

# How Reliable are Gray Matter Disruptions in Specific Reading Disability Across Multiple Countries and Languages? Insights from a Large-Scale Voxel-Based Morphometry Study

Katarzyna Jednoróg,<sup>1\*</sup> Artur Marchewka,<sup>2</sup> Irene Altarelli,<sup>3,4</sup>  
Ana Karla Monzalvo Lopez,<sup>5,6,7</sup> Muna van Ermingen-Marbach,<sup>8,9</sup>  
Marion Grande,<sup>10</sup> Anna Grabowska,<sup>1,11</sup> Stefan Heim,<sup>8,12</sup> and Franck Ramus<sup>3</sup>

<sup>1</sup>Laboratory of Psychophysiology, Department of Neurophysiology, Nencki Institute of Experimental Biology, Warsaw, Poland

<sup>2</sup>Laboratory of Brain Imaging, Neurobiology Center, Nencki Institute of Experimental Biology, Warsaw, Poland

<sup>3</sup>Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL Research University, Paris, France

<sup>4</sup>Brain and Learning Laboratory, FPSE, University of Geneva, Geneva, Switzerland

<sup>5</sup>INSERM, Cognitive Neuroimaging Unit, Gif sur Yvette, 91191 France

<sup>6</sup>Commissariat à l'Energie Atomique, Division of Life Sciences, Institute of BioImaging, Neurospin center, Gif sur Yvette, 91191 France

<sup>7</sup>University Paris 11, Orsay, France

<sup>8</sup>Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical Faculty, RWTH Aachen, Germany

<sup>9</sup>SRH University of Applied Sciences for Health Gera GmbH, Gera, Germany

<sup>10</sup>Section Clinical and Cognitive Neurosciences, Department of Neurology, Uniklinik RWTH Aachen, Germany

<sup>11</sup>University of Social Sciences and Humanities, Warsaw, Poland

<sup>12</sup>Research Centre Jülich, Institute of Neuroscience and Medicine [INM-1], Jülich, Germany



**Abstract:** The neural basis of specific reading disability (SRD) remains only partly understood. A dozen studies have used voxel-based morphometry (VBM) to investigate gray matter volume (GMV) differences between SRD and control children, however, recent meta-analyses suggest that few regions

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Agence Nationale de la Recherche; Contract grant number: ANR-06-NEURO-019-01, ANR-11-BSV4-014-01, ANR-10-LABX-0087 IEC, 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, IP2012 025272; 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 National Science Foundation; Contract grant number: 320030\_135679

\*Correspondence to: Katarzyna Jednoróg, Laboratory of Psychophysiology, Department of Neurophysiology, Nencki Institute of Experimental Biology, Warsaw, Poland.

E-mail: k.jednorog@nencki.gov.pl

Received for publication 17 March 2014; Revised 14 December 2014; Accepted 6 January 2015.

DOI: 10.1002/hbm.22734

Published online 00 Month 2015 in Wiley Online Library (wileyonlinelibrary.com).

are consistent across studies. We used data collected across three countries (France, Poland, and Germany) with the aim of both increasing sample size (236 SRD and controls) to obtain a clearer picture of group differences, and of further assessing the consistency of the findings across languages. VBM analysis reveals a significant group difference in a single cluster in the left thalamus. Furthermore, we observe correlations between reading accuracy and GMV in the left supramarginal gyrus and in the left cerebellum, in controls only. Most strikingly, we fail to replicate all the group differences in GMV reported in previous studies, despite the superior statistical power. The main limitation of this study is the heterogeneity of the sample drawn from different countries (i.e., speaking languages with varying orthographic transparencies) and selected based on different assessment batteries. Nevertheless, analyses within each country support the conclusions of the cross-linguistic analysis. Explanations for the discrepancy between the present and previous studies may include: (1) the limited suitability of VBM to reveal the subtle brain disruptions underlying SRD; (2) insufficient correction for multiple statistical tests and flexibility in data analysis, and (3) publication bias in favor of positive results. Thus the study echoes widespread concerns about the risk of false-positive results inherent to small-scale VBM studies. *Hum Brain Mapp* 00:000–000, 2015. © 2015 Wiley Periodicals, Inc.

**Key words:** voxel-based morphometry; reading impairment; multisite study; magnetic resonance imaging; gray matter volume

## INTRODUCTION

Specific reading disability (SRD, or “developmental dyslexia”) is characterized by a specific and persistent failure to acquire efficient reading skills that cannot be accounted for by low intelligence quotient (IQ), poor educational opportunities or obvious sensory or neurological damage [World Health Organization, 2008]. It occurs in 3–7% of all the children and often persists into adulthood [Lindgren et al., 1985]. SRD can affect any part of the reading process, including difficulty with accurate or fluent word recognition, word decoding, reading rate, and reading comprehension. Although the disorder can have a different expression from person to person, difficulties with spelling, phonological awareness, and rapid naming are often found. SRD can be understood as a neurodevelopmental disorder with a genetic origin that causes deficits at the cognitive level, which produce the behavioral symptoms of SRD. These hypothesized cognitive deficits are still subject to an ongoing debate as the large body of data fails to fit into single coherent theoretical framework [Ramus and Ahissar, 2012].

The search for neuroanatomical differences in SRD started with the histological postmortem observations of Galaburda and Kemper [1979] and Galaburda et al. [1985]. These studies demonstrated neuronal ectopias and architectonic dysplasias mainly in left perisylvian regions, which were interpreted as a consequence of disrupted neuronal migration during the prenatal stage [Galaburda et al., 1985; Humphreys et al., 1990]. Reduced left–right asymmetry of the planum temporale was also found [Galaburda and Kemper, 1979; Galaburda et al., 1985], as well as disruptions in thalamic structures consisting in disorganization and smaller neurons in the magnocellular layers of the lateral geniculate nuclei bilaterally [Livingstone

et al., 1991], and smaller neurons in the left medial geniculate nucleus [Galaburda et al., 1994]. Subsequent work also revealed abnormalities in primary visual cortex [Jenner et al., 1999] and suggested an increased proportion of large neurons and fewer small neurons in the cerebellum of SRD individuals [Finch et al., 2002]. However, the number of analyzed brains in post-mortem studies was quite small (ranging from one to five), hence they might not be representative of the population with SRD.

Since the introduction of magnetic resonance imaging (MRI), it has become possible to study brain structure in vivo in larger samples. Subsequently, Ashburner and Friston [2000] introduced voxel-based morphometry (VBM) allowing the objective automatic analysis of structural T1-weighted (T1w) MR scans. So far a number of VBM studies of gray matter volume (GMV) in SRD were published [Black et al., 2012; Brambati et al., 2004; Brown et al., 2001; Eckert et al., 2005; Evans et al., 2014; Hoeft et al., 2007; Jednoróg et al., 2014; Krafnick et al., 2014; Kronbichler et al., 2008; Liu et al., 2013; Menghini et al., 2008; Pernet et al., 2009a, b; Raschle et al., 2011; Silani et al., 2005; Siok et al., 2008; Steinbrink et al., 2008; Vinckenbosch et al., 2005]. One review of the group differences (increased and decreased GMV) observed across some of these studies, gives an impression of a widely distributed set of bilateral regions differing between individuals with SRD and controls [Richardson and Price 2009]. However, that review did not use meta-analytic methods to assess the consistency of the findings. Two meta-analyses actually revealed limited consistency [Linkersdörfer et al., 2012; Richlan et al., 2013]. This may be partly due to small sample sizes, typically involving 10–20 subjects per group. It is indeed quite remarkable that the largest published study [Pernet et al. 2009a], involving 38 and 39 participants per group, respectively, did not find any significant group

