Neuroanatomy of dyslexia: An allometric approach

Hugo Peyre[1,2,3*], Neha Mohanpuria[1], Katarzyna Jednoróg[4], Stefan Heim[5,6], Marion Grande[7], Muna van Ermingen-Marbach[5,7], Irene Altarelli[8], Karla Monzalvo[9], Camille Williams[1], David Germanaud[2,9,10,11], Roberto Toro[12], Franck Ramus[1]

1. Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Ecole Normale Supérieure, PSL University, Paris, France.
2. Neurodiderot. INSERM UMR 1141, Paris Diderot University, Paris, France.
3. Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France.
4. Laboratory of Language Neurobiology, Nencki Institute of Experimental Biology, Polish Academy of Sciences (PAS), Warsaw, Poland.
5. Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Germany.
6. Institute of Neuroscience and Medicine (INM-1), Forschungszentrum Jülich, Helmholtz-Gemeinschaft Deutscher Forschungszentren (HZ), Germany.
7. Department of Neurology, Medical Faculty, RWTH Aachen University, Germany.
8. Laboratory for the Psychology of Child Development and Education, CNRS UMR 8240, Université de Paris, Paris, France.
9. INSERM, UMR992, CEA, NeuroSpin Center, University Paris Saclay, Gif-sur-Yvette, France.
10. Department of Pediatric Neurology and Metabolic Diseases, Robert Debré Hospital, APHP, Paris, France.
11. INSERM, Sorbonne Paris Cité University (USPC), CEA, UMR 1129, Paris, France.
12. Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France.

Running title: Neuroanatomy of dyslexia.

*CORRESPONDING AUTHOR: Hugo Peyre, MD, PhD: LSCP, Département d'Etudes Cognitives. Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. Email: peyrehugo@yahoo.fr. Phone: 0033(0)144322619. Fax: 0033(0)144322619

ABSTRACT

Despite evidence for a difference in total brain volume between dyslexic and good readers, no previous neuroimaging study examined differences in allometric scaling (*i.e.* differences in the relationship between regional and total brain volumes) between dyslexic and good readers. The present study aims to fill this gap by testing differences in allometric scaling and lobar brain volume differences in dyslexic and good readers.

Object-based morphometry analysis was used to determine grey and white matter volumes of the 4 lobes, the cerebellum, and limbic structures in 130 dyslexic and 106 good readers aged 8 to 14 years. Data were collected across three countries (France, Poland, and Germany). Three methodological approaches were used: Principal Components Analysis (PCA), linear regression, and Multiple Group Confirmatory Factor Analysis (MGCFA).

Difference in total brain volume between good and dyslexic readers was Cohen's d=0.39. We found no difference in allometric scaling, nor in regional brain volume between dyslexic and good readers. Results of our three methodological approaches (PCA, linear regression and MGCFA) were consistent.

This study provides evidence for total brain volume differences between dyslexic and control children, but no evidence for differences in the volumes of the four lobes, the cerebellum or limbic structures, once allometry is properly taken into account. It also finds no evidence for a difference in allometric relationships between the groups. We highlight the methodological interest of the MGCFA approach to investigate such research issues.

Keywords: Dyslexia; Allometry; Brain volume; Sex.

List of abbreviations:

Magnetic Resonance Imaging (MRI)

Voxel-Based Morphometry (VBM)

Object-Based Morphometry (OBM)

Principal Components Analysis (PCA)

Multiple Group Confirmatory Factor Analysis (MGCFA)

Intelligence Quotient (IQ)

Wechsler Intelligence Scale for Children (WISC)

Attention Hyperactivity Deficit Disorder (ADHD)

INTRODUCTION

Dyslexia is characterized by persistent difficulties in learning the written language code that cannot be accounted for by another disorder and by lack of education or sensory deficits (American Psychiatric Association, 2013). Like other neurodevelopmental disorders, the etiology of dyslexia involves complex interactions between multiple genetic and environmental risk factors (Oliver & Plomin, 2007; Mascheretti *et al.*, 2013; Bishop, 2015). At the cognitive level, deficits in phonological processes are thought to be central to the development of dyslexia in most cases (Ramus *et al.*, 2003; Saksida *et al.*, 2016).

There is a vast literature on the neuroanatomical correlates of dyslexia, most of them relying on Voxel-Based Morphometry (VBM) and reporting regional brain volume differences between dyslexic and good readers. However, meta-analyses of those studies reported remarkably few consistent findings after proper statistical corrections (Linkersdörfer *et al.*, 2012; Richlan *et al.*, 2013; Eckert *et al.*, 2016). Indeed, the validity of these results have been questioned due to several methodological limitations (Ramus *et al.*, 2018). The problem of multiple testing is particularly acute in cluster-based VBM studies with small samples, as it is the case for most studies on dyslexia. Comparatively, neuroanatomical studies using predefined segmentation, which reduce the number of multiple comparisons (known as Object-Based Morphometry [OBM] (Mangin *et al.*, 2004)) of tissues and lobes are less concerned by false positive results (Scarpazza *et al.*, 2013). Ramus *et al.* (2018) argued that the only robust result emerging from this literature was a smaller total brain volume in dyslexics compared to good readers (Cohen's d=-0.58 [IC-95%: -0.32 ; -0.85]) (Ramus *et al.*, 2018). Yet, such global differences are only taken into account for the evaluation of regional differences in about half of the studies. This is particularly problematic when examining cerebral differences between groups with differing total brain volumes. For instance, previous findings suggest that sex differences and sex by age interactions in local brain volumes practically disappeared when taking into account brain size

(Jäncke *et al.*, 2015). Brain size should thus be considered when examining volumetric differences across brain regions, as it accounts for more interindividual differences than sex or age.

