THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY
ACAMH

Journal of Child Psychology and Psychiatry *:* (2012), pp **–** doi:10.1111/jcpp.12029

Predictors of developmental dyslexia in European orthographies with varying complexity

Karin Landerl,[1,2,*] Franck Ramus,[3,*] Kristina Moll,[2,4] Heikki Lyytinen,[5,6] Paavo H. T. Leppänen,[5,6] Kaisa Lohvansuu,[5,6] Michael O'Donovan,[7] Julie Williams,[7] Jürgen Bartling,[8] Jennifer Bruder,[8] Sarah Kunze,[8] Nina Neuhoff,[8] Dénes Tóth,[9] Ferenc Honbolygó,[9] Valéria Csépe,[9] Caroline Bogliotti,[3] Stéphanie Iannuzzi,[10,11,15] Yves Chaix,[10,11] Jean-François Démonet,[11,12] Emilie Longeras,[3,13,14] Sylviane Valdois,[13,14] Camille Chabernaud,[3,15] Florence Delteil-Pinton,[15] Catherine Billard,[15] Florence George,[16] Johannes C. Ziegler,[17] Isabelle Comte-Gervais,[18] Isabelle Soares-Boucaud,[18,19] Christophe-Loïc Gérard,[20] Leo Blomert,[21] Anniek Vaessen,[21] Patty Gerretsen,[22] Michel Ekkebus,[22,24] Daniel Brandeis,[23,24,25,26] Urs Maurer,[23,24,27,28] Enrico Schulz,[23,24,29] Sanne van der Mark,[30] Bertram Müller-Myhsok,[31] and Gerd Schulte-Körne[8]

[1]Department of Psychology, University of Graz, Graz, Austria; [2]Department of Psychology, University of Salzburg, Salzburg, Austria; [3]Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, CNRS, EHESS, Paris, France; [4]Department of Psychology, University of York, York, UK; [5]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland; [6]Agora Center, University of Jyväskylä, Jyväskylä, Finland; [7]MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, Cardiff University, Cardiff, UK; [8]Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital Munich, Munich, Germany; [9]Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary; [10]Unité de Neurologie Pédiatrique, Hôpital des Enfants, Toulouse, France; [11]Inserm U825, Hôpital Purpan, Toulouse, France; [12]Leenards Memory Center, Centre Hospitalier Universitaire Vaudois, Université de Lausanne, Lausanne, Switzerland; [13]Laboratoire de Psychologie et NeuroCognition UMR 5105 CNRS, Université Pierre Mendès France, Grenoble, France; [14]Centre référent pour le diagnostic des troubles du langage et des apprentissages, Département de pédiatrie, CHU Nord, Grenoble, France; [15]Centre de Référence sur les Troubles des Apprentissages, Hôpital Bicêtre, Paris, France; [16]Centre de Référence des Troubles d'apprentissage, CHU Timone, Marseille, France; [17]Laboratoire de Psychologie Cognitive, CNRS, Aix-Marseille Université, Marseille, France; [18]Centre de Référence pour les Troubles des Apprentissages, Hospices Civils de Lyon, Hôpital E. Herriot, Lyon, France; [19]Centre Hospitalier Le Vinatier, Bron, France; [20]Service de Psychopathologie de l'enfant et de l'adolescent, Hôpital Robert Debré, APHP, Paris, France; [21]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands; [22]Regional Institute of Dyslexia, Maastricht, The Netherlands; [23]Department of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland; [24]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland; [25]Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland; [26]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany; [27]Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, NY, USA; [28]Institute of Psychology, University of Zurich, Zurich, Switzerland; [29]Department of Neurology, Technische Universität München, München, Germany; [30]MR-Center, University Children's Hospital, University of Zurich, Zurich, Switzerland; [31]Max Planck Institute of Psychiatry, Munich, Germany

Background: The relationship between phoneme awareness, rapid automatized naming (RAN), verbal short-term/working memory (ST/WM) and diagnostic category is investigated in control and dyslexic children, and the extent to which this depends on orthographic complexity. **Methods:** General cognitive, phonological and literacy skills were tested in 1,138 control and 1,114 dyslexic children speaking six different languages spanning a large range of orthographic complexity (Finnish, Hungarian, German, Dutch, French, English). **Results:** Phoneme deletion and RAN were strong concurrent predictors of developmental dyslexia, while verbal ST/WM and general verbal abilities played a comparatively minor role. In logistic regression models, more participants were classified correctly when orthography was more complex. The impact of phoneme deletion and RAN-digits was stronger in complex than in less complex orthographies. **Conclusions:** Findings are largely consistent with the literature on predictors of dyslexia and literacy skills, while uniquely demonstrating how orthographic complexity exacerbates some symptoms of dyslexia. **Keywords:** Dyslexia, phonology, orthography, cross-linguistic.

Introduction

Developmental dyslexia, i.e. a specific deficit in reading that cannot be accounted for by low IQ, poor educational opportunities or obvious sensory or

---

*Equal contributors.
Conflict of interest statement: No conflicts declared.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

neurological damage (World Health Organization, 2008), affects about 3–7% of the population (Lindgren, De Renzi, & Richman, 1985). Research has made tremendous progress in specifying biological and environmental factors associated with this disorder. A central environmental factor that has been identified to influence reading acquisition and dyslexia is the particular orthography that the child is acquiring. All orthographies depict the sound structure of the language they represent, but there is considerable variability in how transparent this relationship is for the learner and how consistently orthographic symbols represent the sounds of a particular language. Both theoretical conceptions (Katz & Frost, 1992; Ziegler & Goswami, 2005) and empirical evidence suggest that transparent orthographies with high symbol–sound consistency are acquired more easily than complex and opaque orthographies with a high proportion of inconsistent and irregular spellings. The most impressive evidence for the impact of orthographic complexity on reading development came from a large European network involving 13 different alphabetic orthographies (Seymour, Aro, & Erskine, 2003). At the end of first grade, reading accuracy was close to ceiling in consistent orthographies (Italian, Icelandic, Norwegian, Spanish, German, Dutch, Finnish), while children acquiring more complex orthographies (English, Danish, French) were still struggling. Complex and opaque mapping systems such as English orthography cause particular problems not only to the young learner, but also to dyslexic individuals (Landerl, Wimmer, & Frith, 1997; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Korne, 2003).