**TABLE I. Demographic and cognitive data**

Children	Sex	Age (months)	Reading accuracy (%)	Reading speed (words/s)	RAN_digits (items/s)	RAN_pictures (items/s)
French						
45 control	23 F 22 M	129.0 ( $\pm 16$ )	93.5 ( $\pm 7$ )	1.38 ( $\pm 0.4$ )	2.37 ( $\pm 0.4$ )	1.39 ( $\pm 0.2$ )
39 SRD	14 F 25M	130.4 ( $\pm 16$ )	47.1 ( $\pm 22$ )	0.40 ( $\pm 0.2$ )	0.65 ( $\pm 0.2$ )	1.02 ( $\pm 0.2$ )
German						
26 control	10 F 16 M	115.3 ( $\pm 7$ )	85.3 ( $\pm 11$ )	1.19 ( $\pm 0.4$ )	1.97 ( $\pm 0.4$ )	1.02 ( $\pm 0.1$ )
45 SRD	22 F 23 M	118.6 ( $\pm 7$ )	68.6 ( $\pm 20$ )	0.43 ( $\pm 0.2$ )	1.50 ( $\pm 0.4$ )	0.86 ( $\pm 0.2$ )
Polish						
35 control	22 F 13 M	123.9 ( $\pm 11$ )	93.5 ( $\pm 5$ )	1.05 ( $\pm 0.3$ )	2.15 ( $\pm 0.4$ )	1.23 ( $\pm 0.2$ )
46 SRD	20 F 26 M	123.2 ( $\pm 11$ )	52.6 ( $\pm 20$ )	0.51 ( $\pm 0.2$ )	1.94 ( $\pm 0.4$ )	1.07 ( $\pm 0.2$ )

difference when properly correcting for multiple tests. There is, therefore, an urgent need for larger VBM studies of SRD. Although VBM has become the most widely used method to study GM volume, interesting insights about reading skills may also be gained by studying its two components separately, that is, cortical thickness and surface [Altarelli et al. 2013; Frye et al. 2010; Goldman and Manis, 2012]. Here, however our focus is GMV as estimated by VBM.

While the aforementioned studies were carried out in a variety of countries and languages (e.g., Austria, China, France, Germany, Italy, Poland, UK, USA), only one VBM study of SRD [Silani et al., 2005] included participants from three different countries: France ( $n = 23$ ), Italy ( $n = 19$ ), and UK ( $n = 22$ ). In a pre-specified region of interest based on a PET study on the same sample, these authors found a reduction of GMV in the left middle temporal cortex (BA21) in SRD subjects, coupled with an increase of GMV in the more posterior adjacent area (BA37). No other regions in the whole brain approach outside this ROI were revealed. Since the authors did not find country-specific differences, they concluded that the neurological disorder underlying SRD is the same across the three languages.

Here we followed this cross-linguistic approach with the aim of both increasing sample size to obtain a clearer picture of group differences in GMV, and of further assessing the consistency of the findings across languages. We constituted a large population ( $n = 236$ ) of SRD and control children from three different countries: France, Germany, and Poland, and we examined group differences in GMV and as well as associations between GMV, literacy, and phonological skills.

Having a cross-linguistic study design, one might expect differences in atypical reading development across languages based on their orthographic transparency. It is well known that transparent orthographies with high symbol-sound consistency (e.g., German) are acquired more easily than complex and opaque orthographies with a high proportion of inconsistent and irregular spellings (e.g., English, French) [Seymour et al., 2003]. However, the cognitive

predictors of reading performance and of SRD (usually phonological awareness and rapid naming) seem to be universal across alphabetic languages, although their relative weight varies systematically as a function of script transparency [Landerl et al., 2013; Ziegler et al., 2010]. It is also likely that atypical reading development will share common neurocognitive etiologies even in languages that vary considerably in orthographic consistency [Paulesu et al., 2001; Silani et al., 2005; Siok et al., 2004]. This study aims to uncover the common gray matter (GM) differences that characterize SRD across most alphabetic languages, but it is not in a position to study differences that depend on the orthographic system (this would require many more languages).

## MATERIALS AND METHODS

### Participants

Participants came from diverse social backgrounds and had at least one and a half years of formal reading instruction to differentiate serious problems in reading acquisition from early delays that are not always persistent. SRD participants were either identified in school, through clinic or were specifically requesting clinical assessment of their reading problems. Most of the studied children already had a clinical diagnosis of SRD and all were screened for inattention/hyperactivity symptoms and other language disorders.

The sample included children from three countries: 81 Polish children—35 control (22 girls) and 46 SRD (20 girls); 84 French children—45 control (23 girls) and 39 SRD (14 girls); 71 German children—26 control (10 girls) and 45 SRD (22 girls) (Table I). Other analyses based on partly overlapping datasets have been published before for the French [Altarelli et al., 2014; Altarelli et al., 2013; Jednoróg et al., 2012; Monzalvo et al., 2012], and the Polish children [Jednoróg et al., 2014]. Participants were recruited following the criteria below: age between 8.5 and 13.7 years, IQ higher than 85, or an age-appropriate scaled score of at least 7 on WISC Block Design, and 6 on WISC Similarities, no language disorders, no formal diagnosis of ADHD, no

reported hearing, sight, or neurological problems. The inclusion criterion for SRD children was defined as more than 1.5 SD below grade level on different language-appropriate standardized test of reading, whereas for controls it was 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.

### Measures

Reading accuracy and speed were assessed using different language-appropriate standardized tests of reading. In French, “L'alouette” [Lefavrais, 1967] and/or ODEDYS were used [Jacquier-Roux et al., 2005]. Participants had to read aloud a meaningless text and a list of single words (respectively) as quickly and accurately as possible, with both accuracy and speed being coded. In the German sample, the Würzburger Leise Leseprobe (WLLP) was used for a subset of 26 children [Küspert and Schneider, 1998]. Participants read words at the beginning of a row and marked with a pencil the one out of four pictures displayed in the same row that is denoted by the word. The names of the three distractor pictures may be semantically related, phonologically related, or semantically and phonologically related to the target word. For a subset of 45 German children, scores for reading single word lists were used [Repscher et al., 2012]. In the Polish sample, Real Word Reading (WDREAD) from the normalized Polish battery of tests was used [Bogdanowicz et al. 2008]. Participants had to read aloud single words as quickly and accurately as possible, with both accuracy and speed being coded.

Reading accuracy scores were available for all the children, while reading speed was available for only a subset of participants ( $n = 185$ ; 59 French, 45 German, and 81 Polish). The same subset also had naming speed assessed via language-specific rapid automatized naming (RAN) tasks requiring children to sequentially name as quickly as possible lists of digits and pictures depicting easily recognizable objects. The dependent measure was the number of items named per minute.

All behavioral measures (reading accuracy, speed, RAN digits, and pictures) were transformed into  $z$ -scores relative to the distribution of the same test in control participants of the same grade and the same country.

### Imaging Procedure

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

#### French group

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

platform using either 12-channels head coil (13 control and 11 SRD) with the following parameters: acquisition matrix:  $256 \times 256 \times 176$ , TR = 2,300 ms, TE = 4.18 ms, flip angle =  $9^\circ$ , FOV = 256 mm, voxel size:  $1 \times 1 \times 1$  mm, or 32-channels head coil (32 control and 28 SRD) with the following parameters: acquisition matrix =  $230 \times 230 \times 202$ , TR = 2,300 ms, TE = 3.05 ms, flip angle =  $9^\circ$ , FOV = 230 mm, voxel size =  $0.9 \times 0.9 \times 0.9$  mm.

#### German group

In the case of 10 control and 35 SRD children, 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:  $256 \times 256 \times 176$ , TR = 1,900 ms, TE = 2.52 ms, flip angle =  $9^\circ$ , FOV = 256 mm, voxel size:  $1 \times 1 \times 1$  mm. For the rest (16 control and 10 SRD), whole brain images were acquired on a 1.5T Siemens Avanto scanner using a standard birdcage head coil with the following parameters: acquisition matrix:  $256 \times 256 \times 170$ , TR = 2,200 ms, TE = 3.93 ms, flip angle =  $15^\circ$ , FOV = 256 mm, voxel size:  $1 \times 1 \times 1$  mm.

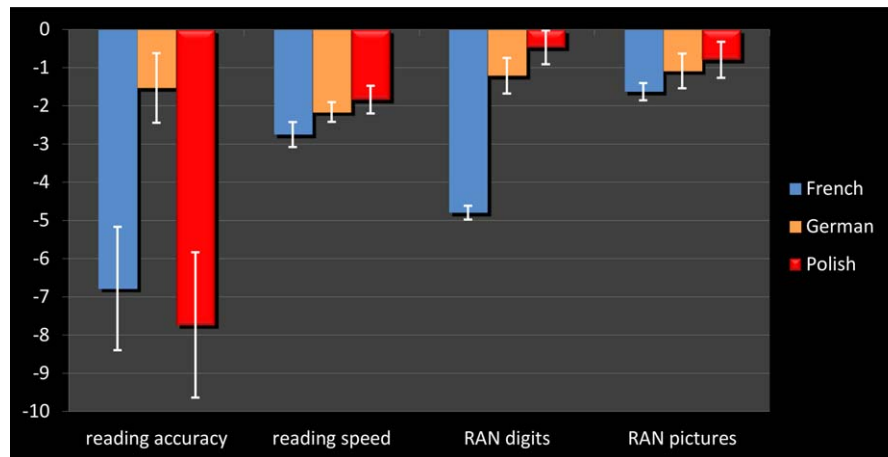
#### Polish group

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:  $256 \times 256 \times 192$ ; TR = 1,720 ms; TE = 2.92 ms; flip angle =  $9^\circ$ , FOV = 256, voxel size  $1 \times 1 \times 1$  mm.