Given that dyslexic and normal readers are thought to differ in both total and regional brain volumes, the question now arises whether they also differ in the relationship between any given regional volume and total brain volume. This scaling relationship between a regional volume (y) and total brain volume (x) can be investigated with the commonly used power equation $y = b\,x^a$ (Finlay *et al.*, 2001). If $a = 1$, the volumes x and y are directly proportional (isometric). If $a \neq 1$, they are not (they are allometric). The allometric scaling coefficient "a" can be easily estimated by linear regression using a log-log transformation $\{\log(y) = a\log(x) + \log(b)\}$. When the regional volume grows disproportionately with total brain volume growth ($a > 1$), this is called positive allometry or hyperallometry. When the regional volume grows more slowly than total brain volume ($a < 1$), this is called negative allometry or hypoallometry. In many cases, "a" has been shown to differ from 1, non-linear relationships being the rule more than the exception between regional and total brain volumes (Jäncke *et al.*, 2015; de Jong *et al.*, 2017; Reardon *et al.*, 2018). Therefore, using a linear or proportional adjustment of total brain volume when comparing regional volumes is theoretically inappropriate and may be misleading when linear approximation is too crude. Indeed, recent studies have shown that omitting brain allometry can lead to overestimating or underestimating regional volumetric group differences and recommend that studies adjust for total brain volume differences using allometric scaling (Germanaud *et al.*, 2014; Reardon *et al.*, 2016; de Jong *et al.*, 2017; Mankiw *et al.*, 2017; Jäncke *et al.*, 2019; Sanchis-Segura *et al.*, 2019).

In the present study, we examined differences in allometric scaling and regional brain volume differences in dyslexic and good readers. Our analyses focused on 24 regional brain volumes (2 hemispheres x 6 regions (frontal, temporal, parietal, cerebellum, limbic, and occipital) x 2

tissue compartments (grey and white matter)) of 130 dyslexic and 106 good readers. Data was collected across three countries (France, Poland, and Germany). Since several sex differences in clinical and neuroanatomical characteristics of dyslexic readers have previously been reported (Altarelli *et al.*, 2014; Evans *et al.*, 2014; Arnett *et al.*, 2017), our analyses were performed on the whole sample as well as in boys and girls separately. We aimed to address the following research questions: Do allometric scaling and regional brain volumes differ between dyslexic and good readers? Do these observed differences (if any) between dyslexic and good readers depend on sex? We had no specific *a priori* hypotheses with regards to these research questions.

Three methodological approaches were applied to address our research questions (Principal Components Analysis (PCA), linear regression, and Multiple Group Confirmatory Factor Analysis (MGCFA)). In theory, MGCFA is more advantageous than PCA since it tests regional allometric scaling group differences as well as global and regional volumetric group differences (Jolicoeur, 1963; Toro *et al.*, 2009) and, unlike linear regression, MGCFA also considers the mutual relationship between regional brain structures and tests global group differences in allometric scaling (Toro *et al.*, 2009; de Jong *et al.*, 2017). However, since MGCFA is rarely conducted in this literature, the present study compared the results of the three approaches to determine whether the results of the less commonly used MGCFA were consistent with the results of the classical linear regression and PCA approaches.

METHODS

PARTICIPANTS

Dyslexic and good readers were recruited in three countries (France, Poland, and Germany). Reading accuracy and speed were assessed using different language-appropriate standardized reading tests (see (Jednoróg *et al.*, 2015; Płoński *et al.*, 2017) for details). Participants came

from diverse social backgrounds and had at least one and a half year of formal reading instruction to differentiate serious problems in reading acquisition from early delays that are not always persistent. They were recruited based on the following criteria: age was between 8.5 and 13.7 years old, Intelligence Quotient (IQ) higher than 85, or an age-appropriate scaled score of at least 7 on Wechsler Intelligence Scale for Children (WISC) Block Design and 6 on WISC Similarities, no spoken language disorders, no formal diagnosis of Attention Hyperactivity Deficit Disorder (ADHD), and no reported hearing, sight, or other neurological problems.

Dyslexic readers were either identified in school, through clinics or were specifically requesting clinical assessment of their reading problems. Most of the studied children already had a clinical diagnosis of dyslexia and all were screened for inattention/hyperactivity symptoms and language disorders. The inclusion criterion for dyslexic readers was defined as more than 1.5 SD below grade level on different appropriate standardized tests of reading, whereas good readers were less than 0.85 SD below grade level.

All studies were approved by local ethics committees (CPP Bicêtre in France; Medical University of Warsaw in Poland; Uniklinik RWTH Aachen in Germany) in compliance with the Code of Ethics of the World Medical Association—Declaration of Helsinki. The children and their parents gave informed written consent to participate in the study.

Together, 130 dyslexic (56 girls) and 106 good readers (55 girls) were included in the study.

IMAGING PROCEDURE

High-resolution T1w images were acquired in five different studies:

French group (studies 1 & 2)

Whole brain T1w images were acquired for the total sample on the same 3 Tesla (3T) Siemens Trio Tim MRI platform.

Study 1 (13 good and 11 dyslexic readers): The MRI had a 12-channels head coil with the following parameters: acquisition matrix: 256 x 256 x 176, TR=2,300 ms, TE=4.18 ms, flip angle=9°, FOV=256 mm, voxel size: 1 x 1 x 1 mm.

