspond to sounds), a fundamental aspect of learning to read in alphabetic scripts. Interventions that involve this type of training were shown to be the most beneficial in ameliorating reading deficits in poor readers (Galuschka et al. 2014) and were also suggested to be beneficial for young children at risk of reading delay (Hatcher et al. 2004; Snowling and Hulme 2011; 2012; Lonigan et al. 2013). For the specific purposes of this study, we adapted the GraphoGame interface (Richardson and Lyytinen 2014) to create a new tablet-based version (GraphoGame-Flemish; GG-FL) which is appropriate for Flemish kindergarteners without formal reading instruction. Briefly, the training started by introducing the graphemes, providing exercises of visual and auditory discrimination of graphemes and phonemes, while gradually progressing to exercises of grapheme-phoneme coupling, phoneme blending and counting, and early reading. See Glatz et al. (2021) for a detailed description of the game content and level progression.

Children in the active control group (AC) received a different type of training. For this purpose, six commercially available tablet-based games (LEGO City My City, LEGO DUPLO Town, LEGO DUPLO Connected Train, LEGO Friends: Heartlake Rush, PLAYMOBIL Horse Farm, and PLAYMOBIL Police) were implemented within a single custom-made application. The games were selected because they were age-appropriate, were motivating enough for a training period of 12 weeks, and did not train early literacy skills. These characteristics were confirmed by pilot data of five-year-old kindergarteners collected before the start of the study.

The intervention was intended for at-home use, whereby all participating families received the necessary equipment (Samsung Galaxy Tab E9.6 tablet and Audiotechnica ATH M20X

headphones). The games were installed on a child-friendly profile and the app icons were modified into colored stars to conceal the group assignment. They were asked to play their game (GG-FL or AC) for 15 minutes a day, six days a week for a total duration of 12 weeks. As part of the larger longitudinal project, both intervention groups received an additional story listening game, which they were asked to use for 10 minutes a day in addition to their normal training schedule. The effects of story listening fall beyond the scope of the current study and will not be discussed further. Note however, that the two groups (GG-FL and AC) did not differ in the amount of story listening ($W = 378.5$, $p = .805$) or the total amount of intervention-related tablet exposure ($W = 361$, $p = .480$). All families received additional support during the intervention period in the form of a detailed instruction manual, intervention calendar to monitor progress, as well as stickers and small rewards for reaching milestones during the training period. Data logs were transmitted daily to a central university server and contact with the parents was maintained throughout the intervention in the case of technical or motivational/compliance issues.

MRI data acquisition

Participants were invited for two MRI sessions, one before and one shortly after the intervention phase. All images were acquired on a 3T Philips Achieva MRI scanner (Philips, Best, The Netherlands) using a 32-channel head coil and identical acquisition parameters. The sequence consisted of seven non-diffusion-weighted volumes, followed by 20, 32 and 60 diffusion-weighted volumes with b-values of 700, 1000 and 2000 s/mm^2 respectively. Additional scanning parameters were as follows: 62 transverse slices, slice thickness = 2.2 mm, voxel size 2.2 x 2.3 x 2.2 mm^3, TR/TE (ms) = 3593/88ms, flip angle = 90°, multi-band factor MB = 2, EPI factor = 43 and acquisition time = 7:38 min. In the same session, a reversed phase encoding direction (posterior-anterior) sequence was acquired for subsequent distortion correction during processing.

Anatomical 3D T1-weighted images were acquired using a CS-SENSE TFE sequence with following parameters: 240 sagittal slices, isotropic voxel size 0.9 mm^3, TR/TE (ms) = 9.1/4.2, acquisition time = 3:30 min. In line with previous MRI studies of young children (Raschle et al. 2009; Theys et al. 2014; Vanderauwera et al. 2017), an age-appropriate preparation protocol was used to familiarize the children with the scanning procedure and minimize motion artifacts. In order to make the experience more pleasant, the scanner room was decorated in the protocol story theme and the children were given time to acclimatize with the surrounding environment. Communication was maintained during the scan via a built-in microphone system and close monitoring was possible using a camera system. To limit head movement, custom-fit inflatable pads were used for head positioning, as well as positive encouragement during scanning. The children watched a cartoon of their choice during scan acquisition. All participants received a small non-monetary reward at the end of each visit.

Of the total sample of 90 children reported in the present study, MRI data are not available for six individuals at pre-test who either did not wish to participate in MRI assessments (n = 2), or were not able to complete any of the MRI examinations (n = 4). At post-test, MRI data are not available for four additional participants who did not plan a second visit and one dataset that was compromised due to technical issues (see Figure 1).

MRI data pre-processing

Diffusion MRI data were pre-processed using the FMRIB Software Library v6.0 (FSL; Jenkinson et al. 2012). A pair of non-diffusion-weighted images in forward- and reversed-encoding directions was used for estimating the off-resonance field (Andersson et al. 2003) and correcting for susceptibility distortions (Smith et al. 2004). The volumes of all shells of the DWI series were subsequently corrected for motion and eddy current distortions (Andersson and Sotiropoulos 2016)

followed by brain mask estimation (Smith 2002). Scalar FA maps were produced using FSL'S DTIFIT, which fits the tensor model on the data. The amount of head motion in each subject was quantified as the mean volume displacement relative to the first volume of the DWI series. This was done by calculating the root mean square of the three translation parameters (x,y,z), similar to Yendiki et al. (2014). Eleven datasets ($n = 5$ at pre-test and $n = 6$ at post-test) with mean translational movement exceeding 2.2 mm (acquisition voxel size) were excluded from further analyses.

Fiber tract identification and quantification

White matter tract identification and quantification was done using the open-source tool Automated Fiber Quantification (AFQ; Yeatman et al. 2012b) implemented in MATLAB 2016a. First, whole-brain deterministic tractography was performed using a streamlines tracking algorithm (STT) with a fourth-order Runge–Kutta path integration method and following parameters: step size = 1, FA threshold = 0.2, min length = 50mm, max length = 250mm, angle = 30°. This and subsequent steps were performed using all acquired diffusion-weighted volumes from all shells. Next, a waypoint ROI-based tract segmentation was used to define white matter tracts (as described in Wakana et al. 2007), followed by tract refinement and cleaning (removal of fibers deviating more than 4 SD above mean fiber length or more than 5 SD from the fiber core). For a detailed description of the AFQ procedure see Yeatman et al. (2012b). The resulting reconstructed tract was then resampled to 100 equidistant nodes spanning the full length of the tract and a weighted FA average was calculated at each node to create a tract profile (ordered anterior to posterior). The weighting procedure was such that fibers deviating from the fiber core count less than those closer to the fiber core. Because the first and last node could not be reliably estimated in all tracts and subjects, they

were not included in further analyses. Both the tract profile as well as summarized tract FA measures were analyzed.