### VBM Analysis

Image processing and analysis were carried out using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) run in MATLAB 7.11 (Mathworks, Sherborn, MA). T1w images were segmented automatically into different tissue classes—GM, white matter, and nonbrain (cerebrospinal fluid, skull), using the “New Segmentation” option in SPM8 [Ashburner and Friston, 2005]. Tissue probability maps were taken from a customized pediatric 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 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 and Friston, 2000]. Finally the data was smoothed with 4-mm full-width at half maximum Gaussian kernel.

First, whole brain anatomical differences between groups (control vs SRD children) were investigated. Since language was largely confounded with scanner/study,



**Figure 1.**

Behavioral performance in reading accuracy, reading speed, and naming speed in the SRD sample. Raw data were transformed into z-scores relative to the distribution of the group of normal controls of the same grade and same country. Error bars represent standard deviation. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

we could not analyze language differences and interactions with the language factor. Next, exploratory linear regressions with group as a factor were performed to analyze correlations between GM volume and behavioral measures converted into z-scores. Across-group main effects and group  $\times$  measure interactions are reported. In both analyses, age, sex, study, and total intracranial volume (TIV) were introduced as nuisance variables.

The correction for multiple comparisons was performed using a Monte Carlo simulation (3dClustSim, AFNI, <http://afni.nimh.nih.gov>). Only clusters with a minimum of 150 contiguous voxels and  $P_{\text{voxel}} < 0.001$  were considered as significant ( $P < 0.05$ ). All the reported clusters sizes were also adjusted according to local roughness using nonstationary correction [Hayasaka et al., 2004]. Clusters meeting nominal significance levels with an extent of at least 50 voxels are reported in tables for the purpose of future meta-analyses, but will not be discussed.

## RESULTS

### Behavioral Results

There was a significant difference in reading accuracy between control and SRD children ( $F(1,223) = 288.76$ ,  $P < 0.001$ , Fig. 1). There was a significant sex effect ( $F(1,223) = 4.23$ ,  $P = 0.041$ ), with girls performing better than boys. There was also a significant effect of Language ( $F(2,223) = 41.75$ ,  $P < 0.001$ ) and a Group  $\times$  Language interaction ( $F(2,223) = 32.21$ ,  $P < 0.001$ ). In the control group, there were no differences between countries (by definition of the z-scores), whereas the German SRD children performed better than the French and Polish ones (both  $P < 0.001$ ). There was no difference between Polish and French SRD children.

Control children read words more quickly than SRD children ( $F(1,172) = 259.72$ ,  $P < 0.001$ ). There was also a significant effect of Language ( $F(2,172) = 6.96$ ,  $P = 0.001$ ) and a Group  $\times$  Language interaction ( $F(2,172) = 5.59$ ,  $P = 0.004$ ). While there were no differences between countries in the control group, French SRD children read more slowly than the German ( $P = 0.001$ ) and Polish ones ( $P < 0.001$ ).

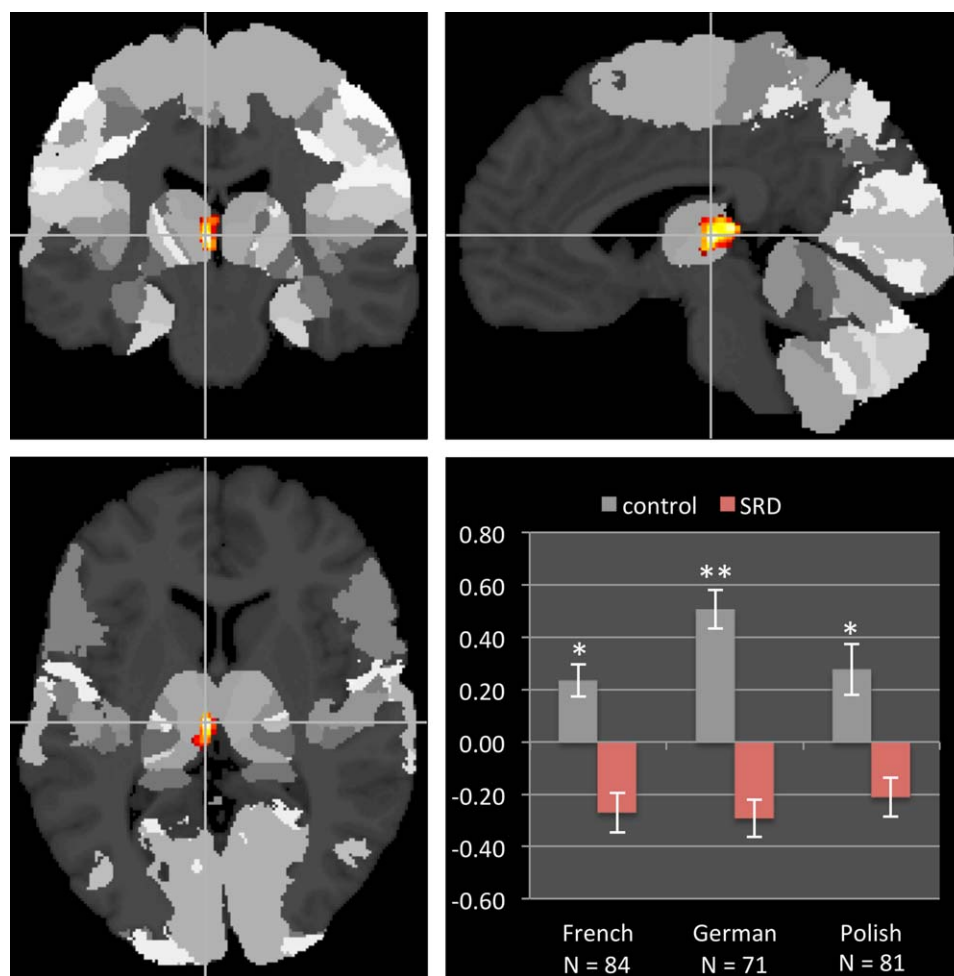
Control children had also better scores than SRD children in RAN tests (digits  $F(1,172) = 187.29$ ,  $P < 0.001$  and pictures  $F(1,172) = 42.23$ ,  $P < 0.001$ ). In the case of digits, a significant effect of Language ( $F(2,172) = 90.51$ ,  $P < 0.001$ ) and a Group  $\times$  Language interaction ( $F(2,172) = 98.20$ ,  $P < 0.001$ ) were revealed. Again, while there were no differences between countries in the control group, French SRD children showed the poorest performance (all  $P < 0.001$ ) and German SRD children performed worse than Polish ones ( $P = 0.003$ ).

We found no significant effect of age on reading accuracy, reading, or naming speed in the current sample.

### VBM Results

#### Group differences

Control children had higher TIV than SRD children ( $F(1,223) = 12.12$ ,  $P = 0.001$ ,  $d = 0.83$ ). The VBM comparison between groups revealed a significant reduction of GMV in the left thalamus of SRD children (pulvinar and medial dorsal nuclei,  $x = -5$ ,  $y = -20$ ,  $z = 5$ ,  $t = 3.99$ ,  $z = 3.92$ , cluster size = 244 voxels, Fig. 2). The reversed contrast showed a cluster in the left inferior parietal lobule ( $x = -41$ ,  $y = -35$ ,  $z = 36$ ,  $t = 4.24$ ,  $z = 4.15$ , cluster size = 103 voxels), but it did not survive the correction for multiple comparisons.