An important research question that arises from the well-documented orthographic differences in early and deficient written language processing is to what extent the cognitive mechanisms underlying reading acquisition and dyslexia might vary as well. English-based research has identified verbal-phonological processing as the central cognitive predictor of typical as well as dyslexic reading development (Vellutino, Fletcher, Snowling, & Scanlon, 2004), but there have been claims that focusing on the complex English orthography may have led to an overestimation of the relevance of phonological processing (Share, 2008). In transparent orthographies, even children who start reading acquisition with deficient phonological skills may be able to understand the mappings between spoken and written language if they are simple enough. Moreover, a simple and transparent representation of the phonological structure may help children to overcome early deficits, even more so if formal reading instruction is strongly phonics-based as is the case in many consistent orthographies.

Recently, a number of large-scale cross-linguistic studies on typical reading acquisition in different orthographies addressed this question. This approach can certainly not eliminate all methodological problems that are inherent in comparisons across different educational, cultural and language backgrounds, but findings are still easier to interpret within such designs than between studies that are carried out in different orthographic systems independently. Ziegler et al. (2010) investigated 1263 2nd graders in five orthographies with increasing degrees of complexity (Finnish, Hungarian, Dutch, Portuguese, French) and indeed found support for the hypothesis that phonology may be less relevant in consistent orthographies as the impact of phonological awareness (PA) was weaker. Nevertheless, PA was significantly associated with reading accuracy and speed in all orthographies and was the strongest concurrent predictor in all orthographies except in the highly transparent Finnish writing system, where vocabulary was the strongest predictor of word reading speed and predicted reading accuracy equally strongly as phoneme deletion. Another well-established predictor of reading, rapid automatized naming (RAN, see Kirby, Georgiou, Martinussen, & Parrila, 2010 for a recent review) did not show marked differences across orthographies and overall showed surprisingly moderate associations with reading. Two reasons may explain this atypically low RAN-reading relationship. First, Ziegler et al. (2010) used sequential naming of pictured objects and there is evidence that alphanumeric RAN tasks (letters, digits) show a stronger relationship with reading than such nonalphanumeric versions (e.g. Bowey, McGuigan, & Ruschena, 2005).

Second, the strength of the RAN-reading association may be relatively weak among Ziegler et al.'s sample of young readers and may increase later in reading development. This is suggested by another recent cross-language study by Vaessen et al. (2010) which focused on the concurrent prediction of reading fluency in three orthographies with increasing complexity (Hungarian, Dutch, Portuguese). Findings indicated a shift of cognitive mechanisms underlying reading fluency during development. In Grades 1 and 2, the association of PA and RAN (objects, letters and digits combined) with reading fluency was largely comparable, while in Grades 3 and 4, RAN was more strongly associated with reading fluency than PA. Importantly, Vaessen et al. (2010) confirmed that cognitive mechanisms underlying reading were similar across the three alphabetic orthographies, but again, the association of reading with PA, but not with RAN or verbal short-term memory was modulated by orthographic complexity.

Moll et al. (submitted)[1] examined whether the cross-linguistic findings resulting from these large-scale European studies could be extended to

---

[1] The analysis by Moll et al. (submitted) is based on the typical readers in the NEURODYS project and has 954 participants in common with the current control group

English, which is assumed to be the most inconsistent alphabetic orthography (Share, 2008). Altogether, 1,062 elementary school children learning to read in English, French, German, Hungarian or Finnish participated. Phonological skills (awareness and short-term memory) accounted for significant variance in reading accuracy and spelling in all orthographies while RAN was the best concurrent predictor of reading fluency. No major language differences were found between patterns of concurrent prediction. However, overall the regression models accounted for more variance in English than in the other orthographies. Interestingly, it was RAN rather than PA that showed a stronger association to literacy measures in English than in the other orthographies. However, a marked outlier-position of English orthography could not be confirmed.

Identifying the cognitive mechanisms underlying reading is of particular relevance in the context of dyslexia. How accurately can the predictor measures that have been identified for typical development differentiate between typical and dyslexic readers? Is the quality of prediction similar across orthographies or are there orthography-dependent differences? These questions were addressed in the context of NEURODYS, an EU-FP6 network focusing on the neurobiological and neurocognitive foundations of dyslexia in different alphabetic orthographies covering a broad range of consistency (English, French, Dutch, German, Hungarian, Finnish).

Methods

Rationale

Large-scale cross-linguistic comparisons are faced with critical methodological problems concerning (a) selection of adequate tasks to measure the relevant cognitive constructs (b) lack of a common metric adequately describing linguistic and orthographic differences and (c) possible differences in the diagnosis of dyslexia between countries. There are enormous differences in how dyslexia is diagnosed in the various national school and health care systems and also in the kind of support systems that are available to dyslexic individuals (Ise et al., 2011). A major advantage of NEURODYS is that within the project, the same ICD-10 (World Health Organization, 2008) based criteria were applied: Children with more general learning, attentional or neurological problems and children whose first language was not the instructional language were not admitted to the study. Reading was assessed by language-specific standardized word reading tests. Dyslexic readers had to perform more than 1.25 *SD*s below grade level which is a pragmatic compromise between the standard criteria of −1 and −1.5 *SD*s that are widely applied in research and clinical practice.