The analysis focused on four bilateral pathways of interest defined *a priori*: the direct fronto-temporal and anterior fronto-parietal segments of the AF (AF_{direct} and $AF_{anterior}$ respectively), the IFOF and the ILF (Vandermosten, Boets, Wouters, et al. 2012; Wandell and Yeatman 2013; Zhang et al. 2014). The investigation of both left- and right-hemispheric pathways is warranted by evidence that 1) reveals bilateral associations between white matter and early language and phonological skills (Vandermosten et al. 2015; Vanderauwera et al. 2018; Walton et al. 2018; Farah et al. 2020) and 2) suggests a potential role of right-hemispheric pathways in promoting better reading outcomes, by means of either compensatory (Hoeft et al. 2011) or protective mechanisms (Wang et al. 2017; Zuk et al. 2020). Here, it is important to clarify that the pathway corresponding to $AF_{anterior}$ in the present study, is referred to as SLF in AFQ, which is consistent with the terminology used in Wakana et al. (2007). Tract identification was performed in 150 datasets from a total sample of 83 subjects. The following tracts could not be identified in certain datasets using the chosen tensor-based tractography and AFQ procedure: left AF_{direct} ($n = 7$), right AF_{direct} ($n = 34$), left $AF_{anterior}$ ($n = 3$), right $AF_{anterior}$ ($n = 2$), left IFOF ($n = 2$), right IFOF ($n = 4$), left ILF ($n = 0$), right ILF ($n = 1$). Notably, a higher number is observed for the right AF_{direct} (fronto-temporal segment), which is largely consistent with previous research showing that this pathway may not be detectable in certain individuals, including studies in both children (Eluvathingal et al. 2007; Van Der Auwera et al. 2021) and adults (Catani et al. 2007; Lebel and Beaulieu 2009; Allendorfer et al. 2016). This inability to detect the right AF_{direct} has been attributed to limitations of deterministic algorithms in regions of crossing fibers (as is the case in the temporo-parietal junction), rather than the anatomical absence of the tract from the brain (Yeatman et al. 2011).

Statistical analyses

All reported analyses were performed in R (version 4.0.0) (R Core Team 2020). For all analyses, the alpha level was set at .05. First, we examined white matter risk-related differences prior to intervention. Note that in these pre-test FA comparisons, for the purpose of assessing the effect of cognitive risk, the two risk groups were merged (i.e. GG-FL and AC groups, $n = 52$) and compared to the typically developing group ($n = 26$). To this end, average tract FA measures (i.e. tract average) and FA values at each node of the tract (i.e. tract profiles consisting of 98 nodes) at pre-test were compared between the at-risk children and typically developing children using a linear model, while controlling for in-scanner subject motion.

Given the inherently correlated node structure within each tract profile, Bonferroni correction for multiple comparison corrections can be considered overly conservative and potentially lead to type-II errors. Therefore, a nonparametric permutation-based multiple testing correction was applied for along-tract analyses, as described in Nichols and Holmes (2002) and implemented in several tractometry papers (Yeatman et al. 2012b; Travis et al. 2017; Farah et al. 2020; Jossinger et al. 2021; Wasserthal et al. 2021). This approach provides the family-wise error (FWE) corrected alpha value for pointwise tests at a given tract, as well as a FWE-corrected cluster size threshold at the user-defined alpha of .05. Additional details on the implementation of this procedure are provided in the Supplementary Material.

Group differences were considered significant if they occurred at nodes where the *p*-value was lower than the FWE-corrected alpha, or if they occurred in a sufficient number of adjacent nodes as determined by the FWE-corrected cluster size, without the need for further *p*-value adjustment. For the baseline risk-related analysis, corrected alpha thresholds for the different tracts ranged from $p < .001$ to $p < .003$, while cluster-based thresholds ranged from 11 to 14 consecutive nodes.

To assess changes in white matter over time and intervention effects, average tract FA values (i.e. tract average) were entered in linear mixed-effects models (LMMs) with fixed-effects of group (GG-FL, AC, typical control), session (pre-test, post-test) and their interaction, as well as by-subject random intercepts. The same model specification was then used to assess these factors at each node (i.e. tract profiles) to identify clusters of nodes with similar effects. In all reported LMMs, in-scanner subject motion was included as a nuisance predictor. All LMMs were fitted using *lmer* via the R package *lme4* (Bates et al. 2015) and *F*-statistics and *p*-values were obtained by calculating type-III analysis of variance tables using Satterthwaite's degrees-of-freedom approximation (R package *lmerTest*, Kuznetsova et al. 2017). In order to determine the appropriate *p*-value or cluster size for the longitudinal analysis, the same permutation-based approach as the one described above was applied. For the intervention analysis, corrected alpha thresholds for the different tracts ranged from $p < .001$ to $p < .002$, while cluster-based thresholds ranged from 8 to 11 consecutive nodes.

Effect sizes for all pre-test comparisons are reported as Cohen's *d* values, while effect sizes for LMM main effects and interactions are reported as Cohen's f values (Cohen 1988).

Data availability

Anonymized data and code to reproduce the figures and findings of this study are publicly available at https://github.com/meconomou/Economou2021_wmintervention.

Results

Participant and group characteristics

Descriptive group statistics are summarized in Table 1. The three groups did not differ significantly with respect to non-verbal intelligence, sex, handedness, socio-economic status, or age at either MRI visit. More children in the GG-FL group had a familial risk for dyslexia, however this was not controlled for further given the very low number of occurrences in our sample. A Kruskal-Wallis test revealed no group differences in head motion at either timepoint (pre-test: $H(2) = 0.29$, $p = .865$; post-test: $H(2) = 0.96$, $p = .618$). There was however a main effect of session ($F(1,74) = 11.12$, $p = .001$), such that less head movement was observed at post-test compared to pre-test. As summarized in Table 2, both intervention groups were overall comparable in terms of training duration and total game exposure. Most children played at least 90% of the intended amount (estimated at 18h) and successful completion of the intervention took slightly longer than the intended 12-week period.

Table 1 Group characteristics

	Group			
Variable	**GraphoGame-Flemish *n* = 31[a]**	**Active Control *n* = 25[a]**	**Typical Control *n* = 27[a]**	***p*-value[b]**
Sex (female)	16/15	11/14	15/12	.700
Familial risk (n)	5	1	0	.035
Non-verbal intelligence	102 (78 - 133)	96 (79 - 126)	101 (78 - 133)	.394
Handedness[c]				.857
Left/Right/Ambidextrous	2/28/1	2/21/2	2/22/3	
Age at pre-test scan (months)	64 (60 - 71)	66 (60 - 72)	66 (62 - 73)	.075
Age at post-test scan (months)	69 (65 - 76)	70 (65 - 76)	70 (65 - 77)	.734
Socio-economic status[d]				.112
Low/Middle/High/Unknown	7/12/12/0	7/13/4/1	3/10/14/0	

[a]Median (Range) or Occurrence (*n*).

[b]Group differences were assessed using a Kruskal-Wallis test for age and non-verbal intelligence, a Pearson's Chi-squared test for sex, handedness and maternal education, and a Fisher's exact test for familial risk.

[c]Based on a subset of questions adapted from the Edinburgh Handedness Inventory (Oldfield, 1971).

[d]Assessment based on maternal educational level (low = no extra degree after secondary school; middle = professional bachelor/academic bachelor; high = Master or PhD).

Table 2 Training characteristics

Variable	Group		
	GraphoGame-Flemish *n* = 31[a]	Active Control *n* = 25[a]	*p*-value[b]
Training exposure (hours)	16 (3 - 25)	18 (6 - 28)	.315
Training period (weeks)	12 (6 - 19)	13 (10 - 20)	.483
Proportion of completed intervention[c] (%)	91 (15 - 137)	101 (32 - 158)	.315

[a]Median (Range).

[b]Group differences were assessed using a Wilcoxon rank sum test.