**Figure 2.**

Group differences in gray matter volume overlaid on the Julich Anatomy Atlas. The region in the left thalamus shown in red–yellow exhibited reduced volume in the dyslexic group (prefrontal—54.6 % probability, temporal—10.5 % probability). Mean gray matter volume extracted from this cluster (displayed in

z-scores) shows consistent group difference across three languages. Error bars represent SEM. \*\* $P < 0.001$ ; \* $P < 0.05$ . [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

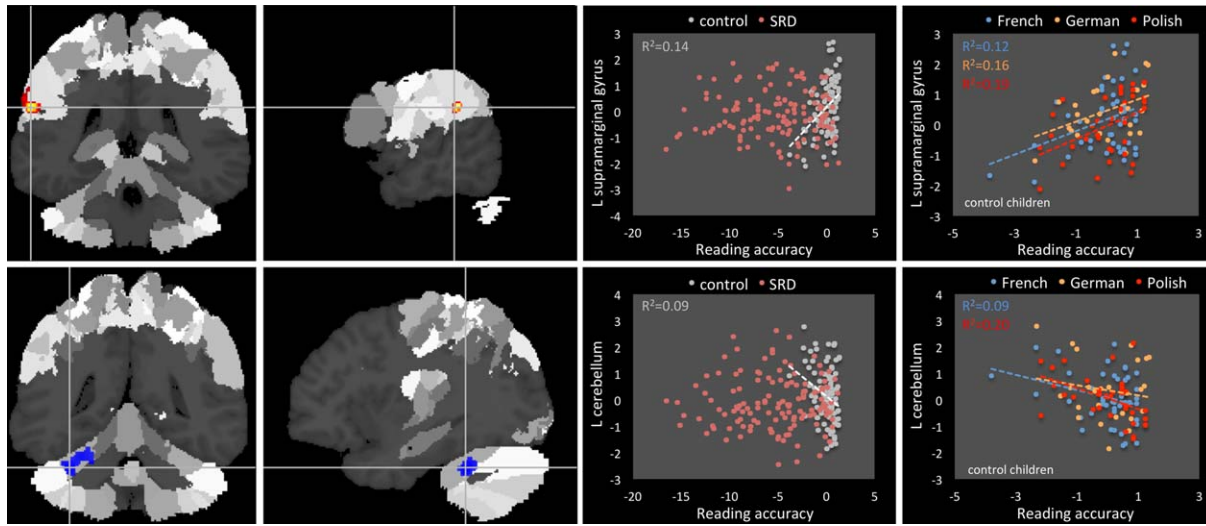
The reduction of GMV in the left thalamus cluster was consistent across the three language groups ( $t_{\text{French}} = 2.62$ ,  $P = 0.011$ ,  $d = 0.58$ ;  $t_{\text{German}} = 3.69$ ,  $P < 0.001$ ,  $d = 0.88$ ;  $t_{\text{Polish}} = 2.02$ ,  $P = 0.047$ ,  $d = 0.45$ ). Separate group analyses in each language revealed few significant clusters and no consistency between languages (Supporting Information Table I). Since there was no significant interaction between group and sex, the latter was treated only as a covariate.

### Linear Regressions

#### Reading accuracy

A significant across-group negative correlation of GMV with reading accuracy was found in the left cerebellum

(lobule VI, Crus 1). There were also significant group  $\times$  reading accuracy interactions in the left supramarginal gyrus (SMG) and the left cerebellum (lobule VI, Crus 1; Fig. 3 and Table II). While in the left SMG controls showed a positive correlation between GMV and reading accuracy, no such relation was observed for the SRD children. Further, the positive correlation in the left SMG in controls was consistent across languages ( $r_{\text{French}} = 0.35$ ,  $P = 0.018$ ;  $r_{\text{German}} = 0.40$ ,  $P = 0.042$ ;  $r_{\text{Polish}} = 0.44$ ,  $P = 0.008$ ). Again for the left cerebellum, a significant negative correlation was found only in control children, but it was less consistent across languages, as it reached significance only in the French and Polish groups ( $r_{\text{French}} = -0.30$ ,  $P = 0.046$ ;  $r_{\text{German}} = -0.17$ ,  $P > 0.1$ ;



**Figure 3.**

Correlation between GMV and reading accuracy. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

$r_{\text{Polish}} = -0.45, P = 0.006$ ). Similarly as in the group analyses, the interaction with sex was not significant, therefore, sex was only included as a covariate.

### Reading speed

No across-group correlations or interactions reached significance (Supporting Information Table II).

### RAN digits

No across group effects or interactions reached significance (Supporting Information Table III).

### RAN pictures

The across-group analysis revealed a significant negative correlation between naming pictures and GMV in the left

**TABLE II. Correlations between gray matter volume and reading accuracy**

Region	MNI			<i>T</i>	<i>R</i>	Cluster size (voxels)
	<i>x</i>	<i>y</i>	<i>z</i>			
Positive correlation—whole group						
L supramarginal gyrus	-60	-39	24	4.75	0.30	50
L superior occipital gyrus	-23	-90	36	4.17	0.26	51
L insula	-36	22	-4	3.82	0.24	65
Negative correlation—whole group						
<i>L cerebellum lobule VI, Crus 1</i>	-32	-42	-27	4.58	-0.28	935
	-38	-45	-34	4.40	-0.28	
R cerebellum lobule VIII	15	-62	-52	3.51	-0.22	96
R_control > R_SRD						
<i>L supramarginal gyrus</i>	-60	-39	24	5.38	0.34	161
	-65	-32	29	3.71	0.23	
L supramarginal/postcentral gyrus	-62	-24	35	4.00	0.24	63
R_SRD > R_control						
R cerebellum lobule VI	27	-44	-34	4.37	-0.21	59
L precentral (BA6)	-41	-2	48	4.26	-0.27	108
<i>L cerebellum lobule VI, Crus 1</i>	-36	-45	-34	4.05	-0.26	652
	-30	-42	-27	4.01	-0.26	

Italics are clusters surviving correction for multiple comparisons  $P < 0.05$ .

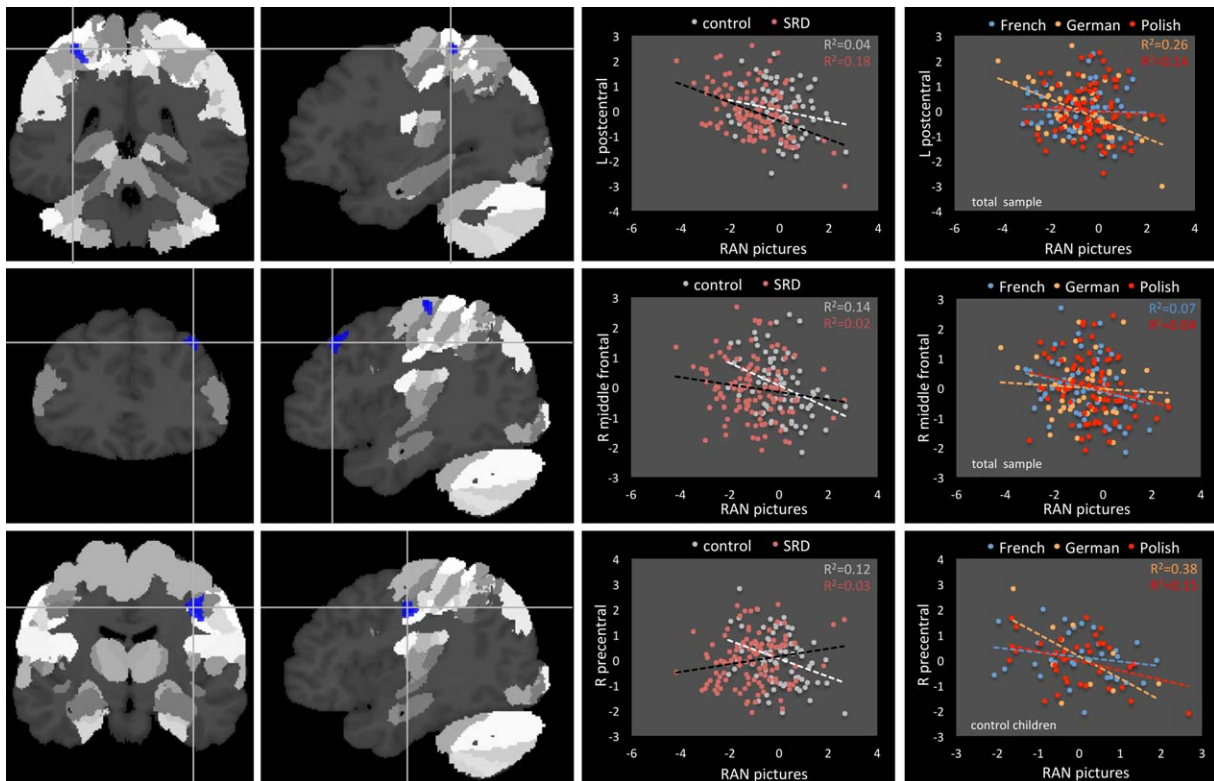
**TABLE III. Correlations between gray matter volume and RAN pictures**