The lack of a common metric to describe orthographic and linguistic differences is a general problem for cross-linguistic studies. Although there is notable agreement on where to place particular writing systems on a continuum of orthographic complexity (e.g., Borgwaldt, Hellwig, & De Groot, 2005; Caravolas, 2005; Seymour et al., 2003), the adequate levels of description are still under discussion and their quantification is a future enterprise for psycholinguistic research. For the present analysis, we classified the six participating orthographies into three categories of orthographic complexity that are based on feedforward and feedback consistency of grapheme-phoneme and phoneme-grapheme correspondences. The three groups are largely inspired by the classification provided by Seymour et al. (2003) (see Table S1 in the online supplement), and are consistent with the data on word recognition in that study. English and French comprised the highest level of orthographic complexity with inconsistencies in grapheme-phoneme as well as phoneme-grapheme correspondences (low feedforward and feedback consistency). Dutch and German represented a medium complexity level as these two Germanic languages have comparable orthographic structures with highly consistent grapheme-phoneme correspondences (high feedforward consistency), but less consistent phoneme-grapheme correspondences (low feedback consistency). Finnish displayed the lowest level of orthographic complexity as it has nonambiguous 1:1 relationships between letters and sounds with equally high feedforward and feedback consistency. Hungarian was not included in the study of Seymour et al., but its linguistic and orthographic properties are similar to Finnish.

The three levels of orthographic complexity are also consistent with the analysis of word-initial letter-to-phoneme mappings provided by Borgwaldt et al. (2005) (see Table S1). Although that study did not include Finnish, this language falls clearly at the lower end of the complexity continuum with no more than 21 letter-to-phoneme mappings. However, we do not feel that the numbers in Table S1 represent a sufficiently valid measure of orthographic complexity to use them directly as a quantitative variable. For instance, Borgwaldt's entropy measure is based only on word onsets and is missing most of the irregularities in many languages, and therefore vastly underestimates some language differences. We therefore decided to use a simpler and more conservative categorical factor with three values, by grouping languages in pairs of similar orthographic complexity.

To identify similarities and differences in the concurrent prediction of dyslexia in orthographies with low, medium and high complexity, standard tasks measuring PA, verbal STM/WM and RAN were administered with all participants. Similarity across languages was relatively easy to achieve for verbal STM/WM as standardized versions of WISC digit span were available in each language. For the sake of similarity across languages, naming speed was measured by language-specific digit RAN tasks requiring children to sequentially name as quickly as possible lists of digits. A second RAN condition required sequential naming of pictured objects representing short high-frequency words in each language. With respect to PA, we decided to follow the example of earlier cross-linguistic studies (Caravolas, Volin, & Hulme, 2005; Vaessen et al., 2010; Ziegler et al., 2010) and administered phoneme deletion, thus ensuring reasonable comparability of findings across studies. Phoneme deletion is a standard paradigm and is included in standardized test batteries of dyslexia (e.g., Wagner, Torgesen, & Rashotte, 1999). Due to the large variability of languages involved, devising one and the same paradigm for all participants was not feasible. Specifying the linguistic structure of presented items across languages might have induced higher typicality in some languages than others (e.g., consonant clusters are atypical in Finnish, polysyllabic words are less typical in English). Thus, it was decided to leave the language-specific characteristics to individual partners who were advised to select words or nonwords with typical linguistic structure and to ask children to delete a specified phoneme.

In summary, two major methodological improvements could be made compared to earlier studies on dyslexia in different orthographies: First, due to extensive recruitment efforts, sample sizes for dyslexic and control groups in each language are clearly larger than in earlier studies. Second, the parallel structure of recruitment and components assessed resulted in comparable datasets across orthographies.

Participants

Participants[2] came from varying social backgrounds and were either identified in school or were specifically requesting clinical assessment of their reading problems. The same inclusion and exclusion criteria were applied in all partner countries. The following criteria were applied:

1. Age between 8 and 12; 11 years.
2. At least one and a half years of formal reading instruction in order to differentiate serious problems in reading acquisition from early delays that are not always persistent.
3. An age-appropriate scaled score of at least 7 on WISC Block design, and 6 on WISC Similarities.
4. An attention score within the 95th percentile of the age-appropriate norm, from either the Child Behavior Check-List (Achenbach, 2001) or the Conners questionnaire (Conners, 1973), filled by parents.
5. A parental questionnaire further asked a number of simple questions, on which basis we applied the following exclusion criteria:

(i) Hearing loss.
(ii) Uncorrected sight problems.
(iii) Test language not spoken by at least one parent since birth.
(iv) Child not schooled in test language.
(v) Child missed school for any period of 3 months or more.
(vi) Formal diagnosis of ADHD.
(vii) Child on medication for epilepsy or behavioural problems.

Inclusion criterion for dyslexic children:

More than 1.25 *SD*s below grade level on a standardized test of word reading.

Inclusion criterion for control children:

Less than 0.85 *SD*s below grade level on the same standardized word reading test.

Participant numbers per grade level are presented for each participating country in Table S2 in the online supplement. There were 1,138 control and 1,114 dyslexic children in total, based in eight different countries and speaking six different languages. In the control group, there were 598 boys and 540 girls, while in the dyslexic group there were 705 boys and 408 girls, which is consistent with the typical gender-ratio of dyslexia.