Study 2 (32 good and 28 dyslexic readers): MRI platform used a 32-channels head coil with the following parameters: acquisition matrix=230 x 230 x 202, TR=2,300 ms, TE=3.05 ms, flip angle=9°, FOV=230 mm, voxel size=0.9 x 0.9 x 0.9 mm.

German group (studies 3 & 4)

Study 3 (10 good and 35 dyslexic readers): Whole brain images were acquired on a 3T Siemens Trio Tim scanner using a standard birdcage head coil. T1w images had the following specifications: acquisition matrix: 256 x 256 x 176, TR=1,900 ms, TE=2.52 ms, flip angle=9°, FOV=256 mm, voxel size: 1 x 1 x 1 mm.

Study 4 (16 good and 10 dyslexic readers): Whole brain images were acquired on a 1.5T Siemens Avanto scanner using a standard bird-cage head coil with the following parameters: acquisition matrix: 256 x 256 x 170, TR=2,200 ms, TE=3.93 ms, flip angle=15°, FOV=256 mm, voxel size: 1 x 1 x 1 mm.

Polish group (Study 5)

Study 5 (35 good and 46 dyslexic readers): Whole brain images were acquired for the total sample on a 1.5T Siemens Avanto platform equipped with 32-channels phased array head coil. T1w images had the following specifications: acquisition matrix: 256 x 256 x 192, TR=1,720 ms, TE=2.92 ms, flip angle=9°, FOV=256, voxel size 1 x 1 x 1 mm.

MORPHOMETRIC ANALYSIS

Image processing and analyses were carried out using SPM8 (http://ww.fil.ion.ucl.ac.uk/spm/) run in MATLAB7.11 (Mathworks, Sherborn, MA). T1w images were automatically segmented into different tissue classes [grey matter, white matter, and nonbrain (cerebrospinal fluid,

skull)], using the "New Segmentation" option in SPM8 (Ashburner & Friston, 2005). Tissue probability maps were taken from a customized paediatric brain generated using Template-O-Matic toolbox (http://dbm.neuro.uni-jena.de/software/tom/). The Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) algorithm was then used to create a study-specific template (Ashburner, 2007; Marchewka *et al.*, 2014). This step was followed by affine registration of the GM maps to the Montreal Neurological Institute (MNI) space scaling the GM probability values with the Jacobian determinants to ensure that the total signal in each tissue class remained constant (*i.e.* "modulation") (Ashburner & Friston, 2000). Binary masks for the main lobes (frontal, temporal, parietal, cerebellum, limbic and occipital) were derived from a cerebral lobe atlas defined in the MNI template space and published by Fonov *et al*. (Fonov *et al.*, 2011).

STATISTICAL ANALYSIS

Three different methodological approaches were used to investigate differences in allometric scaling and regional brain volume differences in dyslexic and good readers. The project was preregistered on OSF (https://osf.io/t2v5h/). All brain volumes were log-transformed because the relation between a regional volume and the total volume is not linear (de Jong *et al.*, 2017).

ANOVA

First, we performed an ANOVA to examine the main effects of group (dyslexic and good readers), sex (girls and boys), hemisphere (left and right), tissue (grey matter and white matter), brain regions (frontal, temporal, parietal, cerebellum, limbic and occipital) and the interactions between group, sex and the other variables (hemisphere, tissue, lobe). Group and sex were between-subject variables and hemisphere, tissue, and lobe were within-subjects variables.

Linear Regression Models

Each regional brain volume was tested in separate models for differences in allometric scaling between groups (see Equation 1) (Mankiw *et al.*, 2017).

Equation 1: $log(RBV) = \beta_0 + \beta_1\ Group + \beta_2\ log(TBV) + \beta_3\ Group \times log(TBV)$.

RBV=Regional Brain Volume (e.g. right grey matter of the frontal lobe) and TBV=Total Brain Volume.

Linear regression models were performed using SAS 9.2 software (SAS Institute, Cary, NC). To reduce type 1 error inflation due to multiple testing, we set the alpha threshold at 0.002 [0.05/24, 24 being the number of observed brain regions]. We also tested the interaction between sex and group (dyslexic and good readers) in linear regression models. Moreover, we performed linear regression models in dyslexic *versus* good readers in girls and boys separately, to follow the same analytic plan than for the MGCFA.

Principal Component Analysis (PCA)

The second method to investigate differences in allometric scaling between groups is PCA. A PCA of the log-transformed regional brain volumes was performed separately in the two groups. The loadings of the first principal components of the groups are considered as the coordinates of two vectors (Krzanowski, 1979). The angle between these two vectors corresponds to a global test of the differences in allometric scaling between these groups (Jolicoeur, 1963). PCA and permutation tests (10,000 iterations) of the angle between vectors were performed by running the *psych* package in R software (R Foundation for Statistical Computing, n.d.). The statistical significance of the difference in angle was estimated by comparing it with a null distribution obtained from 10,000 random permutations. In each of these permutations, two groups of subjects were created with sample sizes corresponding to the number of dyslexic and control participants, independently of their diagnostic status. A p-value

was estimated by counting the proportion of random permutations where the angle difference was more extreme than the one observed in the original groups (Toro *et al.*, 2009). If the proportion is smaller than alpha=0.05, then there is a significant group difference and follow up analyses were conducted with a post-hoc examination of the individual exponents. Differences in allometric scaling in dyslexic *versus* good readers in girls and boys were compared separately and the interaction between sex and group (dyslexic and good readers) was tested.