[c]The proportion of completed intervention is calculated as the number of hours spent on-task divided by the total amount of hours they were instructed to play during 12 weeks (estimated at 18h).

Risk-related differences in white matter properties before the intervention

Tract average

Here we examined the differences in FA between at-risk and typically developing children (Figure 2). Overall, the at-risk group had significantly lower FA in the left AF_{direct} (β = -0.01, t = -2.02, p = .047, d = 0.53), left IFOF (β = -0.01, t = -2.15, p = .035, d = 0.62) and right IFOF (β = -0.01, t = -2.23, p = .029, d = 0.59), but this difference did not reach significance in the right AF_{direct} (β = -0.02, t = -1.97, p = .054, d = 0.55). We did not find any evidence for risk-related differences in global tract FA in bilateral $AF_{anterior}$ (left: β = -0.01, t = -1.52, p = .132, d = 0.38; right: β = -0.01, t = -1.13, p = .262, d = 0.32) or bilateral ILF (left: β = -0.01, t = -1.56, p = .123, d = 0.40; right: β = -0.01, t = -0.97, p = .335, d = 0.23). Tracts for which group differences were found in the tract average are depicted with bold black panel borders in Figure 2.

Tract profiles

The results of the tract profile analysis are shown in Figure 2. Based on the permutation-based multiple testing correction described in the Methods section, significant group differences between at-risk and typically developing children were observed in localized node clusters of both dorsal and ventral pathways. Lower FA for the at-risk group was found in the right AF_{direct} (nodes 11-13, average $d = 0.47$), right ILF (nodes 8-9, average $d = 0.30$) and right IFOF (node 37, $d = 0.58$). These locations are visualized as red bands on individual tract renderings and as grey-shaded bars on the tract profile plots in Figure 2. No further localized group differences were found in the left AF_{direct}, left ILF, left IFOF and bilateral $AF_{anterior}$. Along-tract effect sizes for these group comparisons are visualized in Supplementary Figure 1. Note that the analyses on the right AF_{direct} have been conducted with a considerably smaller sample size compared to the other tracts, due to the high number of failed reconstructions. As a result, it is possible that findings concerning the right AF_{direct} might not be as reliable.

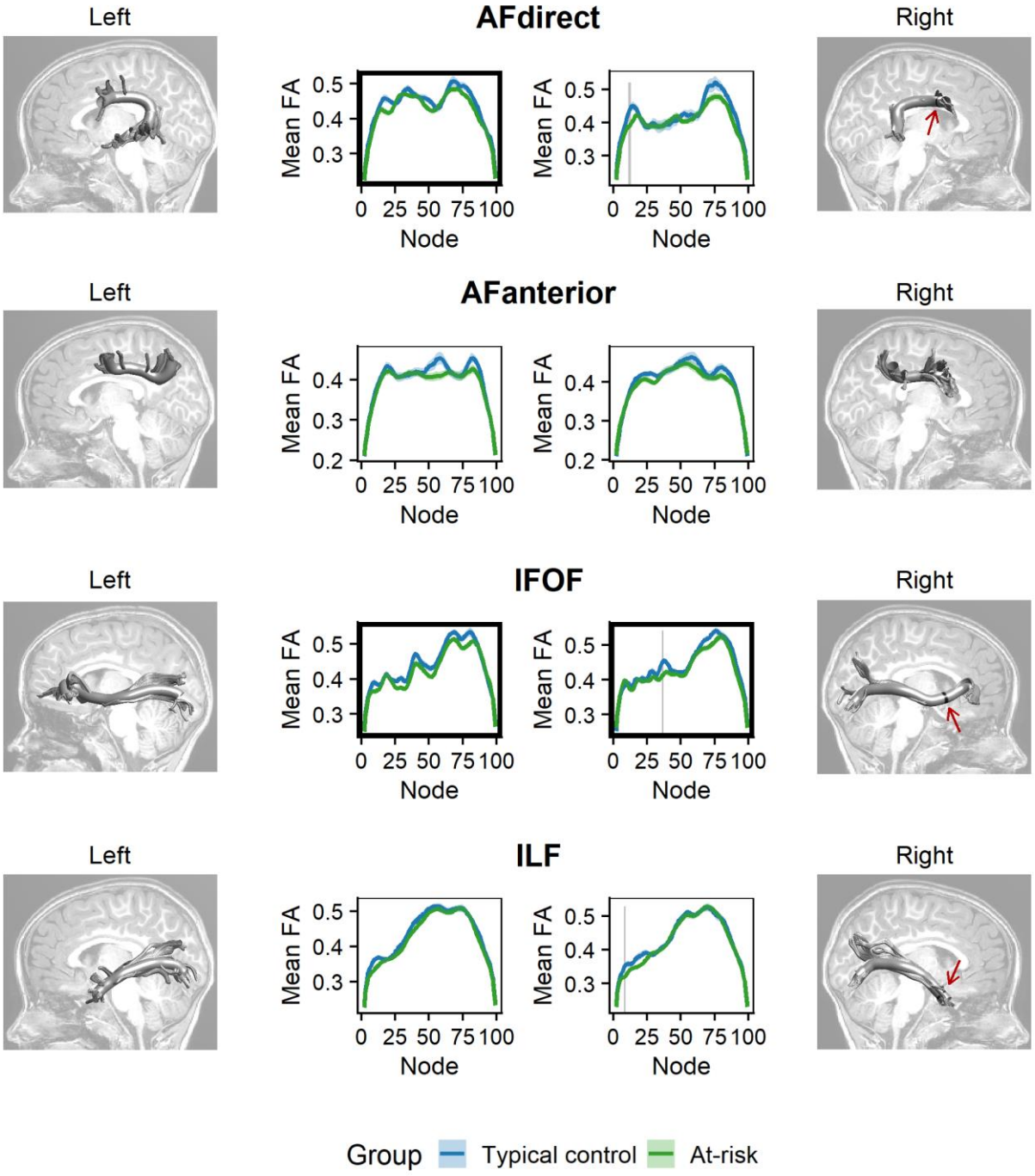


Figure 2 Tract FA profiles before the intervention.

Each plot shows the mean tract FA profile ± 1 standard error of the mean for the two groups (at-risk and typical control). Bold black panel borders indicate the tracts where global tract FA differences were found. Localized risk-related differences are indicated by grey-shaded regions on the tract profile. The red arrows indicate the position of the corresponding anatomical locations, which are visualized in black on three-dimensional tract renderings next to each plot.

Training effects on white matter properties

Tract average

A main effect of session (pre-test, post-test) was present in several pathways, which was always positive, indicating a systematic increase in FA over time. This effect was significant in the left AF_{direct} ($F(1,63) = 8.09$, $p = .006$, $f = 0.36$), left $AF_{anterior}$ ($F(1,68) = 5.34$, $p = .024$, $f = 0.28$), left IFOF ($F(1,69) = 4.75$, $p = .033$, $f = 0.26$) and left ILF ($F(1,76) = 5.86$, $p = .018$, $f = 0.28$), but did not reach significance in the right AF_{direct} ($F(1,51) = 3.96$, $p = .052$, $f = 0.28$), right $AF_{anterior}$ ($F(1,70) = 2.16$, $p = .146$, $f = 0.18$), right IFOF ($F(1,70) = 3.86$, $p = .053$, $f = 0.23$) or right ILF ($F(1,72) = 2.04$, $p = .158$, $f = 0.17$). This analysis revealed no evidence of intervention-related group differences in global tract FA, as indicated by the absence of significant group-by-session interactions in global FA for any of the tracts (right $AF_{anterior}$: $F(2, 67) = 2.65$, $p = .078$; all other $p > .213$). The individual tract average FA changes across sessions are shown in Supplementary Figure 2.