Region	MNI			T	R	Cluster size (voxels)
	x	y	z			
Positive correlation—whole group						
R superior temporal pole	39	18	-30	3.79	0.27	62
Negative correlation with—whole group						
L lingual gyrus	-23	-57	-7	4.66	-0.31	89
<i>L postcentral gyrus</i>	-30	-32	69	4.55	-0.32	241
<i>R precentral gyrus</i>	41	-23	63	4.46	-0.30	260
<i>R middle frontal gyrus</i>	39	33	41	4.32	-0.31	237
L insula (BA 13)	-30	-21	17	3.99	-0.27	115
R_control > R_SRD						
—						
R_SRD > R_control						
<i>R precentral gyrus</i>	45	-6	36	4.45	-0.31	355
	48	-9	47	4.20	-0.29	
R cerebellum (VI)	33	-42	-37	3.68	-0.26	50

Italics are clusters surviving correction for multiple comparisons  $P < 0.05$ .

postcentral gyrus, right precentral gyrus, and right middle frontal gyrus (Table III and Fig. 4). In the case of left postcentral gyrus, the correlation was significant for German ( $r = -0.51, P < 0.001$ ) and Polish ( $r = -0.37, P = 0.001$ ) sam-

ples, whereas for the right middle frontal gyrus the correlation was significant in the French group only ( $r = -0.26, P = 0.047$ ) and there was a trend in Polish (Polish  $r = -0.21, P = 0.065$ ). There was also a significant group  $\times$



**Figure 4.**

Correlation between GMV and RAN pictures. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



naming interaction in the right precentral gyrus. While control children showed a negative correlation between GMV in this region and picture naming ( $r = -0.35$ ,  $P = 0.002$ ), there was a trend for positive correlation in SRD children ( $r = -0.35$ ,  $P = 0.059$ ). The negative correlation in controls was significant in Polish and German samples ( $r = -0.39$ ,  $P = 0.021$  and  $r = -0.62$ ,  $P = 0.057$ ). No interaction with sex was revealed.

## DISCUSSION

Despite more than a decade of studies investigating GMV in SRD, few consistent results have emerged. To assess group differences in GMV with greater confidence, we have pulled together data acquired at four different sites from three different countries: France, Germany, and Poland. In this large sample of 130 SRD and 106 control children, we first observed behavioral differences, most probably related to the orthographic transparency of the studied languages. French children were slower on reading tasks and rapid naming than their German and Polish peers, whereas German children had the best performance in reading accuracy [consistent with the observations of Ziegler et al., 2003].

At the neural level, the SRD group was characterized by a significant TIV reduction compared with control age-matched subjects, in line with several earlier studies [Casanova et al., 2004; Eckert et al., 2003, 2005; Eliez et al., 2000]. Beyond this global difference, a local reduction of GMV was also observed in the left thalamus (pulvinar and medial dorsal nuclei), consistently in each language. No other GMV differences were found between the two groups.

Furthermore, we observed significant correlations between GMV, reading accuracy and rapid naming of pictures. In controls, reading accuracy correlated positively with GMV in the left SMG and negatively with GMV in the left cerebellum lobule VI, whereas no such relation was present in SRD children. Performance in RAN pictures was negatively correlated with GMV in the right precentral gyrus in controls, but positively so in SRD children. RAN pictures also correlated negatively across both groups with left postcentral and right middle frontal gyri GM. We failed to find any significant GMV correlations for reading speed and rapid naming of digits.

VBM group differences revealed in this study stand in marked contrast to most previous anatomical studies of reading disability. Although the thalamus has been implicated in several studies of SRD, in particular with cellular anomalies [Galaburda et al. 1994; Livingstone et al. 1991] and although connections to acquired reading deficits with lesions specific to pulvinar have been reported [Crosson, 1999], the reduction of GMV in this region is consistent with only one previous VBM study of SRD using liberal significance thresholds [Brown et al., 2001]. At the functional level, the left thalamus was shown to be consistently

activated in a meta-analysis of reading studies [Turkeltaub et al., 2002]. In agreement, the left thalamus was found to be underactivated in SRD in a meta-analysis done by Maisog et al. [2008], as well as in a more recent study by Díaz et al. [2012]. Moreover, in an fMRI study of linguistic stimuli processing in early and late talkers, the thalamus showed activations that distinguished children at-risk for literacy problems from other subjects [Preston et al., 2010]. The authors argued that if articulatory proficiency, language learning, phonological skills, and reading rely on the integrity of subcortical structures (such as the thalamus) in young children, and these structures are underengaged during simple language processing tasks, early language development would be likely to suffer. Furthermore, in adolescents performing online print learning, changes in activity of the left thalamus discriminated typically developing from reading disabled learners [Pugh et al., 2008]. Finally, brain-behavior analyses in beginning readers showed that reading skills measures were positively correlated with print-related activation levels in bilateral posterior thalamus, with greater involvement of the left pulvinar [Pugh et al., 2013]. It has been suggested that the pulvinar mediate interactions between visual language and attentional regions, shaping the functional organization of the ventral visual pathways for orthographic form learning [Pugh et al., 2013]. This argument was based on studies showing rich structural connectivity between pulvinar and distributed cortical systems including frontoparietal, superior temporal, and visual cortex [Baleydier and Morel, 1992; Casanova et al., 2004], as well as reports implicating pulvinar in the control of visually guided attention [Posner and Raichle, 1995]. Similarly, the medial dorsal nucleus with its connections to the prefrontal cortex [Giguere and Goldman-Rakic, 1998] plays an important role in attentional focus [Buchsbaum et al., 2006; Crosson, 1999]. In this context, it seems possible that the pulvinar and the medial dorsal nuclei of the left hemisphere mediate selective attention to features that shape orthographic learning with the input of regions sensitive to linguistic forms and thus their disruption in SRD is plausible. All things considered, while the left thalamus is not a structure that is consistently observed to structurally differ in SRD, if this result was more broadly replicated it could be weaved into a relatively plausible story of speech and literacy development.

With respect to correlations, our most plausible result is the association between GMV in the left SMG and reading accuracy. It is well known that the SMG is important for reading words and pseudowords. The evidence comes from both lesion studies and structural imaging studies in healthy populations. In a perfusion study of acute stroke patients, the SMG was one of two areas (together with the fusiform gyrus) in which hypoperfusion predicted impairment in reading and spelling words and nonwords. In the same line, transcranial magnetic stimulation applied over the SMG selectively disrupts phonological decoding [Hartwigsen et al., 2010]. Interestingly, in adults, anatomical variations in the left SMG account for individual differences in reading

performance among typical readers in their first [Goldman and Manis, 2013; Lee et al., 2007] and second language [He et al., 2013]. Our results thus extend these findings to typically reading French, German, and Polish children in their first language.

The negative correlation between reading accuracy and GMV in the left cerebellum lobule VI is also consistent with recent neuroimaging studies suggesting that the cerebellum may be part of the reading network in typically developing readers. Reading-related activity is typically focused in lobules VI and VII [for a review, see Stoodley and Stein, 2011]. It has been suggested that the left cerebellum is activated while processing the morphology of word forms (reading low-frequency words), whereas the right is more active during phonological processing (reading non-words) [Joubert et al., 2004]. Additionally, a recent meta-analysis of eye movement fMRI and PET studies [Jamadar et al., 2013] pointed that specific eye movements consistently activated left cerebellar lobule VI. This lobule is likely to be involved in a range of cognitive processes, including spatial tasks [Stoodley and Schmahmann, 2010], requiring one to visually follow the continuum of text during reading. Some VBM studies comparing SRD and control groups reported reduced GMV bilaterally in the cerebellar nuclei [Brambati et al., 2004], lateral lobule VII [Brown et al., 2001], and the anterior cerebellum extending to lobule VI [Kronbichler et al., 2008]. However, we do not find any group differences in the cerebellum, and our negative correlation would on the contrary predict increased GMV in SRD children. More consistently, Pernet et al. [2009a] in their large study of French adults did observe a negative correlation between GMV in the cerebellum and reading-related task performance (pseudoword reading and phoneme deletion) in controls. However, this was in the left cerebellar declive (as opposed to left lobule VI). Thus, while the association of the cerebellum with reading is plausible, the specific results obtained are not particularly consistent with earlier studies. Furthermore, given that cerebellar cortex is typically only a single voxel thick (in contrast with cerebral cortex and subcortical structures) and varies considerably in shape, making it difficult to register to a template, one should be wary of concluding that morphometric differences observed here reflect volume differences alone.