Multiple Group Confirmatory Factor Analysis (MGCFA)

MGCFA is a popular method for measurement invariance which goes beyond the main limitations of linear regression and PCA. It is an extension of Confirmatory Factor Analysis (CFA) (Davidov *et al.*, 2018). In CFA, observed variables are considered to be indicators of an unobserved (latent) variable (Davidov *et al.*, 2014). Systems of equations were thus used to describe the relationship between the observed variables (local regional volumes) and the latent concept (total brain volume) that these variables are supposed to measure. The observed variables $log(RBV)_i$ were modelled as linear functions of a latent variable ξ (see the following equation 2). The τ_i, λ_i and δ_i refer, respectively, to the intercept, the slope (factor loading) and the error term in these functions. The superscript (g) indicates the group (in our case: dyslexic and good readers).

Equation 2: $x_i^{(g)} = \tau_i^{(g)} + \lambda_i^{(g)}\xi^{(g)} + \delta_i^{(g)}$

The MGCFA starts with the determination of a well-fitting multi-group baseline model, and continues by testing, in a hierarchical fashion, the metric invariance *(i.e.* equality of factor loadings) and the scalar invariance *(i.e.* equality of intercepts) between groups. To do this, some parameters are constrained to be the same across groups and this model is compared to a model that is unconstrained on these parameters, in terms of fit, by computing a $\chi 2$ difference test.

Metric invariance is tested by constraining the loadings of the factor across groups (intercept being unconstrained) (Steenkamp & Baumgartner, 1998). If a global test (comparison of the fit of a model with all loadings constrained to be equal across groups to an unconstrained model) indicates that the metric invariance hypothesis is rejected, then each regional brain volume is examined one after the other (comparison of the fit of a model with all loadings constrained to be equal across groups with that of a model with the loading of one regional brain volume unconstrained). Similarly, scalar invariance is tested by constraining the intercepts across groups and comparing this model with an unconstrained model (loadings are also constrained if metric invariance was found in the previous step) (Steenkamp & Baumgartner, 1998; Hong *et al.*, 2003; Davidov *et al.*, 2014). Again, if this global test indicates that the scalar invariance hypothesis is rejected, then each regional brain volume is examined one after the other (comparison of the fit of a model with all intercepts constrained to be equal across groups to a model with the intercept of one regional brain volume unconstrained).

MGCFA models were applied using procedures implemented in *Mplus* (Schnabel *et al.*, 2015). The maximum likelihood estimation with robust standard errors (MLMV) estimator was used throughout for the estimation of parameters (Asparouhov & Muthén, 2010). MGCFA models were also performed on four groups: dyslexic boys, dyslexic girls, non-dyslexic boys and non-dyslexic girls. We additionally compared dyslexic *versus* good readers in girls and boys separately and in non-dyslexic girls & dyslexic boys *versus* dyslexic girls & non-dyslexic boys to identify an interaction between sex and group (dyslexic and good readers). The sample sizes were sufficiently large to perform MGCFA on these subgroups ($n > 200$) (Mundfrom *et al.*, 2005).

In order to evaluate the concordance between linear regression and MGCFA methods, we estimated, using both methods, the slope and intercept differences between groups in the 24 regional brain volumes and then computed a Spearman rank order correlation.

RESULTS

ANOVA

The analysis of variance showed a main effect of group ($F(1, 5400)=10.5$, p-value=0.001, $\eta^2=0.1\%$; good > dyslexic readers), sex ($F(1, 5400)=45.6$, p-value<0.001, $\eta^2=0.4\%$; boys > girls), hemisphere ($F(1, 5400)=5.5$, p-value=0.019, $\eta^2<0.1\%$; left > right hemisphere), tissue ($F(1, 5400)=27832.7$, p-value<0.001, $\eta^2=31.6\%$; grey > white matter) and brain regions ($F(5, 5400)=10542.5$, p-value <0.001, $\eta^2=59.8\%$; frontal > temporal > parietal > occipital > cerebellum > limbic region). There was no significant interaction between group and sex and hemisphere, tissue and brain regions.

Linear regression

In linear regression models, we found no difference in allometric scaling (**Table 1**) nor in regional brain volume (**Table 2**) between dyslexic and good readers, even in the subsamples of boys and girls or when the interaction between sex and group was considered. In **Suppl. Figure 1**, we present, as an example, the relationship between the left temporal lobe grey matter and the total brain volumes in dyslexic and good readers. Together, these results indicate no differences in allometric scaling nor regional brain volumes between dyslexic and good readers.

PCA

In PCA models, we found no differences in allometric scaling (**Table 3**) between dyslexic and good readers, even in the subsamples of boys and girls or when interaction between sex and group were considered.

MGCFA

Group differences were reported with Cohen's d and uncorrected p-values. Differences in total brain volume between good and readers were 0.39 in the whole sample, 0.34 in girls, and 0.43 in boys. Differences in total brain volume between girls and boys were -0.88 in the whole sample, -0.94 in good readers, and -0.72 in dyslexic readers. No differences between dyslexic and good readers were found with the test of equality of all loadings (metric invariance), in the entire sample and each subgroup (dyslexic boys, dyslexic girls, non-dyslexic boys and non-dyslexic girls; **Table 4**). No differences between dyslexic and good readers were reported from the test of equality of all intercepts (scalar invariance) in the entire sample. Interestingly, the test of equality of all intercepts in the four groups was significant (p-value=0.004; **Table 5**). Further analyses revealed scalar invariance in boys (good readers and dyslexic readers; p-value=0.3), girls (p-value=0.7), and when interaction between sex and group were considered (p-value=0.5). A supplementary test was performed to confirm that scalar invariance between the four groups was rejected because of sex (p-value<0.001). These results indicate sex differences in regional brain volume differences between girls and boys, after proper adjustment (in log-log scale) of total brain volume differences (see **Figure 1**). Girls had larger frontal grey and white matter (d=0.1) and smaller cerebellar grey matter (d=-0.2) than boys, relative to total brain volume.