Tract profiles

Longitudinal changes in FA were subsequently examined at each node along the pathways, comparing the three groups. A main effect of session was present in the left $AF_{anterior}$ (nodes 44-57, average $f = 0.10$), such that an increase in FA was observed from pre-test to post-test. Similar to the global tract average analysis, no significant group-by-session interactions were found. Along-tract effect sizes for the session main effect and the group-by-session interaction effect are visualized in Supplementary Figures 3 and 4 respectively.

Discussion

The present study examined white matter properties in pre-reading kindergarten children with a cognitive risk for developing dyslexia during an early literacy intervention. The main aim of this study was to investigate the effects of preventive training on white matter development. To this end, at-risk children were assigned to either a digital phonics-based training (GG-FL group) or engaged with other games unrelated to reading (AC group). Before the intervention, we found evidence of global and local risk-related differences in bilateral dorsal and ventral tracts, suggesting that white matter alterations related to dyslexia predate the formal onset of reading. Our longitudinal analysis revealed that while white matter properties changed from pre- to post-test, no evidence was found to support literacy intervention in more detail in the following paragraphs.

Concerning risk-related white matter differences before the intervention, we observed lower FA in bilateral dorsal and ventral pathways in children who performed poorly on reading-related precursors, when compared to typically developing children. More specifically, global average FA differences were found in the left AF_{direct}, left IFOF and right IFOF, as well as more localized differences in the right AF_{direct}, right IFOF and right ILF. Our findings are consistent with recent evidence reporting pre-reading differences in children with (a familial risk for) dyslexia (Langer et al. 2017; Vanderauwera et al. 2017; Wang et al. 2017). Together, these observations support the view that early underlying anatomical differences represent a neurobiological predisposition for struggling to read and are not purely driven by reading difficulties.

With respect to location, previous studies in adults and school-aged children have predominantly emphasized the role of left-hemispheric pathways in skilled reading and dyslexia (Yeatman et al. 2011; Vandermosten, Boets, Poelmans, et al. 2012; Zhang et al. 2014). In the present study, our analyses indicate additional risk-related differences in the right homologues, revealing a more

bilateral pattern. This is less commonly reported in older children and adults, but is largely in line with pre-reading evidence supporting more bilateral involvement in younger children (see Vandermosten et al. 2016 for a review). This bilateral involvement might shift towards a left-lateralized pattern within the first few years of formal reading instruction and/or once dyslexia can be diagnosed, however additional follow-up assessments would be needed to specifically investigate this hemispheric specialization within our sample.

In examining these pre-reading differences, a tractometry approach was utilized to map FA along the trajectory of white matter tracts (Bells et al. 2011; Yeatman et al. 2012b; Chamberland et al. 2019; Chandio et al. 2020). Such a fine-grained approach may be more appropriate in investigating regional variations in FA than using global average tract measures, especially given the underlying influence of local anatomical features on FA estimation. Indeed, studies have demonstrated that FA varies significantly along the course of white matter tracts, with these variation patterns remaining relatively consistent across subjects (Yeatman et al. 2012b; Johnson et al. 2014; Farah et al. 2020). This observation is also confirmed by the present data (see tract profile visualization in Figure 2).

Notably, both global tract FA differences (in left AF_{direct} and bilateral IFOF) as well as localized FA differences (in right AF_{direct}, right ILF and right IFOF) were associated with the presence of a cognitive risk for dyslexia in our study. These localized group differences were identified following a rigorous permutation-based approach to control the FWE rate (Nichols and Holmes 2002) and were observed only in small clusters, ranging from one to three nodes. It is important to clarify here that the number of nodes analyzed, as well as the approach for determining the portion of the tract that will be analyzed, vary considerably among studies. As a result, the size of reported clusters is often not directly comparable. With this in mind, our results are in agreement with

previously reported cluster sizes in the literature (see for example Wang et al. 2017; Banfi et al. 2019), although it remains uncertain how anatomically meaningful these small clusters are.

In order to aid interpretation of our findings, effect sizes are provided for the risk-related comparisons as Cohen's *d* values (see Supplementary Figure 1). Based on suggested thresholds (Cohen 1988), a number of along-tract moderate effects are observed in bilateral AF_{direct}, bilateral ILF and right IFOF. Despite these moderate effects, the localized group differences in the left AF_{direct} and left ILF were not identified as significant following correction for multiple comparisons. It is important not to entirely dismiss these results, as they provide relevant insights on the practical implications of our findings, which are independent of significance and sample size (Sullivan and Feinn 2012). Moreover, reporting effect sizes enables comparison of similar effects across studies, which is preferred over solely relying on arbitrary thresholds for interpretation (Lakens 2013).

After 12 weeks of training, an effect of session was observed in several tracts, indicating an increase in FA over the course of the intervention. The absence of any group-by-session interaction however, suggests that this increase is not uniquely related to the early literacy training and likely reflects developmental effects. This finding is consistent with previous reports of significant FA development in early childhood (Reynolds et al. 2019; Dimond et al. 2020). Although the available studies capture these developmental FA changes in the course of years, our results point out that changes in shorter timescales (i.e. a few months) may also be detectable at this age.

The absence of longitudinal GG-specific changes in FA may suggest that the training provided in this study is not associated with measurable changes in white matter. This interpretation however, warrants caution. Currently available evidence from reading intervention studies is limited and findings are mixed regarding the location and magnitude of training-induced white matter changes. Following intensive reading instruction, Keller and Just (2009) reported a localized FA increase in

the left anterior centrum semiovale, whereas Huber et al. (2018) described widespread white matter growth during an intensive reading intervention, including changes in the reading network that we investigated. In contrast to both of these studies and therefore more in line with the findings of the current study, Partanen et al. (2020) reported no measurable effects of a three-month reading intervention on white matter properties of dyslexic readers.

Several factors contribute to the observed variability among studies, including intervention-related aspects or divergence in study design and analysis. For instance, interventions that target specific aspects of reading acquisition may lead to more regional changes, as was the case in the study of Keller and Just (2009) (and might be less straightforward to detect with a region-of-interest analysis). On the other hand, widespread plasticity may be a reflection of the general learning experience resulting from receiving multi-componential training, as was the case in Huber et al. (2018). Similarly, the inclusion of an active control group (such us in our study and that of Keller and Just (2009)) or lack thereof (such as in Huber et al. (2018) and Partanen et al. (2020)) might yield different patterns of intervention effects. A recent meta-analysis has shown that the lack of randomization or comparison with a business-as-usual group instead of an active control group, has an impact on estimated effect sizes in early literacy intervention research (Verhoeven et al. 2020). As such, research design is a very relevant factor to consider when interpreting (null) intervention effects Intervention characteristics such as intensity and exposure are also likely to play a key role in explaining the findings of the current study. Both Keller and Just (2009) and Huber et al. (2018) report more than a 100 hours of intervention received by the participants, which is markedly higher than the total intervention exposure in our study and could explain why we did not find an effect on white matter properties. The intervention reported in the current study led to improvements in letter sound knowledge and word decoding (Vanden Bempt et al. 2021). Yet, given that extensive interventions comprising 100 or more sessions are put forward as the standard

for effective remediation (Wanzek and Vaughn 2007), the training provided in our study might have only induced behavioral effects. Lastly, the role of the statistical framework used in the present study needs to be addressed when drawing conclusions about the findings. Other methodologically similar studies often use less conservative approaches to adjust for multiple along-tract comparisons, compared to the approach described in the current study. The methodological choice can have an influence on both the results and the conclusions and therefore should be factored in when interpreting findings within and across studies. For completeness, uncorrected *p*-values for the analyses are shown in Supplementary Figures 5-7.