Measures

IQ was estimated based on one verbal and one nonverbal WISC subtest: Similarities and Block Design (Wechsler, 1992, 2003). Scaled scores ($M = 10$, $SD = 3$) were calculated based on age-specific norms.

Reading was assessed using language-appropriate standardized tests of word reading. All tests are listed in Appendix S1 in the online supplement. In all languages, but English, word lists were presented under a speeded instruction ('Read as quickly as possible without making mistakes'). The number of words read correctly per minute was converted into *Z*-scores based on grade-appropriate norms. In English, reading was not timed.

PA was assessed by a phoneme deletion task, requiring children to delete a specified phoneme from a word or nonword (e.g. 'Say/bli:k/without the/k/'). Number of correct responses was scored.

Naming speed was assessed via language-specific 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.

Verbal ST/WM was measured by WISC Digit span (forwards and backwards). Scaled scores ($M = 10$, $SD = 3$) were calculated based on age-specific norms.

Data analysis

Our main analytical approach is a multilevel logistic regression analysis, allowing nesting students within orthographic complexity groups. This allows us to estimate the effect of orthographic complexity on diagnostic category, as well as the effect of cognitive variables, both across and within orthographic complexity groups. Crucially, this analysis allows us to test to what extent the predictive value of cognitive variables differs between languages with different orthographic complexity. The exact procedures of data treatment and data analysis are described in the online supplement.

Results

Means, *SD*s, minima and maxima for each variable and each group are reported separately for the three levels of orthographic complexity in Table S3 in the online supplement. Age was lower in the low than in the medium complexity group and again lower in the medium than in the high complexity group. Within the high complexity group, dyslexic children were on average 3 months older than control children. Table S2 indicates an overrepresentation of dyslexic children in the higher grade levels. Therefore, grade

---

[2] There is an overlap of 43 Dutch participants (all controls) and 76 Hungarian participants (32 dyslexic and 44 controls) with Ziegler et al. (2010) and 69 Dutch and 178 Hungarian children with Vaessen et al. (2010).

level was controlled for in the following statistical analysis. Dyslexic children's scores on Block Design and Similarities were systematically lower than those of the control children, but average compared to the norm. The standardized word reading measure was used as sole group selection variable and the large group differences on Phoneme deletion, RAN and Digit span confirm that these are relevant predictor measures for dyslexia.

Multilevel logistic regressions of diagnostic category

The age range in the full European sample was unusually large including children from Grades 2–7. Therefore, we reran the analyses with an age-limited sample including Grades 3, 4 and 5, thereby eliminating the age differences between orthographic complexity groups (116, 117 and 116 months for low, medium and high complexity). The intraclass correlation of 0.0395 (95% CI 0.0197–0.067) for all students and 0.040 (0.016–0.075) for the age-limited sample (Grades 3–5), indicates that about 4% of the variance of the trait is attributable to the nesting variable orthographic complexity level. The median odds ratio (MOR) was estimated to be 1.423 (95% CI 1.278–1.588), for the full sample and 1.425 (95% CI 1.247–1.638) for children in Grades 3, 4 and 5, meaning that in median a given child is 40% more likely to be dyslexic in high- than in low complexity orthographies. Further inclusion of country as an additional random effect did not lead to a significant increase of the model likelihood ($\chi^2(35) = 30.782$ and 15.027 for the full and age-limited samples, respectively, $p > .1$), providing some empirical support for the validity of our three levels of orthographic complexity. Hence, the variable country was not retained in the model.

Next, Phoneme Deletion, RAN digits and Digit Span were introduced as predictor variables. RAN digits rather than RAN pictures was included because this measure is more similar across languages and correlations with reading tended to be stronger (RAN digits: .53, RAN pictures: .47, controlling for age). Grade, Gender, Block Design and Similarities were introduced to control for the group differences in these measures. The corresponding odds ratios and *p*-values derived from the Wald statistics are presented in Table 1 for the full sample and in Table S4 in the online supplementary for the age-limited sample (Grades 3–5). As expected, Phoneme Deletion, RAN digits and Digit Span were reliable predictors of dyslexia status. Gender and Similarities also had a significant influence on dyslexia status. The probability of being dyslexic is multiplied by about 2.6 (1/0.354) for each point decrease (in *z* units) of Phoneme deletion and 2.8 for RAN. A child with Phoneme Deletion = −1 and RAN = −1 has therefore roughly seven times the risk of being dyslexic than with a score of 0 on both variables (with odds ratio converted to relative risks assuming a dyslexia prevalence of 5%). On the other hand one *z*-unit of digit span only increases the relative risk by 41%.

We did, however, note significant differences in the estimates for some predictors. Tables 1 and S4 show significant evidence for heterogeneity for Grade, Phoneme Deletion and RAN digits. This heterogeneity, equivalent to an interaction effect between orthographic complexity groups and these predictors, prompted us to also perform analyses separately for each orthographic complexity subgroup (see Table 2). The results of this analysis for all children showed that the effect sizes of Phoneme Deletion and RAN increased with orthographic complexity. Differences for Phoneme Deletion were significant for low versus medium complexity groups ($Q = 8.14$, $df = 1$, $p = .004$), high versus low complexity groups ($Q = 13.92$, $df = 1$, $p = < .001$), and of borderline significance for medium versus high complexity groups ($Q = 3.68$, $df = 1$, $p = .055$). For RAN, the heterogeneity was driven by the low complexity group, which showed significantly higher odds ratios than both the medium ($Q = 14.56$, $df = 1$, $p < .001$) and high complexity groups ($Q = 6.28$, $df = 1$, $p = .012$), indicating that its impact is weaker in the low than in the other two complexity groups (see Figure 1). In all three orthographic complexity groups, Digit Span and verbal IQ were also significant, but more moderate predictors. In addition, grade level was of moderate predictive relevance in the low and high complexity group, Block Design in the medium complexity group and