Comparison of the three methodological approaches

Results of the three methodological approaches (linear regression, PCA and MGCFA) were largely consistent. The ranking of Δ slopes (good readers – dyslexic readers) and Δ intercepts (good readers - dyslexic readers) between linear regression and MGCFA methods in the 24 regional brain volumes using Spearman's correlation coefficients were respectively 0.96 and 0.97. Moreover, the PCA method was concordant with the two other approaches since it did not

detect differences in allometric scaling between dyslexic and good readers (even in the subsamples of boys and girls or when interaction between sex and group were considered).

DISCUSSION

In this study, we investigated differences in allometric scaling and regional brain volume between dyslexic and control children using three methodological approaches (linear regression, PCA and MGCFA). We replicated the well-established finding that total brain volume differs between dyslexic and good readers. Although we did not find differences in allometric scaling or regional brain volume between dyslexic and good readers, the present study highlights the methodological advantage of the MGCFA approach to investigate allometric scaling and volumetric differences between groups.

Since previous reports on the brain anatomy of dyslexia overlooked brain allometry, this study is the first to investigate allometric scaling differences in dyslexia. Such a study was warranted given that the differences in total brain volume between dyslexic and good readers are robustly established (Ramus *et al.*, 2018) and were replicated in the present study (Cohen's d=0.39; delta=28 cm^3). Omitting allometric scaling has been found to overestimate or underestimate some volumetric group differences (Reardon *et al.*, 2016; Mankiw *et al.*, 2017). Adjusting for total brain volume using allometry is thus crucial to reduce the odds of false positive or negative results. In light of the lack of allometric scaling group differences, our results support the idea that the brain of dyslexic readers follows the same structural organization as the typical brain, despite being slightly smaller. A smaller brain volume is clearly not specific to dyslexia, as it is also found in many but not all neurodevelopmental disorders. On the contrary, a larger brain volume is reported in neurodevelopmental disorders such as fragile X syndrome (Hazlett *et al.*, 2012) or in autism during the first 2–4 years of life (Redcay & Courchesne, 2005). Thus, the

interpretation of such global differences remains unclear. One prominent hypothesis was that global brain differences in dyslexia may stem from the correlation between total brain volume and IQ. and to a lower mean IQ in children with dyslexia. However, this hypothesis can be refuted considering that total group differences remain significant after adjusting for IQ in the present data (Ramus et al. 2018). Perhaps lower brain volume is a general risk factor for neurodevelopmental disorders or perhaps it is a secondary consequence of certain types of early regional disruption.

In line with the recent literature review by Ramus *et al.* (2018), we did not find regional brain volume differences between dyslexic and good readers at the "lobar level". Studies that reported differences in regional brain volumes between dyslexics and controls have mainly been conducted using a VBM approach in relatively small samples. However, in the present study, we conducted an OBM approach (*i.e.* predefined segmentation of tissues and lobes *versus* cluster-based in VBM) which is thought to reduce the rate of false positives (Smith & Nichols, 2009). Since several studies reported an increased gyrification of the brain of dyslexic readers (Im *et al.*, 2016; Płoński *et al.*, 2017; Williams *et al.*, 2018), other structural measures besides volume may be associated with dyslexia and could be further investigated using an OBM approach (for instance the folds as segmented by Morphologist (Fischer *et al.*, 2012)).

The present study additionally highlights the methodological advantage of the MGCFA approach to investigate allometric scaling and volumetric group differences by summarizing the benefits of MGCFA over the more frequently used methodological approaches (linear regression and PCA methods; **Table 6**).

Finally, we incidentally found that girls had larger frontal grey and white matter (d=0.1) and smaller cerebellar grey matter (d=-0.2) than boys (relatively to total brain volume; see **Figure 1**). which is largely inconsistent with results of adult neuroanatomical studies (Chen *et al.*, 2007; Ruigrok *et al.*, 2014; Ritchie *et al.*, 2018; Lotze *et al.*, 2019). To our knowledge, Lotze *et al.* were the only to report a larger prefrontal grey matter in women (Lotze *et al.*, 2019) and Chen *et al.* were the only to report smaller cerebellar hemispheres grey matter in men (Chen *et al.*, 2007). The study by Ritchie et al., which examined a much larger sample of 5216 UK Biobank participants did not found sex differences in these brain regions (as well as the meta-analysis by Ruigrok *et al.* (Ruigrok *et al.*, 2014)). Therefore, our incidental results regarding sex dimorphism in brain structure should be considered with great caution.

Strengths and limitations

The two major strengths of this study are the large sample of female and male dyslexic and good readers and the comparison of three methods (linear regression, PCA and MGCFA) used to examine volumetric group differences. However, the current study also has several limitations that must be considered when evaluating the results. First, total brain volume estimates were not identical across methods. While it remains a measured variable in the linear regressions, total brain volume is a latent variable estimated by the shared variance of lobar volumes in the MGCFA and corresponds to the first principal component in PCA, which reflects the most shared and unshared variance of the lobar volumes. While estimates may differ across methods, the consistency of our findings suggest that the MGCFA and PCA are nonetheless advantageous methods to investigate overall neuroanatomical group differences in future studies.