The greatest limitation of our correlation analyses is that the tests used to measure reading accuracy and speed differed substantially between the languages. While for Polish, French, and a subset of German children timed single word reading tests were used, for a subset of German children a cloze sentences procedure was applied. This was in part due to the fact that each study was initially carried out independently, and in part related to the way the manifestations of SRD vary depending on language [Landerl et al., 2013; Ziegler et al., 2003]. These correlation analyses were, therefore, exploratory and the corresponding results are to be taken with caution. Along the same lines, given that the manifestations of SRD vary depending on

orthographic transparency, there is no guarantee that sets of children defined as being reading disabled are 100% the same across languages, and any such disparity will diminish chances of finding common deficits in cross-linguistic studies. Nevertheless, similarities of SRD across alphabetic languages seem to be sufficient for cross-linguistic studies to report largely overlapping findings [e.g., Landerl et al., 2013; Paulesu et al., 2001; Silani et al., 2005].

Overall, our most striking result is the failure to observe most of the differences previously reported in VBM studies of SRD (such as left inferior frontal, temporoparietal, or fusiform regions), including the outcome of meta-analyses [Linkersdörfer et al., 2012; Richlan et al., 2012]. Yet our sample size ( $n = 236$ ) is almost as large as that of these meta-analyses ( $n = 277$  and  $266$ , respectively). One would thus expect not only to replicate at least some of the previous findings, but also to observe new differences of effect sizes too small to be observed in smaller studies.

Part of the explanation may come from the limitations of our study. For one thing, the fact that our large sample size arises from the merging of five different studies carried out in four MRI scanners in three countries does introduce some heterogeneity, which may not entirely be adjusted for by simply entering study as a covariate in the analysis [Focke et al., 2011; Marchewka et al., 2014; Stonnington et al., 2008]. Thus our effective statistical power may not be as large as the sheer number of participants suggests. Also a potential confound of different head coils with multiple channels and the difference in head size between the groups should be acknowledged. This might induce a bias in the current study, as head coil performance is a key factor in determining sensitivity, while the head size influences the distance from the coil. Introducing both site-specific and TIV regressors minimized the influence of these two factors, however, we cannot exclude the potential interaction between them. Furthermore, the heterogeneity of the sample included in this study may be increased by the children being selected using different batteries of reading measures in each language. It is thus likely that the present SRD group drawn from different countries (i.e., speaking languages with varying orthographic transparencies) is more heterogeneous than the ones studied in previous single-language VBM studies and, therefore, less comparable to them. It is also possible that due to differences in assessment batteries within the SRD group, different reading deficits are mixed together, such as decoding, comprehension, or fluency deficits.

Nevertheless, our Polish and our second French studies are each larger than all previously published VBM studies of SRD children, and as homogeneous, so they would be expected to replicate some of the earlier results, yet they do not. Thus, each of them constitutes an independent nonreplication of previous studies, with larger sample sizes, and unaffected by heterogeneity factors such as language and scanning site. Similarly, our cross-linguistic study far exceeds the sample size of the only previous cross-linguistic study [ $n = 9-12$  per group and per

language in Silani et al. 2005], and we do not replicate their group difference (but they studied adults only). It may also be noted that our study includes far fewer scanning sites and languages, and is therefore, far less heterogeneous than the set of studies included in meta-analyses, so we should be able to replicate at least the results of meta-analyses, yet we do not.

Another potential limitation, that we share with all the previous studies but one [Jednoróg et al., 2014], is that we treated all SRD participants as a homogeneous group, whereas SRD may be a heterogeneous disorder, with subtypes having different brain bases [Hadzibeganovic et al., 2010]. If SRD is indeed that heterogeneous, this is a potential concern for any study of SRD. However, at this moment, there are many proposals of typologies of SRD subtypes [Boder, 1973; Bosse et al., 2007, Castles and Coltheart, 1993; Heim et al., 2008 and many others], and none of them has gained widespread acceptance. Another view [e.g., Ramus, 2003] is that the heterogeneity is confined to a few minority subtypes, with one majority subtype (that with a phonological deficit) accounting for the average results obtained in most studies. Such limited heterogeneity would increase the noise and skew the results, but should not preclude consistently replicating some results reflecting the majority subtype, and indeed, in the area of functional brain imaging of SRD, heterogeneity, and putative subtypes do not seem to prevent a reasonable degree of consistency across studies [Richlan et al. 2011], including across different languages [Paulesu et al., 2001]. Only stronger evidence for putative subtypes and even larger neuroimaging studies taking those subtypes into account will have a chance to provide a clearer picture.

There may be more general explanations for the inconsistencies observed. One might be that the VBM method is not well-suited to reveal subtle structural brain disruptions that characterize SRD children (as opposed to major neurological diseases). This might be due to failure to align matching cortical regions across subjects, resulting in a lack of power to localize differences [Scanlon et al., 2011]. Second, it is possible that some of the previous results of VBM studies of SRD, obtained on smaller samples, might have spuriously emerged from underestimation of individual variability. Adding the many degrees of freedom available in data analysis to make a significant effect emerge, and the general publication bias for positive findings, it is perfectly possible that most of the previously reported findings might be false positives. Indeed, the largest study before the present one [Pernet et al. 2009a;  $n = 77$ ] also failed to find significant group differences, and thus focused on correlations.

There is a growing awareness of the problem of poor replication in biomedical research [Ioannidis, 2005], and neuroscience and brain imaging may be particularly affected [Button et al. 2013; Fusar-Poli et al., 2014; Ioannidis, 2011]. Very recently, a large-scale study, performed on over 900 images taken from Autism Brain Imaging Data Exchange, failed to replicate most of the previous anatomical findings

in autistic spectrum disorder done on smaller samples [Haar et al., in press]. Another recent study attempting a confirmatory replication of published structural brain-behavior correlations using VBM and diffusion-tensor imaging also reported a general failure [Boekel et al. in press].

This study may well illustrate this phenomenon in the particular case of VBM and SRD. It suggests that if there are real local GMV differences between reading disabled and control children that hold across alphabetic languages, they must be very small to be so parsimoniously detected with more than 100 participants per group. Even the group differences and correlations reported here should be considered with caution and will require replication by even larger studies to be considered established.

## ACKNOWLEDGMENTS

The authors thank Ghislaine Dehaene, Catherine Billard, Joel Fluss, Nadège Villiermet, Stéphanie Iannuzzi, Natalia Gawron, Łukasz Żurawski and Julia Pape-Neumann for their contribution to study design and data collection. The authors thank all the participants and their families, technical and clinical staff at Hôpital Bicêtre, Neurospin centre, MRlab in Warsaw, and the Jülich-Aachen Research Alliance (Uniklinik RWTH Aachen and Forschungszentrum Jülich). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCES

- Altarelli I, Monzalvo K, Iannuzzi S, Fluss J, Billard C, Ramus F, Dehaene-Lambertz G (2013): A functionally guided approach to the morphometry of occipitotemporal regions in developmental dyslexia: Evidence for differential effects in boys and girls. *J Neurosci* 33:11296–11301.
- Altarelli I, Leroy F, Monzalvo K, Fluss J, Billard C, Dehaene-Lambertz G, Galaburda AM, Ramus F (2014): Planum temporale asymmetry in developmental dyslexia: Revisiting an old question. *Hum Brain Mapp* 35: 5717–5735.
- Ashburner J (2007): A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113.
- Ashburner J, Friston KJ (2000): Voxel-based morphometry—The methods. *Neuroimage* 11:805–821.
- Ashburner J, Friston KJ (2005): Unified segmentation. *Neuroimage* 26:839–851.
- Baleydier C, Morel A (1992): Segregated thalamocortical pathways to inferior parietal and inferotemporal cortex in macaque monkey. *Vis Neurosci* 8:391–405.
- Black JM, Tanaka H, Stanley L, Nagamine M, Zakerani N, Thurston A, Kesler S, Hulme C, Lyytinen H, Glover GH, Serrone C, Raman MM, Reiss AL, Hoefft F (2012): Maternal history of reading difficulty is associated with reduced language-related gray matter in beginning readers. *Neuroimage* 59:3021–3032.
- Boder E (1973): Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. *Dev Med Child Neurol* 15:663–687.