Table 1 Multilevel logistic model for full sample (n = 1,998)

		Confidence interval				
	Odds ratio	Lower	Upper	*t*-value	$r (>\|t\|)$	het
Grade	1.007	0.886	1.143	0.100	.921	<.001
Gender	0.706	0.565	0.882	−3.067	.002	.223
Block design	1.040	0.992	1.090	1.632	.103	.584
Similarities	0.885	0.849	0.922	−5.822	<.001	.603
Phoneme deletion	0.354	0.308	0.408	−14.529	<.001	<.001
RAN digits	0.356	0.311	0.407	−15.08	<.001	<.001
Digit span	0.694	0.616	0.783	−5.963	<.001	.131

het = *p*-value of heterogeneity.

Table 2 Multilevel logistic models separately for each level of orthographic complexity

	Odds ratio	Confidence interval Lower	Confidence interval Upper	*z*-value	Pr (>|*z*|)
Low complexity (*n* = 682)					
Grade	0.638	0.476	0.856	−3.004	.003
Gender	0.753	0.521	1.089	−1.506	.132
Block design	1.022	0.946	1.104	0.555	.579
Similarities	0.909	0.849	0.974	−2.715	.007
Phoneme deletion	0.481	0.388	0.596	−6.681	<.001
RAN digits	0.491	0.400	0.604	−6.744	<.001
Digit span	0.594	0.478	0.737	−4.731	<.001
Medium complexity (*n* = 932)					
Grade	0.972	0.781	1.211	−0.250	.803
Gender	0.830	0.584	1.180	−1.040	.299
Block design	1.096	1.018	1.181	2.419	.016
Similarities	0.881	0.824	0.941	−3.737	<.001
Phoneme deletion	0.307	0.244	0.386	−10.095	<.001
RAN digits	0.285	0.228	0.356	−11.071	<.001
Digit Span	0.773	0.644	0.929	−2.746	.006
High complexity (*n* = 384)					
Grade	1.455	1.107	1.912	2.693	.007
Gender	0.317	0.159	0.635	−3.243	.001
Block design	0.961	0.827	1.116	−0.520	.603
Similarities	0.865	0.769	0.975	−2.385	.017
Phoneme deletion	0.187	0.116	0.303	−6.836	<.001
RAN digits	0.262	0.169	0.404	−6.040	<.001
Digit span	0.639	0.450	0.906	−2.515	.012

Cox & Snell R^2 = 0.294, 0.401 and 0.489 for low, medium and high complexity; Nagelkerke R^2 = 0.394, 0.537 and 0.672 for low, medium and high complexity; AUC = 0.817, 0.877 and 0.929 for low, medium and high complexity.

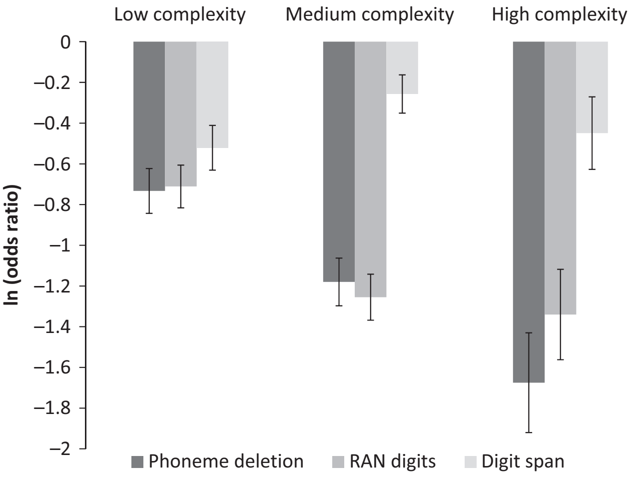

Figure 1 Estimates (ln OR) and their 95% confidence limits per orthographic complexity group for Phoneme Deletion, RAN Digits and Digit Span respectively

gender in the high complexity group. This heterogeneity can be illustrated by noting that one *z*-unit of phoneme deletion multiplies the probability of dyslexia by 2 (relative risk of 1.97 from the odds ratio of 1/0.481) in low complexity orthographies, it multiplies it by almost 3 (relative risk of 2.92 from the odds ratio of 1/0.307) in medium- and by more than 4 (relative risk of 4.39 from the odds ratio of 1/0.187) in high complexity orthographies.

Results for the age-limited sample (Grades 3–5) are reported in Table S4 and S5 of the online supplementary and are consistent with the full-sample analysis. Differences for Phoneme Deletion were significant for high versus low complexity groups (Q = 6.73, df = 1, p = .009), for high versus medium/low complexity combined (Q = 5.32, df = 1, p = .023) and low versus medium/high complexity combined (Q = 5.39, df = 1, p = .021) and showed a trend for medium versus high (Q = 2.93, df = 1, p = .086). For RAN, the low complexity group was again significantly different from both medium (Q = 18.84, df = 1, p < .001) and high complexity groups (Q = 8.01, df = 1, p = .005).