Second, our total sample gathered participants from 5 studies conducted in 3 different countries with different languages and different school systems, although good and dyslexic readers were

recruited in comparable proportions in each of these studies. Since the PCA and MGCFA cannot simultaneously examine the effect of group, sex, scanner site, and language, we were not able to correct for the non-independency of data collected per scanning site nor could we investigate how different languages influence the present findings (**Table 6**). Future large-scale studies should nevertheless investigate the impact of different cultures and languages to obtain a more precise estimate of volumetric group differences in dyslexia.

Third, in theory, the relationship between total brain volume and each lobar volume could correspond to a more complex equation than the power function we employ. However, there is an extensive literature on brain allometry, and the power equation is widely considered as a sufficiently good fit and is the current state of the art (Sanchis-Segura et al., 2019).

Finally, the brain region segmentations used in this study were quite coarse. It remains entirely possible that allometric scaling and regional brain volume differences between dyslexic and good readers might emerge when considering smaller brain regions (*e.g.* superior temporal gyrus of the left temporal lobe). Of course, the smaller the brain regions considered, the more numerous they are, thus the higher the risk of false positive results and the more stringent corrections for multiple tests should be. We suggest that MGCFA is a powerful approach to the study general group differences across a large number of brain regions since it allows for a global test of allometric differences, which does not necessitate correction for multiple comparisons. Our study therefore paves the way for more fine-grain investigations of regional brain volume differences in dyslexia, taking allometry into account.

Conclusions

This study provides further evidence that the brain of dyslexic readers has the same structural organization than a typical brain, at the "lobar" spatial resolution, despite being slightly smaller.

It also emphasizes the methodological advantages of the MGCFA approach to investigate differences in allometric scaling.

Acknowledgements

Contract grant sponsor: Agence Nationale de la Recherche; Contract grant number: ANR-06-NEURO-019-01, ANR-11-BSV4-014-01, ANR-17-EURE-0017, ANR-10-IDEX-0001-02 PSL; Contract grant sponsor: Ecole des Neurosciences de Paris, the Bettencourt-Schueller Foundation, Polish Ministry of Science and Higher Education; Contract grant numbers: IP2010 015170, IP2011 020271 (Ministry of Science and Higher Education) and 2011/03/D/HS6/05584, 2016/22/E/HS6/00119 (National Science Center) ; Contract grant sponsor: National Science Center; Contract grant number: DEC-2011/03/D/HS6/05584; Contract grant sponsor: German Bundesministerium für Bildung und Forschung; Contract grant number: BMBF 01GJ0613, 01GJ0614, 01GJ0804; Contract grant sponsor: Swiss Natonal Science Foundation; Contract grant number: 320030_135679.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author contributions

KJ, SH, MG, MvEM, KM and FR: Data collection.

HP, NM, CW, DG, RT and FR: Data analysis and interpretation.

HP, CW, DG, RT, FR: Drafting the article.

All authors: Critical revision of the article and final approval of the version to be published.

Ethical standards statement

All studies were approved by regional ethics committees (CPP Bicêtre in France; Medical University of Warsaw in Poland; Uniklinik RWTH Aachen in Germany) in compliance with the Code of Ethics of the World Medical Association—Declaration of Helsinki. The children and their parents gave informed written consent to participate in the study.

Data Accessibility

Anonymized data and details about preprocessing/analyses will be made available to colleagues if requested.

REFERENCES

Altarelli, I., Leroy, F., Monzalvo, K., Fluss, J., Billard, C., Dehaene-Lambertz, G., Galaburda, A.M., & Ramus, F. (2014) Planum temporale asymmetry in developmental dyslexia: Revisiting an old question. *Hum. Brain Mapp.*, **35**, 5717–5735.

American Psychiatric Association (2013) Draft DSM-V Criteria for ADHD.

Arnett, A.B., Pennington, B.F., Peterson, R.L., Willcutt, E.G., DeFries, J.C., & Olson, R.K. (2017) Explaining the sex difference in dyslexia. *J. Child Psychol. Psychiatry*, **58**, 719–727.

Ashburner, J. (2007) A fast diffeomorphic image registration algorithm. *NeuroImage*, **38**, 95–113.

Ashburner, J. & Friston, K.J. (2000) Voxel-based morphometry--the methods. *NeuroImage*, **11**, 805–821.

Ashburner, J. & Friston, K.J. (2005) Unified segmentation. *NeuroImage*, **26**, 839–851.

Asparouhov, T. & Muthén, B.O. (2010) Weighted Least Squares Estimation With Missing Data. *Mplus Tech. Append.*,.

Bishop, D.V.M. (2015) The interface between genetics and psychology: lessons from developmental dyslexia. *Proc R Soc B*, **282**, 20143139.

Chen, X., Sachdev, P.S., Wen, W., & Anstey, K.J. (2007) Sex differences in regional gray matter in healthy individuals aged 44-48 years: a voxel-based morphometric study. *NeuroImage*, **36**, 691–699.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014) Measurement Equivalence in Cross-National Research. *Annu. Rev. Sociol.*, **40**, 55–75.