A final consideration when discussing the present findings and the variation among reported intervention results, concerns developmental effects. While some evidence exists to support plasticity in school-aged children following reading interventions, less is known about white matter changes during early (preschool) interventions. In the present study, we specifically investigate the pre-reading brain and therefore focus on structural changes during the sensitive period of learning to read (Ozernov-Palchik and Gaab 2016), whereas the studies described above examine intervention mechanisms in later stages of reading development. Hence, it might be that the impact of such intervention on white matter is more profound after the onset of primary school compared to the pre-reading stage. Another possibility is that the timing of the intervention provided in our study might be associated with other structural changes besides white matter. For instance, recent evidence shows that cortical gray matter volume in left temporal and temporo-parietal regions linked to reading is still increasing within the first few years of primary school (Phan et al. 2021). This provides an indication that gray matter structure is amenable to (intervention-induced) plasticity during early childhood.

A number of limitations are recognized in the present study. First, while FA has been extensively used in studies of reading- and dyslexia-related differences (Ben-Shachar et al. 2007;

Vandermosten, Boets, Wouters, et al. 2012) and has been established as a valuable index of neuroplasticity (Sagi et al. 2012; Zatorre et al. 2012), its biological interpretation remains limited. Due to the demonstrably low specificity of FA for a single neurobiological process, the underlying neural changes that might influence FA change cannot be understood based on this index alone. Access to complementary information about underlying tissue architecture, ideally from multi-modal investigations, would be an advantage of future studies and could uncover a different pattern of results. Recent evidence mapping intervention-related white matter changes to more specific neurobiological processes (Huber et al. 2021) further corroborates this view. A second consideration pertains to the cohort sample size, despite best efforts and a large-scale screening in more than 1000 pre-readers. This is especially relevant for our analyses in the right AF_{direct}, where lower FA was observed in at-risk children. Given that the right AF_{direct} was not successfully reconstructed in a relatively large subset of participants (30%), this finding is potentially less robust and should be interpreted with caution. Owing to similar constraints, exploring individual differences was not possible and we thus report only intervention effects at the group level. The sample size, dropout and data quality are undoubtedly major practical challenges of longitudinal neuroimaging studies in young children up to the age of five, such as the present one (Turesky et al. 2021). They nevertheless advance our understanding of intervention mechanisms during this sensitive period of early reading acquisition.

To summarize, here we report on a longitudinal investigation of white matter properties in 5-year-old kindergarten children following an early literacy training. Our findings extend previous research supporting an early link between structural white matter alterations and a cognitive risk for developing reading difficulties. The results provide evidence of an overall change in white matter properties within the course of the intervention, however no specific training-related effects were observed. The present study is one of the first to investigate the neurocognitive mechanisms

of preventive intervention in young pre-reading children. Longitudinal follow-up of the children's reading skills will help clarify the long-lasting impact of preventive training, which in turn has the potential to inform future remediation strategies for at-risk readers at a time when they are most beneficial.

Acknowledgements

We would like to thank Cara Verwimp, Fran Vanfleteren, Klara Schevenels, Lauren Blockmans, Justine Soetaert, Caitlan Nys, Valerie Geerits and Aymara Taillieu for their assistance with MRI protocol preparation and data collection. We are grateful to Ron Peeters for technical MRI support. We are most thankful to all the participating families and children without whom this research would not be possible. This work was supported by the Research Council of KU Leuven (C14/17/046). J.V. was a postdoctoral fellow of the Research Foundation Flanders (12T4818N).

References

Allendorfer JB, Hernando KA, Hossain S, Nenert R, Holland SK, Szaflarski JP. 2016. Arcuate fasciculus asymmetry has a hand in language function but not handedness. Human Brain Mapping. 37(9):3297–3309. doi:10.1002/hbm.23241.

Andersson JLR, Skare S, Ashburner J. 2003. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. NeuroImage. 20(2):870–888. doi:10.1016/S1053-8119(03)00336-7.

Andersson JLR, Sotiropoulos SN. 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage. 125:1063–1078. doi:10.1016/j.neuroimage.2015.10.019. http://dx.doi.org/10.1016/j.neuroimage.2015.10.019.

Banfi C, Koschutnig K, Moll K, Schulte-Körne G, Fink A, Landerl K. 2019. White matter alterations and tract lateralization in children with dyslexia and isolated spelling deficits. Human Brain Mapping. 40(3):765–776. doi:10.1002/hbm.24410.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 67(1):1–48. doi:10.18637/jss.v067.i01.

Bells S, Cercignani M, Deoni S, Assaf Y. 2011. "Tractometry" – comprehensive multi-modal quantitative assessment of white matter along specific tracts. Proceedings of the International Society for Magnetic Resonance in Medicine. 19(2009):678. http://cds.ismrm.org/protected/11MProceedings/files/678.pdf.

Ben-Shachar M, Dougherty RF, Wandell BA. 2007. White matter pathways in reading. Current Opinion in Neurobiology. 17(2):258–270. doi:10.1016/j.conb.2007.03.006.

Borchers LR, Bruckert L, Dodson CK, Travis KE, Marchman VA, Shachar MB, Feldman HM. 2019. Microstructural properties of white matter pathways in relation to subsequent reading abilities in children : a longitudinal analysis. Brain Structure and Function. 224(2):891–905. doi:10.1007/s00429-018-1813-z. http://dx.doi.org/10.1007/s00429-018-1813-z.

Caravolas M, Lervåg A, Mousikou P, Efrim C, Litavský M, Onochie-Quintanilla E, Salas N, Schöffelová M, Defior S, Mikulajová M, et al. 2012. Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. Psychological Science. 23(6):678–686. doi:10.1177/0956797611434536.

Catani M, Allin MPG, Husain M, Pugliese L, Mesulam MM, Murray RM, Jones DK. 2007. Symmetries in human brain language pathways correlate with verbal recall. Proceedings of the National Academy of Sciences of the United States of America. 104(43):17163–17168. doi:10.1073/pnas.0702116104.

Catani M, Jones DK, Ffytche DH. 2005. Perisylvian language networks of the human brain. Annals of Neurology. 57(1):8–16. doi:10.1002/ana.20319.

Chamberland M, Raven EP, Genc S, Duffy K, Descoteaux M, Parker GD, Tax CMW, Jones DK. 2019. Dimensionality reduction of diffusion MRI measures for improved tractometry of the

human brain. NeuroImage. 200(March):89–100. doi:10.1016/j.neuroimage.2019.06.020. https://doi.org/10.1016/j.neuroimage.2019.06.020.

Chandio BQ, Risacher SL, Pestilli F, Bullock D, Yeh FC, Koudoro S, Rokem A, Harezlak J, Garyfallidis E. 2020. Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. Scientific Reports. 10(1):1–18. doi:10.1038/s41598-020-74054-4. https://doi.org/10.1038/s41598-020-74054-4.