- Boekel W, Wagenmakers, EJ, Belay L, Verhagen J, Brown S, Forstmann BU: A purely confirmatory replication study of structural brain-behavior correlations. *Cortex* (in press).
- Bogdanowicz M, Jaworowska A, Krasowicz-Kupis G, Matczak A, Pelc-Pekala O, Pietras I, Stańczak J, Szczerbiński M (2008): Diagnostyka dysleksji u uczniów klasy III szkoły podstawowej. Warszawa: Pracownia Testów Psychologicznych.
- Bosse ML, Tainturier MJ, Valdois S (2007): Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition* 104:198–230.
- Brambati SM, Termine C, Ruffino M, Stella G, Fazio F, Cappa SF, Perani D (2004): Regional reductions of gray matter volume in familial dyslexia. *Neurology* 63:742–745.
- Brown WE, Eliez S, Menon V, Rumsey JM, White CD, Reiss AL (2001): Preliminary evidence of widespread morphological variations of the brain in dyslexia. *Neurology* 56:781–783.
- Buchsbaum MS, Buchsbaum BR, Chokron S, Tang C, Wei TC, Byne W (2006): Thalamocortical circuits: fMRI assessment of the pulvinar and medial dorsal nucleus in normal volunteers. *Neurosci Lett* 404:282–287.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013): Power failure: Why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14:365–376.
- Casanova, C. (2004): The visual functions of the pulvinar. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (pp. 592–608). Cambridge, USA: The MIT Press.
- Casanova MF, Araque J, Giedd J, Rumsey JM. (2004): Reduced brain size and gyrification in the brains of dyslexic patients. *J Child Neurol* 19:275–281.
- Castles A, Coltheart M (1993): Varieties of developmental dyslexia. *Cognition* 71:231–255.
- Crosson B (1999): Subcortical mechanisms in language: Lexical-semantic mechanisms and the thalamus. *Brain Cogn* 40:414–438.
- Díaz B, Hintz F, Kiebel SJ, von Kriegstein K (2012): Dysfunction of the auditory thalamus in developmental dyslexia. *Proc Natl Acad Sci USA* 109:13841–13846.
- Eckert MA, Leonard CM, Richards TL, Aylward EH, Thomson J, Berninger VW (2003): Anatomical correlates of dyslexia: Frontal and cerebellar findings. *Brain* 126:482–494.
- Eckert MA, Leonard CM, Wilke M, Eckert M, Richards T, Berninger VW (2005): Anatomical signatures of dyslexia in children: Unique information from manual and voxel based morphometry brain measures. *Cortex* 41:304–315.
- Eliez S, Rumsey JM, Giedd JN, Schmitt JE, Patwardhan AJ, Reiss AL (2000): Morphological alteration of temporal lobe gray matter in dyslexia: An MRI study. *J Child Psychol Psychiatry* 41: 637–644.
- Evans TM, Flowers DL, Napoliello EM, Eden GF (2014): Sex-specific gray matter volume differences in females with developmental dyslexia. *Brain Struct Funct*. 219:1041–1054.
- Finch AJ, Nicolson RI, Fawcett AJ (2002): Evidence for a neuroanatomical difference within the olivo-cerebellar pathway of adults with dyslexia. *Cortex* 38:529–539.
- Focke NK, Helms G, Kaspar S, Diederich C, Tóth V, Dechent P, Mohr A, Paulus W (2011): Multi-site voxel-based morphometry—Not quite there yet. *Neuroimage* 56:1164–1170.
- Frye RE, Liederman J, Malmberg B, McLean J, Strickland D, Beauchamp MS (2010): Surface area accounts for the relation of gray matter volume to reading-related skills and history of dyslexia. *Cereb Cortex* 20:2625–2635.
- Fusar-Poli P, Radua J, Frascarelli M, Mechelli A, Borgwardt S, Di Fabio F, Biondi M, Ioannidis JPA, David SP (2014): Evidence of reporting biases in voxel-based morphometry (VBM) studies of psychiatric and neurological disorders. *Hum Brain Mapp*, 35: 3052–3065.
- Galaburda AM, Kemper TL (1979): Cytoarchitectonic abnormalities in developmental dyslexia: A case study. *Ann Neurol* 6: 94–100.
- Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N (1985): Developmental dyslexia: Four consecutive patients with cortical anomalies. *Ann Neurol* 18: 222–233.
- Galaburda AM, Menard MT, Rosen GD (1994): Evidence for aberrant auditory anatomy in developmental dyslexia. *Proc Natl Acad Sci USA* 91:8010–8013.
- Giguere M, Goldman-Rakic PS (1988): Mediodorsal nucleus (Ariall, laminar, and tangential distribution of afferents and efferents in the frontal lobe of Rhesus monkeys). *J Comp Neurol* 277: 195–213.
- Goldman JG, Manis FR (2013): Relationships among cortical thickness, reading skill, and print exposure in adults. *Sci Stud Read* 17:163–176.
- Haar S, Berman S, Behrmann M, Dinstein I: Anatomical abnormalities in Autism? *Cereb Cortex* (in press). doi:10.1093/cercor/bhu242.
- Hadzibeganovic T, van den Noort M, Bosch P, Perc M, van Kralingen R, Mondt K, Coltheart M (2010): Cross-linguistic neuroimaging and dyslexia: A critical view. *Cortex* 46:1312–1316.
- Hartwigsen G, Baumgaertner A, Price CJ, Koehnke M, Ulmer S, Siebner HR (2010): Phonological decisions require both the left and right supramarginal gyri. *Proc Natl Acad Sci USA* 107: 16494–16499.
- Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE (2004): Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage* 22:676–687.
- He Q, Xue G, Chen C, Chen C, Lu ZL, Dong Q (2013): Decoding the neuroanatomical basis of reading ability: A multivoxel morphometric study. *J Neurosci* 33:12835–12843.
- Heim S, Tschierse J, Amunts K, Wilms M, Vossel S, Willmes K, Grabowska A, Huber W (2008): Cognitive subtypes of dyslexia. *Acta Neurobiol Exp (Wars)* 68:73–82.
- Hoelt F, Meyler A, Hernandez A, Juel C, Taylor-Hill H, Martindale JL, McMillon G, Kolchugina G, Black JM, Faizi A, Deutsch GK, Siok WT, Reiss AL, Whitfield-Gabrieli S, Gabrieli JD (2007): Functional and morphometric brain dissociation between dyslexia and reading ability. *Proc Natl Acad Sci USA* 104:4234–4239.
- Humphreys P, Kaufmann WE, Galaburda AM (1990): Developmental dyslexia in women: Neuropathological findings in three patients. *Ann Neurol* 28:727–738.
- Ioannidis JP (2005): Why most published research findings are false. *PLoS Med* 2:e124.
- Ioannidis JP (2011): Excess significance bias in the literature on brain volume abnormalities. *Arch Gen Psychiatry* 68:773–780.
- Jacquier-Roux M, Valdois S, Zorman M, Lequette C, Pouget G. 2005. *L'Odédys, outil de dépistage des dyslexies*. Grenoble: Laboratoire cogni-sciences, IUFM de Grenoble.
- Jamadar SD, Fielding J, Egan GF (2013): Quantitative meta-analysis of fMRI and PET studies reveals consistent activation in fronto-striatal-parietal regions and cerebellum during anti-saccades and prosaccades. *Front Psychol* 4:749.