Thus, the direction of the observed heterogeneity effects pointed to more pronounced ORs in the higher orthographic complexity group indicating an increase of predictive capability of the model with increasing orthographic complexity. In accordance with this, the area under the ROC curve (AUC), reflecting the predictive power of the model across the whole range of classification criteria, was higher in the high complexity than in the low complexity group, with medium complexity in-between. Figure 2 shows the corresponding ROC curves. Aiming to test whether differences in AUCs between the orthographic complexity groups were meaningful, we ran a bootstrap analysis with 10,000 replicates. For each bootstrap sample, the full logistic regression analysis was run with all random and fixed effects and AUCs in each of the three complexity groups for each of the replicated were recorded. The 95% confidence intervals for all differences did not include zero, indicating that all three differences had a certain level of meaningfulness.

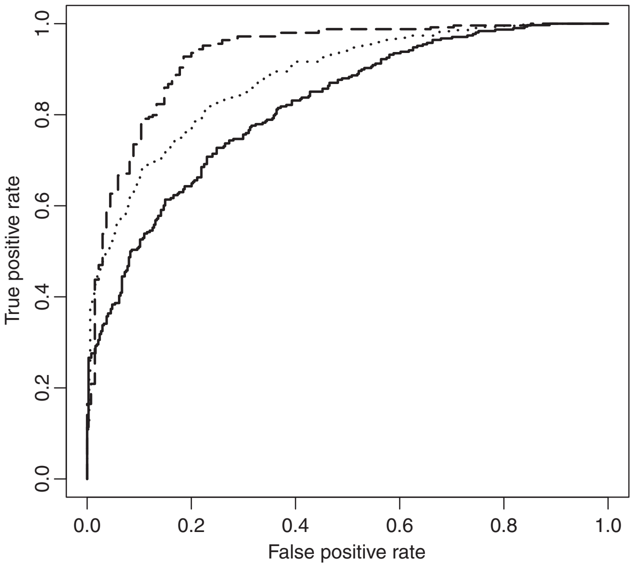

Figure 2 ROC curves (high complexity: dashed, medium complexity: dotted, low complexity: full)

Finally, we analysed whether predictive patterns would change when RAN pictures instead of RAN digits was introduced by rerunning the full-sample analysis and the analyses separately for each orthographic complexity level. Findings (Tables S6 and S7) were fully consistent with the earlier analysis.

Discussion

We investigated the cognitive and literacy skills of about 1,000 dyslexic and 1,000 control children aged 8–12, from eight European countries, learning to read in six different languages varying widely in terms of orthographic complexity. We specifically tested to what extent various cognitive variables predicted children's diagnostic status, and to what extent this differed between languages of varying orthographic complexity.

What are the predictors of dyslexia and are they similar for different levels of orthographic complexity?

Standard predictors of reading skills and dyslexia were introduced in this large-sample analysis. Findings confirmed that phoneme deletion and RAN are strong concurrent predictors of developmental dyslexia, while verbal ST/WM and general verbal abilities played a significant, but comparatively minor role.

Interestingly, our statistical model classified more participants correctly when orthography was more complex. This may be a consequence of the larger variance in phonological and reading skills in high than in low complexity orthographies. Indeed, despite normalization of all variables, dyslexic children often reached more extreme negative z-scores in high than in low complexity languages. In low complexity orthographies, reading achievement is generally higher (Seymour et al., 2003), even for dyslexic children (Landerl et al., 1997). Thus, variance in reading skills is reduced, and so is variance in phonological skills, given the influence of the former on the latter. The little variance that remains to be explained, amplified by rescaling into z-scores, may be much noisier and reflect more idiosyncratic factors in low than in high complexity orthographies, where variance in phonological skills can show its full impact on reading skills and on dyslexia.

Specifically, the impact of phoneme deletion and RAN was stronger in more than in less complex orthographies. No such heterogeneity was found for verbal ST/WM. Phoneme deletion as a standard measure of phoneme awareness is generally seen as an important predictor of dyslexia (Vellutino et al., 2004). Ziegler et al. (2010) showed for young readers (Grade 2) in five alphabetic orthographies that the predictive power of phoneme deletion increases with the degree of orthographic complexity and in the present study, this important finding could be confirmed for dyslexia. Thus, there is accumulating evidence in favour of Share's (2008) opacity by transparency hypothesis.

Our finding that RAN is a strong predictor is at odds with Ziegler et al. (2010), but fits well with the rest of the literature (Kirby et al., 2010). We have introduced both RAN digits and RAN pictures in our regression models. Both were reliable predictors of diagnostic status at all orthographic complexity levels. Thus, this study does not confirm earlier assumptions that processes involved in alphanumeric RAN (naming letters and digits) are different from nonalphanumeric task versions like naming colours or objects (Bowey et al., 2005). Similar to Phoneme Deletion, RAN also showed a stronger impact on relatively more complex orthographies. Thus, earlier claims that RAN may be a stronger predictor in orthographies with low compared to high complexity, as in such orthographies the variance in reading skills is usually determined by reading fluency rather than accuracy (Kirby et al., 2010), are not supported by the current study.

Limitations and caveats

The present cross-linguistic dyslexia study covers an unprecedented number of languages representing a large range of orthographic complexity. Nevertheless, our conclusions remain limited to the alphabetic orthographies included, which all use the Latin alphabet and a left-to-right writing direction. Any generalization to other alphabetic orthographies, and all the more so to nonalphabetic orthographies would be tentative.

In this study, the constructs of interest, PA, RAN and verbal ST/WM were assessed by only one measure each. Having several tasks per construct would obviously have increased construct validity. Unfortunately, it was not feasible to significantly increase testing time across all countries in such a large-scale project.