Cohen J. 1988. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates.

Coppock A. 2019. Randomizr: Easy-to-use tools for common forms of random assignment and sampling. https://CRAN.R-project.org/package=randomizr.

Dehaene S, Cohen L. 2011. The unique role of the visual word form area in reading. Trends in Cognitive Sciences. 15(6):254–262. doi:10.1016/j.tics.2011.04.003. http://dx.doi.org/10.1016/j.tics.2011.04.003.

Dimond D, Rohr CS, Smith RE, Dhollander T, Cho I, Lebel C, Dewey D, Connelly A, Bray S. 2020. Early childhood development of white matter fiber density and morphology. NeuroImage. 210(October 2019):116552. doi:10.1016/j.neuroimage.2020.116552. https://doi.org/10.1016/j.neuroimage.2020.116552.

Dougherty C. 2003. Numeracy, literacy and earnings: Evidence from the National Longitudinal Survey of Youth. Economics of Education Review. 22(5):511–521. doi:10.1016/S0272-7757(03)00040-2.

Ekerdt CEM, Kühn C, Anwander A, Brauer J, Friederici AD. 2020. Word learning reveals white matter plasticity in preschool children. Brain Structure and Function. 225(2):607–619. doi:10.1007/s00429-020-02024-7. http://link.springer.com/article/10.1007/s00429-020-02024-7?utm{_}source=researcher{_}app{\&}utm{_}medium=referral{\&}utm{_}campaign=RESR{_}MRKT{_}Researcher{_}inbound.

Eluvathingal TJ, Hasan KM, Larry K, Fletcher JM, Ewing-Cobbs L. 2007. Quantitative Diffusion Tensor Tractography of Association and Projection Fibers in Normally Developing Children and Adolescents. Cerebral Cortex. 17(12):2760–2768. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf.

Farah R, Tzafrir H, Horowitz-Kraus T. 2020. Association between diffusivity measures and language and cognitive-control abilities from early toddler's age to childhood. Brain Structure and Function. 225(3):1103–1122. doi:10.1007/s00429-020-02062-1. https://doi.org/10.1007/s00429-020-02062-1.

Ferrer E, Shaywitz BA, Holahan JM, Marchione KE, Michaels R, Shaywitz SE. 2015. Achievement Gap in Reading Is Present as Early as First Grade and Persists through Adolescence. Journal of Pediatrics. 167(5):1121–1125.e2. doi:10.1016/j.jpeds.2015.07.045. http://dx.doi.org/10.1016/j.jpeds.2015.07.045.

Forkel SJ, Thiebaut de Schotten M, Kawadler JM, Dell'Acqua F, Danek A, Catani M. 2014. The anatomy of fronto-occipital connections from early blunt dissections to contemporary

tractography. Cortex. 56(iii):73–84. doi:10.1016/j.cortex.2012.09.005. http://dx.doi.org/10.1016/j.cortex.2012.09.005.

Galuschka K, Ise E, Krick K, Schulte-Körne G. 2014. Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. PLoS ONE. 9(2). doi:10.1371/journal.pone.0089900. http://arxiv.org/abs/e89900.

Gebauer D, Fink A, Filippini N, Johansen-Berg H, Reishofer G, Koschutnig K, Kargl R, Purgstaller C, Fazekas F, Enzinger C. 2012. Differences in integrity of white matter and changes with training in spelling impaired children : a diffusion tensor imaging study. Brain Structure and Function. 217(3):747–760. doi:10.1007/s00429-011-0371-4.Differences.

Glatz T, Vanderauwera J, Vanden Bempt F, Economou M, Wouters J, Vandermosten M, Ghesquière P. 2021. Design of GraphoGame-Flemish. doi:10.17605/OSF.IO/CYKED.

Grills-Taquechel AE, Fletcher JM, Vaughn SR, Stuebing KK. 2012. Anxiety and reading difficulties in early elementary school: evidence for unidirectional- or bi-directional relations? Child psychiatry and human development. 43(1):35–47. doi:10.1007/s10578-011-0246-1.

Gullick MM, Booth JR. 2014. Individual Differences in Crossmodal Brain Activity Predict Arcuate Fasciculus Connectivity in Developing Readers. Journal of Cognitive Neuroscience. 26(7):1331–1346. doi:10.1162/jocn.

Hatcher PJ, Hulme C, Snowling MJ. 2004. Explicit phoneme training with phonic reading instructions helps young children at risk. Journal of Child Psychology and Psychiatry. 43:338–358. doi:10.1111/j.1469-7610.2004.00225.x.

Hoeft F, McCandliss BD, Black JM, Gantman A, Zakerani N, Hulme C, Lyytinen H, Whitfield-Gabrieli S, Glover GH, Reiss AL, et al. 2011. Neural systems predicting long-term outcome in dyslexia. Proceedings of the National Academy of Sciences. 108(1):361–366. doi:10.1073/pnas.1008950108. http://www.pnas.org/cgi/doi/10.1073/pnas.1008950108.

Huber E, Donnelly PM, Rokem A, Yeatman JD. 2018. Rapid and widespread white matter plasticity during an intensive reading intervention. Nature Communications. 9(1):1–13. doi:10.1038/s41467-018-04627-5. http://dx.doi.org/10.1038/s41467-018-04627-5.

Huber E, Mezer A, Yeatman JD. 2021. Neurobiological underpinnings of rapid white matter plasticity during intensive reading instruction. NeuroImage. 243(August):118453. doi:10.1016/j.neuroimage.2021.118453. https://doi.org/10.1016/j.neuroimage.2021.118453.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. FSL. NeuroImage. 62:782–790. doi:10.1016/j.neuroimage.2011.09.015. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki.

Johnson RT, Yeatman JD, Wandell BA, Buonocore MH, Amaral DG, Nordahl CW. 2014. Diffusion properties of major white matter tracts in young, typically developing children. Neuroimage. 88:143–154. doi:10.1016/j.neuroimage.2013.11.025.Diffusion.

Jolles D, Wassermann D, Chokhani R, Richardson J, Tenison C, Bammer R, Fuchs L, Supekar K, Menon V. 2016. Plasticity of left perisylvian white-matter tracts is associated with individual

differences in math learning. Brain Structure and Function. 221(3):1337–1351. doi:10.1007/s00429-014-0975-6. http://dx.doi.org/10.1007/s00429-014-0975-6.

Jossinger S, Kronfeld V, Avital D, Ofer Z, Michal A, Shachar B. 2021. Speech rate association with cerebellar white - matter diffusivity in adults with persistent developmental stuttering. Brain Structure and Function.(0123456789). doi:10.1007/s00429-020-02210-7. https://doi.org/10.1007/s00429-020-02210-7.

Keller TA, Just MA. 2009. Altering Cortical Connectivity: Remediation-Induced Changes in the White Matter of Poor Readers. Neuron. 64(5):624–631. doi:10.1016/j.neuron.2009.10.018. http://dx.doi.org/10.1016/j.neuron.2009.10.018.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software. 82(13):1–26. doi:10.18637/jss.v082.i13.

Lakens D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology. 4(NOV):1–12. doi:10.3389/fpsyg.2013.00863.

Langer N, Peysakhovich B, Zuk J, Drottar M, Sliva DD, Smith S, Becker BLC, Grant PE, Gaab N. 2017. White Matter Alterations in Infants at Risk for Developmental Dyslexia. Cerebral Cortex.(February):bhv281. doi:10.1093/cercor/bhv281. http://cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/bhv281.