- Jednoróg K, Altarelli I, Monzalvo K, Fluss J, Dubois J, Billard C, Dehaene-Lambertz G, Ramus F (2012): The influence of socioeconomic status on children's brain structure. *PLoS One* 7:e42486.
- Jednoróg K, Gawron N, Marchewka A, Heim S, Grabowska A (2014): Cognitive subtypes of dyslexia are characterized by distinct patterns of grey matter volume. *Brain Struct Funct* 219:1697–1707.
- Jenner A, Rosen GD, Galaburda AM (1999): Neuronal asymmetries in the primary visual cortex of dyslexic and non-dyslexic brains. *Ann Neurol* 46:189–196.
- Joubert S, Beauregard M, Walter N, Bourgouin P, Beaudoin G, Leroux JM, Karama S, Lecours AR (2004): Neural correlates of lexical and sublexical processes in reading. *Brain Lang* 89:9–20.
- Krafnick AJ, Flowers DL, Luetje MM, Napoliello EM, Eden GF (2014): An investigation into the origin of anatomical differences in dyslexia. *J Neurosci* 34:901–908.
- Kronbichler M, Wimmer H, Staffen W, Hutzler F, Mair A, Ladurner G (2008): Developmental dyslexia: Gray matter abnormalities in the occipitotemporal cortex. *Hum Brain Mapp* 29:613–625.
- Küspert P, Schneider W. 1998. Würzburger Leise Leseprobe (WLLP). Göttingen: Hogrefe.
- Landerl K, Ramus F, Moll K, Lyytinen H, Leppänen PH, Lohvansuu K, O'Donovan M, Williams J, Bartling J, Bruder J, Kunze S, Neuhoff N, Tóth D, Honbolyó F, Csépe V, Bogliotti C, Iannuzzi S, Chaix Y, Démonet JF, Longeras E, Valdois S, Chabernaud C, Delteil-Pinton F, Billard C, George F, Ziegler JC, Comte-Gervais I, Soares-Boucaud I, Gérard CL, Blomert L, Vaessen A, Gerretsen P, Ekkebus M, Brandeis D, Maurer U, Schulz E, van der Mark S, Müller-Myhsok B, Schulte-Körne G (2013): Predictors of developmental dyslexia in European orthographies with varying complexity. *J Child Psychol Psychiatry* 54:686–694.
- Lee H, Devlin JT, Shakeshaft C, Stewart LH, Brennan A, Glensman J, Pitcher K, Crinion J, Mechelli A, Frackowiak RS, Green DW, Price CJ (2007): Anatomical traces of vocabulary acquisition in the adolescent brain. *J Neurosci* 27:1184–1189.
- Lefavrais P. 1967. Test de l'aluette: manuel. Paris: les éditions du centre de psychologie appliquée.
- Lindgren SD, De Renzi E, Richman LC (1985): Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Dev* 56:1404–1417.
- Linkersdörfer J, Lonnemann J, Lindberg S, Hasselhorn M, Fiebach CJ (2012): Grey matter alterations co-localize with functional abnormalities in developmental dyslexia: An ALE meta-analysis. *PLoS One* 7:e43122.
- Livingstone MS, Rosen GD, Drislane FW, Galaburda AM (1991): Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proc Natl Acad Sci USA* 88:7943–7947.
- Maisog JM, Einbinder ER, Flowers DL, Turkeltaub PE, Eden GF (2008): A meta-analysis of functional neuroimaging studies of dyslexia. *Ann N Y Acad Sci* 1145:237–259.
- Marchewka A, Kherif F, Krueger G, Grabowska A, Frackowiak R, Draganski B; The 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. *Human Brain Mapp* 35:1865–1874.
- Menghini D, Hagberg GE, Petrosini L, Bozzali M, Macaluso E, Caltagirone C, Vicari S. (2008): Structural correlates of implicit learning deficits in subjects with developmental dyslexia. *Ann N Y Acad Sci* 1145:212–221.
- Monzalvo K, Fluss J, Billard C, Dehaene S, Dehaene-Lambertz G (2012): Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status. *Neuroimage* 61:258–274.
- Paulesu E, Démonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, Frith U (2001): Dyslexia: Cultural diversity and biological unity. *Science* 291:2165–2167.
- Pernet C, Andersson J, Paulesu E, Demonet JF (2009a): When all hypotheses are right: a multifocal account of dyslexia. *Hum Brain Mapp* 30:2278–2292.
- Pernet CR, Poline JB, Demonet JF, Rousselet GA (2009b): Brain classification reveals the right cerebellum as the best biomarker of dyslexia. *BMC Neurosci* 10:67.
- Posner MI, Raichle ME (1995): Precis of images of mind. *Behav Brain Sci* 18:327–383.
- Preston JL, Frost SJ, Mencl WE, Fulbright RK, Landi N, Grigorenko E, Jacobsen L, Pugh KR (2010): Early and late talkers: School-age language, literacy and neurolinguistic differences. *Brain* 133:2185–2195.
- Pugh KR, Frost SJ, Sandak R, Landi N, Rueckl JG, Constable RT, Seidenberg MS, Fulbright RK, Katz L, Mencl WE (2008): Effects of stimulus difficulty and repetition on printed word identification: An fMRI comparison of nonimpaired and reading-disabled adolescent cohorts. *J Cogn Neurosci* 20:1146–1160.
- Pugh KR, Landi N, Preston JL, Mencl WE, Austin AC, Sibley D, Fulbright RK, Seidenberg MS, Grigorenko EL, Constable RT, Molfese P, Frost SJ (2013): The relationship between phonological and auditory processing and brain organization in beginning readers. *Brain Lang* 125:173–183.
- Ramus F. (2003): Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Curr Opin Neurobiol* 13:212–218.
- Ramus F, Ahissar M (2012): Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cogn Neuropsychol* 29:104–122.
- Raschle NM, Chang M, Gaab N (2011): Structural brain alterations associated with dyslexia predate reading onset. *Neuroimage* 57:742–749.
- Repscher S, Grande M, Heim S, van Ermingen M, Pape-Neumann J (2012): Entwicklung parallelisierter Wortlisten zur Verlaufsdiagnostik bei dyslektischen Kindern, Sprache, Stimme, Gehör 36:33–39.
- Richardson FM, Price CJ (2009): Structural MRI studies of language function in the undamaged brain. *Brain Struct Funct* 213:511–523.
- Richlan F, Kronbichler M, Wimmer H (2011): Meta-analyzing brain dysfunctions in dyslexic children and adults. *Neuroimage* 56:1735–1742.
- 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.
- Scanlon C, Mueller SG, Tosun D, Cheong I, Garcia P, Barakos J, Weiner MW, Laxer KD (2011): Impact of methodologic choice for automatic detection of different aspects of brain atrophy by using temporal lobe epilepsy as a model. *AJNR Am J Neuroradiol* 32:1669–1676.
- Seymour PH, Aro M, Erskine JM (2003): Foundation literacy acquisition in European orthographies. *Br J Psychol* 94:143–174.
- Silani G, Frith U, Demonet JF, Fazio F, Perani D, Price C, Frith CD, Paulesu E (2005): Brain abnormalities underlying altered

- activation in dyslexia: A voxel based morphometry study. *Brain* 128:2453–2461.
- Siok WT, Perfetti CA, Jin Z, Tan LH (2004): Biological abnormality of impaired reading is constrained by culture. *Nature* 431:71–76.
- Siok WT, Niu Z, Jin Z, Perfetti CA, Tan LH (2008): A structural-functional basis for dyslexia in the cortex of Chinese readers. *Proc Natl Acad Sci USA* 105:5561–5566.
- Steinbrink C, Vogt K, Kastrup A, Müller HP, Juengling FD, Kassubek J, Riecker A (2008): The contribution of white and gray matter differences to developmental dyslexia: Insights from DTI and VBM at 3.0 T. *Neuropsychologia* 46:3170–3178.
- Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack CR Jr, Chen K, Ashburner J, Frackowiak RS (2008): Interpreting scan data acquired from multiple scanners: A study with Alzheimer’s disease. *Neuroimage* 39:1180–1185.
- Stoodley CJ, Schmahmann JD (2010): Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex* 46:831–844.
- Stoodley CJ, Stein JF (2011): The cerebellum and dyslexia. *Cortex* 47:101–116.
- Turkeltaub P, Eden G, Jones K, Zeffiro T (2002): Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16:765–780.
- Vinckenbosch E, Robichon F, Eliez S (2005): Gray matter alteration in dyslexia: converging evidence from volumetric and voxel-by-voxel MRI analyses. *Neuropsychologia* 43:324–331.
- World Health Organization (2008): International statistical classification of diseases and related health problems—tenth revision, 2nd ed. Geneva, Switzerland: World Health Organization.
- Ziegler JC, Perry C, Ma-Wyatt A, Ladner D, Schulte-Körne G (2003): Developmental dyslexia in different languages: Language-specific or universal? *J Exp Child Psychol* 86:169–193.
- Ziegler JC, Bertrand D, Tóth D, Csépe V, Reis A, Faísca L, Saine N, Lyytinen H, Vaessen A, Blomert L (2010): Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. *Psychol Sci* 21:551–559.