As in any cross-linguistic study, a major issue is the extent to which linguistic tasks designed in different languages tap similar cognitive processes and similar levels of difficulty. One possibility is to choose languages that are sufficiently close (such as English and German) and to match all the material (e.g., Landerl et al., 1997; Ziegler et al., 2003). When a broader range of languages is studied, this is not an option. There is simply no way to design equivalent lists of words or nonwords across languages as different as English and Finnish. Furthermore, even if the material was matched, the difficulty of certain tasks (such as word reading or phoneme deletion) might not be, as this is partly dependent on the orthographic complexity of the language. Thus, it is inevitable that in this study some tasks tap different levels of performance and thus have different sensitivities in different languages. This issue is most crucial with respect to the role of phoneme deletion. Indeed, it has been suggested that in languages with transparent orthographies, both reading accuracy and PA are easily acquired, even by dyslexic children, so that the former is a minor issue, and the latter is of minor importance for reading acquisition and dyslexia. Nevertheless, we found that PA was a significant predictor of dyslexia, even in the most transparent languages. Thus, to the extent that there is variation in performance in PA (and there is, even in Finnish), this variation is still meaningful in terms of the prediction of diagnostic category. Even in a language where dyslexic children reach near-perfect performance on phoneme deletion, the fact that their performance is slightly less perfect than that of controls' predicts dyslexia status.

The present cross-sectional study can only provide suggestive information regarding causality. Our use of the word 'predictor' is strictly statistical, i.e., the extent to which knowing the value of one variable allows to predict the value of another variable. We do not imply that the pattern of predictions observed here would necessarily hold across different time points. Our present findings are most likely to reflect bidirectional causality between predictors and reading skills. Longitudinal studies on the prediction of reading and dyslexia in different orthographies will be seminal to specify causality (see, for example, Caravolas et al., 2012; Furnes & Samuelsson, 2010; Georgiou, Parrila, & Papadopoulos, 2008).

Finally, an obvious limitation of the current study is that at present, comparable statistical descriptions of the languages and orthographies involved that would allow methodologically adequate quantification of orthographic differences are unavailable. Certainly, the absence of such statistics cannot mean that cross-linguistic studies cannot produce valuable results. The European network NEURODYS created the unique opportunity to analyse unusually large samples of dyslexic and control children with diagnostic and assessment procedures that were parallelized as much as possible across various orthographies. Our rough categorization of orthographic complexity based on feedforward/feedback consistency certainly underestimates the differences between the orthographies involved and can only be a first and cautious step towards the important enterprise to fully understand the implications of orthographic and linguistic structure on reading development and dyslexia.

Supporting information

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

Appendix S1 Standardized word reading tests applied in the six languages for group selection

Table S1 Rankings of orthographic complexity according to Seymour et al. (2003) and Borgwaldt et al. (2005)

Table S2 Number of participants per country, per grade and per group

Table S3 Mean, standard deviation, minimum and maximum for all variables, separately for each level of orthographic complexity

Table S4 Multilevel logistic model for age-limited sample (Grades 3, 4 and 5, n = 1,486)

Table S5 Multilevel logistic models for age-limited sample (Grades 3, 4 and 5), separately for each level of orthographic complexity

Table S6 Multilevel logistic model including RAN pictures (n = 1,998)

Table S7 Multilevel logistic models including RAN pictures, separately for each level of orthographic complexity

Acknowledgements

This project was funded by the EU FP6 grant to the Neurodys consortium, Agence Nationale de la Recherche, Ville de Paris, Austrian Science Fund (18351-B02), SNSF grant 32-108130 and Stiftung für wissenschaftliche Forschung der Universität Zürich. We thank Gary Hill, Silke Streiftau, Laure Bricout, Sophie Jery, Geneviève Lina, Christine Nossent, Corinne Bruchet, Lionel Collet, Marie-Christine Mouren.

Correspondence

Karin Landerl, Department of Psychology, University of Graz, Universitaetsplatz 2, 8010 Graz, Austria, Email: karin.landerl@uni-graz.at; Franck Ramus, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. Email: franck.ramus@ens.fr

Key points

- Phoneme awareness and rapid automatized naming are the main cognitive predictors of dyslexia in the six languages studied, across all levels of orthographic complexity.
- Their predictive power increases with orthographic complexity.
- However, phoneme awareness and rapid automatized naming have roughly equal relative importance in the prediction of dyslexia, and this does not differ between languages with different levels of orthographic complexity.
- More general verbal abilities play a significant, but more minor role, and a constant one across levels of orthographic complexity.

References

Achenbach, T. (2001). *Child behavior check-list.* Burlington: ASEBA.

Borgwaldt, S.R., Hellwig, F.M., & De Groot, A.M.B. (2005). Onset entropy matters: Letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*, 211–229.

Bowey, J.A., McGuigan, M., & Ruschena, A. (2005). On the association between serial naming speed for letters and digits and word-reading skill: towards a developmental account. *Journal of Research in Reading*, *28*, 400–422.

Caravolas, M. (2005). The nature and causes of dyslexia in different languages. In M.J. Snowling, & C. Hulme (Eds.), *The science of reading* (pp. 336–355). Oxford: Blackwell.

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., … & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*, 678–686.

Caravolas, M., Volin, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal of Experimental Child Psychology*, *92*, 107–139.

Conners, C.K. (1973). Rating scales for use in drug studies with children. *Psychopharmacology Bulletin*, *9*, 24–84.

Furnes, B., & Samuelsson, S. (2010). Predicting reading and spelling difficulties in transparent and opaque orthographies: A comparison between Scandinavian and US/Australian children. *Dyslexia*, *16*, 119–142.