Lebel C, Beaulieu C. 2009. Lateralization of the arcuate fasciculus from childhood to adulthood and its relation to cognitive abilities in children. Human Brain Mapping. 30(11):3563–3573. doi:10.1002/hbm.20779.

Lebel C, Shaywitz B, Holahan J, Shaywitz S, Marchione K, Beaulieu C. 2013. Diffusion tensor imaging correlates of reading ability in dysfluent and non-impaired readers. Brain and Language. 125(2):215–222. doi:10.1016/j.bandl.2012.10.009. http://dx.doi.org/10.1016/j.bandl.2012.10.009.

Lonigan CJ, Purpura DJ, Wilson SB, Walker PM, Clancy-Menchetti J. 2013. Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. Journal of Experimental Child Psychology. 114(1):111–130. doi:10.1016/j.jecp.2012.08.010. http://dx.doi.org/10.1016/j.jecp.2012.08.010.

Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping. 15(1):1–25. doi:10.1002/hbm.1058.

Ozernov-Palchik O, Gaab N. 2016. Tackling the 'dyslexia paradox': Reading brain and behavior for early markers of developmental dyslexia. Wiley Interdisciplinary Reviews: Cognitive Science. 7(2):156–176. doi:10.1002/wcs.1383. http://arxiv.org/abs/15334406.

Partanen M, Kim DHC, Rauscher A, Siegel LS, Giaschi DE. 2020. White matter but not grey matter predicts change in reading skills after intervention. Dyslexia.(April):1–21. doi:10.1002/dys.1668.

Peterson RL, Pennington BF. 2015. Developmental Dyslexia. Annual Review of Clinical Psychology. 11(1):283–307. doi:10.1146/annurev-clinpsy-032814-112842. http://www.annualreviews.org/doi/10.1146/annurev-clinpsy-032814-112842.

Phan TV, Sima D, Smeets D, Wouters J, Vandermosten M. 2021. Structural brain dynamics across reading development : A longitudinal MRI study from kindergarten to grade 5. Human Brain Mapping.(January):1–13. doi:10.1002/hbm.25560.

Pugh KR, Mencl WE, Jenner AR, Katz L, Frost SJ, Lee JR, Shaywitz SE, Shaywitz BA. 2001. Neurobiological studies of reading and reading disability. Journal of Communication Disorders. 34(6):479–492. doi:10.1016/S0021-9924(01)00060-0.

Puolakanaho A, Ahonen T, Aro M, Eklund K, Leppänen PHT, Poikkeus AM, Tolvanen A, Torppa M, Lyytinen H. 2007. Very early phonological and language skills: Estimating individual risk of reading disability. Journal of Child Psychology and Psychiatry and Allied Disciplines. 48(9):923–931. doi:10.1111/j.1469-7610.2007.01763.x.

Raschle NM, Lee M, Buechler R, Christodoulou JA, Chang M, Vakil M, Stering PL, Gaab N. 2009. Making MR imaging child's play - Pediatric neuroimaging protocol, guidelines and procedure. Journal of Visualized Experiments.(29):1–5. doi:10.3791/1309.

Raven JC, Court JH, Raven J. 1984. Manual for Raven's progressive matrices and vocabulary scales. London: Lewis.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reynolds JE, Grohs MN, Dewey D, Lebel C. 2019. Global and regional white matter development in early childhood. NeuroImage. 196(March):49–58. doi:10.1016/j.neuroimage.2019.04.004. https://doi.org/10.1016/j.neuroimage.2019.04.004.

Richardson U, Lyytinen H. 2014. THE GRAPHOGAME METHOD: THE THEORETICAL AND METHODOLOGICAL BACKGROUND OF THE TECHNOLOGY-ENHANCED LEARNING ENVIRONMENT FOR LEARNING TO READ. Human Technology - An Interdisciplinary Journal on Humans in ICT Environments. 10(1):39–60. doi:10.17011/ht/urn.201405281859.

Romeo RR, Christodoulou JA, Halverson KK, Murtagh J, Cyr AB, Schimmel C, Chang P, Hook PE, Gabrieli JDE. 2018. Socioeconomic status and reading disability: Neuroanatomy and plasticity in response to intervention. Cerebral Cortex. 28(7):2297–2312. doi: https://doi.org/10.1093/cercor/bhx131

Sagi Y, Tavor I, Hofstetter S, Tzur-Moryosef S, Blumenfeld-Katzir T, Assaf Y. 2012. Learning in the Fast Lane: New Insights into Neuroplasticity. Neuron. 73(6):1195–1203. doi:10.1016/j.neuron.2012.01.025. http://dx.doi.org/10.1016/j.neuron.2012.01.025.

Saur D, Kreher BW, Schnell S, Kümmerera D, Kellmeyera P, Vrya MS, Umarova R, Musso M, Glauche V, Abel S, et al. 2008. Ventral and dorsal pathways for language. Proceedings of the National Academy of Sciences of the United States of America. 105(46):18035–18040. doi:10.1073/pnas.0805234105.

Saygin ZM, Norton ES, Osher DE, Beach SD, Cyr AB, Ozernov-Palchik O, Yendiki A, Fischl B, Gaab N, Gabrieli JDE. 2013. Tracking the Roots of Reading Ability: White Matter Volume and Integrity Correlate with Phonological Awareness in Prereading and Early-Reading Kindergarten Children. Journal of Neuroscience. 33(33):13251–13258. doi:10.1523/JNEUROSCI.4383-12.2013. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4383-12.2013.

Schlaggar BL, McCandliss BD. 2007. Development of neural systems for reading. Annual Review of Neuroscience. 30:475–503. doi:10.1146/annurev.neuro.28.061604.135645.

Smidts DP, Oosterlaan J. 2007. GvK, Gedragsvragenlijst voor Kleuters: Handleiding. Amsterdam: Harcourt Test Publishers.

Smith SM. 2002. Fast robust automated brain extraction. Human Brain Mapping. 17(3):143–155. doi:10.1002/hbm.10062.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 23(SUPPL. 1):208–219. doi:10.1016/j.neuroimage.2004.07.051.

Snowling MJ, Hulme C. 2011. Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. British Journal of Educational Psychology. 81(1):1–23. doi:10.1111/j.2044-8279.2010.02014.x.

Snowling MJ, Hulme C. 2012. Interventions for children's language and literacy difficulties. International Journal of Language and Communication Disorders. 47(1):27–34. doi:10.1111/j.1460-6984.2011.00081.x.

Stanovich KE. 1986. Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. Reading Research Quarterly. 21(4):360–407. doi:10.1598/rrq.21.4.1.

Sullivan GM, Feinn R. 2012. Using Effect Size—or Why the P Value Is Not Enough. Journal of Graduate Medical Education. 4(3):279–282. doi:10.4300/jgme-d-12-00156.1.

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.

Travis KE, Adams JN, Kovachy VN, Ben-Shachar M, Feldman HM. 2017. White matter properties differ in 6-year old Readers and Pre-readers. Brain Structure and Function. 222(4):1685–1703. doi:10.1007/s00429-016-1302-1.

Turesky TK, Vanderauwera J, Gaab N. 2021. Imaging the rapidly developing brain: Current challenges for MRI studies in the first five years of life. Developmental Cognitive Neuroscience. 47:100893. doi:10.1016/j.dcn.2020.100893. https://doi.org/10.1016/j.dcn.2020.100893.