Davidov, E., Schmidt, P., Billiet, J., & Meuleman, B. (2018) *Cross-Cultural Analysis: Methods and Applications, Second Edition*. Routledge.

de Jong, L.W., Vidal, J.-S., Forsberg, L.E., Zijdenbos, A.P., Haight, T., Sigurdsson, S., Gudnason, V., van Buchem, M.A., & Launer, L.J. (2017) Allometric scaling of brain regions to intra-cranial volume: an epidemiological MRI study. *Hum. Brain Mapp.*, **38**, 151–164.

Eckert, M.A., Berninger, V.W., Vaden, K.I., Gebregziabher, M., & Tsu, L. (2016) Gray Matter Features of Reading Disability: A Combined Meta-Analytic and Direct Analysis Approach. *eNeuro*, **3**.

Evans, T.M., Flowers, D.L., Napoliello, E.M., & Eden, G.F. (2014) Sex-specific Gray Matter Volume Differences in Females with Developmental Dyslexia. *Brain Struct. Funct.*, **219**, 1041–1054.

Finlay, B.L., Darlington, R.B., & Nicastro, N. (2001) Developmental structure in brain evolution. *Behav. Brain Sci.*, **24**, 263–278; discussion 278-308.

Fischer, C., Operto, G., Laguitton, S., Perrot, M., Denghien, I., Rivière, D., & Mangin, J. (2012) Morphologist 2012: the new morphological pipeline of BrainVISA. *Hum Brain Mapp*,.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., & Collins, D.L. (2011) Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *NeuroImage*, **54**, 313–327.

Germanaud, D., Lefèvre, J., Fischer, C., Bintner, M., Curie, A., des Portes, V., Eliez, S., Elmaleh-Bergès, M., Lamblin, D., Passemard, S., Operto, G., Schaer, M., Verloes, A., Toro, R., Mangin, J.F., & Hertz-Pannier, L. (2014) Simplified gyral pattern in severe developmental microcephalies? New insights from allometric modeling for spatial and spectral analysis of gyrification. *NeuroImage*, **102 Pt 2**, 317–331.

Hazlett, H.C., Poe, M.D., Lightbody, A.A., Styner, M., MacFall, J.R., Reiss, A.L., & Piven, J. (2012) Trajectories of Early Brain Volume Development in Fragile X and Autism RH: Trajectory of Brain Volume in Fragile X. *J. Am. Acad. Child Adolesc. Psychiatry*, **51**, 921–933.

Hong, S., Malik, M.L., & Lee, M.-K. (2003) Testing Configural, Metric, Scalar, and Latent Mean Invariance Across Genders in Sociotropy and Autonomy Using a Non-Western Sample. *Educ. Psychol. Meas.*, **63**, 636–654.

Im, K., Raschle, N.M., Smith, S.A., Ellen Grant, P., & Gaab, N. (2016) Atypical Sulcal Pattern in Children with Developmental Dyslexia and At-Risk Kindergarteners. *Cereb. Cortex N. Y. N 1991*, **26**, 1138–1148.

Jäncke, L., Liem, F., & Merillat, S. (2019) Scaling of brain compartments to brain size. *Neuroreport*, **30**, 573–579.

Jäncke, L., Mérillat, S., Liem, F., & Hänggi, J. (2015) Brain size, sex, and the aging brain. *Hum. Brain Mapp.*, **36**, 150–169.

Jednoróg, K., Marchewka, A., Altarelli, I., Monzalvo Lopez, A.K., van Ermingen-Marbach, M., Grande, M., Grabowska, A., Heim, S., & Ramus, F. (2015) How reliable are gray matter disruptions in specific reading disability across multiple countries and languages? Insights from a large-scale voxel-based morphometry study. *Hum. Brain Mapp.*, **36**, 1741–1754.

Jolicoeur, P. (1963) 193. Note: The Multivariate Generalization of the Allometry Equation. *Biometrics*, **19**, 497–499.

Krzanowski, W.J. (1979) Between-Groups Comparison of Principal Components. *J. Am. Stat. Assoc.*, **74**, 703–707.

Linkersdörfer, J., Lonnemann, J., Lindberg, S., Hasselhorn, M., & Fiebach, C.J. (2012) Grey Matter Alterations Co-Localize with Functional Abnormalities in Developmental Dyslexia: An ALE Meta-Analysis. *PLOS ONE*, **7**, e43122.

Lotze, M., Domin, M., Gerlach, F.H., Gaser, C., Lueders, E., Schmidt, C.O., & Neumann, N. (2019) Novel findings from 2,838 Adult Brains on Sex Differences in Gray Matter Brain Volume. *Sci. Rep.*, **9**.

Mangin, J.F., Rivière, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., Collins, D.L., Evans, A.C., & Régis, J. (2004) Object-based morphometry of the cerebral cortex. *IEEE Trans. Med. Imaging*, **23**, 968–982.

Mankiw, C., Park, M.T.M., Reardon, P.K., Fish, A.M., Clasen, L.S., Greenstein, D., Giedd, J.N., Blumenthal, J.D., Lerch, J.P., Chakravarty, M.M., & Raznahan, A. (2017) Allometric Analysis Detects Brain Size-Independent Effects of Sex and Sex Chromosome Complement on Human Cerebellar Organization. *J. Neurosci. Off. J. Soc. Neurosci.*, **37**, 5221–5231.

Marchewka, A., Kherif, F., Krueger, G., Grabowska, A., Frackowiak, R., Draganski, B., & Alzheimer's Disease Neuroimaging Initiative (2014) Influence of magnetic field strength and image registration strategy on voxel-based morphometry in a study of Alzheimer's disease. *Hum. Brain Mapp.*, **35**, 1865–1874.