Valas H. 1999. Students with learning disabilities and low-achieving students: Peer acceptance, loneliness, self-esteem, and depression. Social Psychology of Education. 3(3):173–192. doi:10.1023/A:1009626828789.

Vanden Bempt F, Economou M, Van Herck S, Vanderauwera J, Glatz T, Vandermosten M, Wouters J, Ghesquière P. 2021. Digital Game-Based Phonics Instruction Promotes Print Knowledge in Pre-Readers at Cognitive Risk for Dyslexia. Frontiers in Psychology. 12(September):1–18. doi:10.3389/fpsyg.2021.720548.

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. Human Brain Mapping. 36(8):3273–3287. doi:10.1002/hbm.22848.

Vanderauwera J, Vos AD, Forkel SJ, Catani M, Wouters J, Vandermosten M, Ghesquière P. 2018. Neural organization of ventral white matter tracts parallels the initial steps of reading development : A DTI tractography study. Brain and Language. 183(February):32–40. doi:10.1016/j.bandl.2018.05.007. https://doi.org/10.1016/j.bandl.2018.05.007.

Vanderauwera J, Wouters J, Vandermosten M, Ghesquière P. 2017. Early dynamics of white matter deficits in children developing dyslexia. Developmental Cognitive Neuroscience. 27(June):69–77. doi:10.1016/j.dcn.2017.08.003. https://doi.org/10.1016/j.dcn.2017.08.003.

Van Der Auwera S, Vandermosten M, Wouters J, Ghesquière P, Vanderauwera J. 2021. A three-time point longitudinal investigation of the arcuate fasciculus throughout reading acquisition in children developing dyslexia. NeuroImage. 237(October 2020):118087. doi:10.1016/j.neuroimage.2021.118087. https://doi.org/10.1016/j.neuroimage.2021.118087.

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. Neuroscience and Biobehavioral Reviews. 36(6):1532–1552. doi:10.1016/j.neubiorev.2012.04.002. http://dx.doi.org/10.1016/j.neubiorev.2012.04.002.

Vandermosten M, Hoeft F, Norton ES. 2016. Integrating MRI brain imaging studies of pre-reading children with current theories of developmental dyslexia: A review and quantitative meta-analysis. Current Opinion in Behavioral Sciences. 10:155–161. doi:10.1016/j.cobeha.2016.06.007. http://dx.doi.org/10.1016/j.cobeha.2016.06.007.

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. Developmental Cognitive Neuroscience. 14:8–15. doi:10.1016/j.dcn.2015.05.006. http://dx.doi.org/10.1016/j.dcn.2015.05.006.

Van Herck S, Economou M, Vanden Bempt F, Vanderauwera J, Glatz T, Dieudonne B, Vandermosten M, Ghesquière P, Wouters J. 2021. Ahead of Maturation: Enhanced Speech Envelope Training Boosts Rise Time Discrimination in Pre-Readers at Cognitive Risk for Dyslexia. doi:doi.org/10.1101/2021.04.12.439411.

Verhoeven L, Voeten M, Setten E van, Segers E. 2020. Computer-supported early literacy intervention effects in preschool and kindergarten: A meta-analysis. Educational Research Review. 30(January). doi:10.1016/j.edurev.2020.100325.

Verwimp C, Vanden Bempt F, Kellens S, Economou M, Vandermosten M, Wouters J, Ghesquière P, Vanderauwera J. 2020. Pre-literacy heterogeneity in Dutch-speaking kindergartners: latent profile analysis. Annals of Dyslexia. doi:10.1007/s11881-020-00207-9.

Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, et al. 2007. Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage. 36(3):630–644. doi:10.1016/j.neuroimage.2007.02.049. http://dx.doi.org/10.1016/j.neuroimage.2007.02.049.

Walton M, Dewey D, Lebel C. 2018. Brain white matter structure and language ability in preschool-aged children. Brain and Language. 176(September 2016):19–25. doi:10.1016/j.bandl.2017.10.008. https://doi.org/10.1016/j.bandl.2017.10.008.

Wandell BA, Yeatman JD. 2013. Biological development of reading circuits. Current Opinion in Neurobiology. 23(2):261–268. doi:10.1038/jid.2014.371. http://arxiv.org/abs/NIHMS150003.

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

Wanzek J, Vaughn S. 2007. Research-based implications from extensive early reading interventions. School Psychology Review. 36(4):541–561.

Yeatman JD, Dougherty RF, Ben-Shachar M, Wandell BA. 2012a. Development of white matter and reading skills. Proceedings of the National Academy of Sciences of the United States of America. 109(44). doi:10.1073/pnas.1206792109.

Yeatman JD, Dougherty RF, Myall NJ, Wandell BA, Feldman HM. 2012b. Tract Profiles of White Matter Properties: Automating Fiber-Tract Quantification. PLoS ONE. 7(11). doi:10.1371/journal.pone.0049790.

Yeatman J, Dougherty RF, Rykhlevskaia E, Sherbondy A, Deutsch GK, Wandell BA, Ben-Shachar M. 2011. Anatomical Properties of the Arcuate Fasciculus Predict Phonological and Reading Skills in Children. Journal of Cognitive Neuroscience. 23(11):3304–3317. doi:10.1162/jocn. http://arxiv.org/abs/1511.04103.

Yeatman JD, Rauschecker AM, Wandell BA. 2013. Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. Brain Lang. 125(2):146–55. doi:10.1016/j.bandl.2012.04.010.Anatomy. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf.

Yendiki A, Koldewyn K, Kakunoori S, Kanwisher N, Fischl B. 2014. Spurious group differences due to head motion in a diffusion MRI study. NeuroImage.(88):79–90. doi:10.1016/j.neuroimage.2013.11.027.Spurious. http://arxiv.org/abs/15334406.

Zatorre RJ, Fields RD, Johansen-Berg H. 2012. Plasticity in gray and white: Neuroimaging changes in brain structure during learning. Nature Neuroscience. 15(4):528–536. doi:10.1038/nn.3045. http://dx.doi.org/10.1038/nn.3045.

Zhang M, Chen C, Xue G, Lu Z lin, Mei L, Xue H, Wei M, He Q, Li J, Dong Q. 2014. Language-general and -specific white matter microstructural bases for reading. NeuroImage. 98:435–441. doi:10.1016/j.neuroimage.2014.04.080. http://dx.doi.org/10.1016/j.neuroimage.2014.04.080.

Zuk J, Dunstan J, Norton E, Yu X, Ozernov-Palchik O, Wang Y, Hogan TP, Gabrieli JDE, Gaab N. 2020. Multifactorial pathways facilitate resilience among kindergarteners at risk for dyslexia: A longitudinal behavioral and neuroimaging study. Developmental Science.:1–18. doi:10.1111/desc.12983.

Figure captions

Figure 1 Adapted CONSORT flow diagram.

Figure 2 Tract FA profiles before the intervention. Each plot shows the mean tract FA profile ± 1 standard error of the mean for the two groups (at-risk and typical control). Bold black panel borders indicate the tracts where global tract FA differences were found. Localized risk-related differences are indicated by grey-shaded regions on the tract profile. The red arrows indicate the position of the corresponding anatomical locations, which are visualized in black on three-dimensional tract renderings next to each plot.