Mascheretti, S., Bureau, A., Battaglia, M., Simone, D., Quadrelli, E., Croteau, J., Cellino, M.R., Giorda, R., Beri, S., Maziade, M., & Marino, C. (2013) An assessment of gene-by-environment interactions in developmental dyslexia-related phenotypes. *Genes Brain Behav.*, **12**, 47–55.

Mundfrom, D.J., Shaw, D.G., & Ke, T.L. (2005) Minimum Sample Size Recommendations for Conducting Factor Analyses. *Int. J. Test.*, **5**, 159–168.

Oliver, B.R. & Plomin, R. (2007) Twins' Early Development Study (TEDS): A Multivariate, Longitudinal Genetic Investigation of Language, Cognition and Behavior Problems from Childhood Through Adolescence. *Twin Res. Hum. Genet.*, **10**, 96–105.

Płoński, P., Gradkowski, W., Altarelli, I., Monzalvo, K., van Ermingen-Marbach, M., Grande, M., Heim, S., Marchewka, A., Bogorodzki, P., Ramus, F., & Jednoróg, K. (2017) Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Hum. Brain Mapp.*, **38**, 900–908.

R Foundation for Statistical Computing, R.C. (n.d.) R: a language and environment for statistical computing.

Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & Scotto di Covella, L. (2018) Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neurosci. Biobehav. Rev.*, **84**, 434–452.

Ramus, F., Rosen, S., Dakin, S.C., Day, B.L., Castellote, J.M., White, S., & Frith, U. (2003) Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, **126**, 841–865.

Reardon, P.K., Clasen, L., Giedd, J.N., Blumenthal, J., Lerch, J.P., Chakravarty, M.M., & Raznahan, A. (2016) An Allometric Analysis of Sex and Sex Chromosome Dosage Effects on Subcortical Anatomy in Humans. *J. Neurosci. Off. J. Soc. Neurosci.*, **36**, 2438–2448.

Reardon, P.K., Seidlitz, J., Vandekar, S., Liu, S., Patel, R., Park, M.T.M., Alexander-Bloch, A., Clasen, L.S., Blumenthal, J.D., Lalonde, F.M., Giedd, J.N., Gur, R.C., Gur, R.E., Lerch, J.P., Chakravarty,

M.M., Satterthwaite, T.D., Shinohara, R.T., & Raznahan, A. (2018) Normative brain size variation and brain shape diversity in humans. *Science*, **360**, 1222–1227.

Redcay, E. & Courchesne, E. (2005) When Is the Brain Enlarged in Autism? A Meta-Analysis of All Brain Size Reports. *Biol. Psychiatry*, **58**, 1–9.

Richlan, F., Kronbichler, M., & Wimmer, H. (2013) Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Hum. Brain Mapp.*, **34**, 3055–3065.

Ritchie, S.J., Cox, S.R., Shen, X., Lombardo, M.V., Reus, L.M., Alloza, C., Harris, M.A., Alderson, H.L., Hunter, S., Neilson, E., Liewald, D.C.M., Auyeung, B., Whalley, H.C., Lawrie, S.M., Gale, C.R., Bastin, M.E., McIntosh, A.M., & Deary, I.J. (2018) Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cereb. Cortex*, **28**, 2959–2975.

Ruigrok, A.N.V., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M.V., Tait, R.J., & Suckling, J. (2014) A meta-analysis of sex differences in human brain structure. *Neurosci. Biobehav. Rev.*, **39**, 34–50.

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J.-F., Bricout, L., Billard, C., Nguyen-Morel, M.-A., Le Heuzey, M.-F., Soares-Boucaud, I., George, F., Ziegler, J.C., & Ramus, F. (2016) Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Dev. Psychol.*, **52**, 1503–1516.

Sanchis-Segura, C., Ibañez-Gual, M.V., Adrián-Ventura, J., Aguirre, N., Gómez-Cruz, Á.J., Avila, C., & Forn, C. (2019) Sex differences in gray matter volume: how many and how large are they really? *Biol. Sex Differ.*, **10**.

Scarpazza, C., Sartori, G., De Simone, M.S., & Mechelli, A. (2013) When the single matters more than the group: very high false positive rates in single case Voxel Based Morphometry. *NeuroImage*, **70**, 175–188.

Schnabel, D.B.L., Kelava, A., van de Vijver, F.J.R., & Seifert, L. (2015) Examining psychometric properties, measurement invariance, and construct validity of a short version of the Test to Measure Intercultural Competence (TMIC-S) in Germany and Brazil. *Int. J. Intercult. Relat.*, **49**, 137–155.

Smith, S.M. & Nichols, T.E. (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, **44**, 83–98.

Steenkamp, J.-B.E.M. & Baumgartner, H. (1998) Assessing Measurement Invariance in Cross-National Consumer Research. *J. Consum. Res.*, **25**, 78–90.

Toro, R., Chupin, M., Garnero, L., Leonard, G., Perron, M., Pike, B., Pitiot, A., Richer, L., Veillette, S., Pausova, Z., & Paus, T. (2009) Brain volumes and Val66Met polymorphism of the BDNF gene: local or global effects? *Brain Struct. Funct.*, **213**, 501–509.

Williams, V.J., Juranek, J., Cirino, P., & Fletcher, J.M. (2018) Cortical Thickness and Local Gyrification in Children with Developmental Dyslexia. *Cereb. Cortex N. Y. N 1991*, **28**, 963–973.