Georgiou, G.K., Parrila, R., & Papadopoulos, T.C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, *100*, 566–580.

Ise, E., Blomert, L., Bertrand, D., Faísca, L., Poulakanaho, A., Saine, N.L., … & Schulte-Körne, G. (2011). Support systems for poor readers: Empirical data from six EU member states. *Journal of Learning Disabilities*, *44*, 228–245.

Katz, L., & Frost, R. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost, & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier.

Kirby, J.R., Georgiou, G.K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly*, *45*, 341–362.

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*, 315–334.

Lindgren, S.D., De Renzi, E., & Richman, L.C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development*, *56*, 1404–1417.

Moll, K., Ramus, F., Barting, J., Bruder, J., Kunze, S., Neuhoff, N., … & Landerl, K. (Submitted). Cognitive mechanisms underlying reading and spelling development in five European orthographies.

Seymour, P.H.K., Aro, M., & Erskine, J.M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.

Share, D.L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin*, *134*, 584–615.

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology*, *102*, 827–842.

Vellutino, F.R., Fletcher, J.M., Snowling, M.J., & Scanlon, D.M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1999). *Comprehensive Test of Phonological Processing (CTOPP).* Austin, TX: PRO-ED.

Wechsler, D. (1992). *The Wechsler Intelligence Scale for children* (3rd edn). London: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for children* (4th edn). San Antonio: The Psychological Corporation.

World Health Organization. (2008). *International statistical classification of diseases and related health problems – Tenth revision* (2nd edn). Geneva: World Health Organization.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., … & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559.

Ziegler, J.C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *13*, 3–29.

Ziegler, J.C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Korne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, *86*, 169–193.

Accepted for publication: 02 October 2012

Supplementary online material

Data analysis

For the word reading variable and the WISC subtests (Block Design, Similarities, and Digit Span), standardised scores were used for statistical analysis. For phoneme deletion and RAN, raw scores were converted into z-scores based on the control group's mean and standard deviation, within each country and each grade level. Phoneme deletion showed particularly high skewness in some languages (English: -.88; French: -2.98; Dutch: -2.49; German: -1.22; Hungarian: -1.19; Finnish: -1.48). In order to reduce these distortions, we applied the following procedure: Each variable in each country was converted into ranks, then rescaled on a 0-100 interval, then applied the normal distribution function to convert them into z-scores, and then finally rescaled based on the control group's mean and standard deviations. Although skewness was less marked for RAN digits (ranging from .01 in French to -.85 in Finnish) and RAN pictures (ranging from .04 in French to -1.42 in Finnish), the same normalisation procedure was applied. In other words, the thus normalized variables were inverse rank-quantile normalized and scaled to have equal variance in each of the orthographic complexity groups in order to more fully approach normality and homocedasticity desirables for the logistic regression, which help to find a more stable solution and avoid bias in the estimates.

We used a multilevel logistic regression analysis as implemented in the R-packages hglm (http://cran.r-project.org/web/packages/hglm/index.html, V1.2-2) and lme4 (http://cran.r-project.org/web/packages/lme4//index.html, V0.999999-0), allowing nesting students within orthographic complexity groups. We determined the intraclass correlation (ICC), which gives an estimate of the variance attributable to the nesting variable, following the equation $ICC = V_a/(V_a+\pi^2/3)$, where V_a is the variance attribued to the nesting variable. We also calculated the median odds-ratio (MOR) for the nesting variable as described by Merlo, Chaix, Yang, Lynch, & Råstam (2005). The 95% confidence intervals for the MOR and the ICC were calculated using a total of 10.000 bootstrap samples. In a second step we tested by means of a likelihood ratio test whether including a random effect for country would significantly increase the model likelihood.

Further variables were then entered into the model as fixed effects with odds ratios and confidence intervals calculated by standard procedures. We also allowed for heterogeneity of effects between orthographic subgroups by coding them as random effects in the hglm model. The test for heterogeneity was performed using the R-package rmeta (http://cran.r-project.org/web/packages/rmeta/index.html, V 2.1.6). This implements a Q-test for heterogeneity of effects. In our setting this is equivalent to allowing for interaction effects.

We investigated classifier performance assessed via receiver operating characteristics (ROC) curves using the R-package ROCR (http://cran.r-project.org/web/packages/ROCR/index.html, V 1.0-4) (Sing, Sander, Beerenwinkel, & Lengauer (2005). To assess whether differences in the area under the curve (AUC) values of the ROCs between the orthographic complexity groups were meaningful we again performed a bootstrap analysis with 10.000 bootstrap samples.

The age range in the full European sample was unusually large including children from Grades 2 to 7 and there were worries that the predictive pattern might vary with age. Therefore, we reran the regression model with a reduced sample including Grades 3, 4, and 5, thereby eliminating the age differences between orthographic complexity groups (116, 117 and 116 months for low, medium and high complexity).

References

Merlo, J., Chaix, B., Yang, M., Lynch, J. &. Råstam, L. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiological Community Health, 59*, 1022–1028.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics, 21*, 3940-3941.

Appendix SA. Standardized word reading tests applied in the six languages for group selection.

Finnish:

Häyrinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse.* Helsinki: Psykologien Kustannus Oy.

Hungarian:

Csépe, V., Tóth, D., Vaessen, A., & Blomert, L. (2012). *3DM Hungarian version.* Manuscript in preparation.

Dutch:

Blomert, L., & Vaessen, A. (2009). *Differentiaal Diagnostiek van Dyslexie: Cognitieve analyse van lezen en spellen* [Dyslexia Differential Diagnosis: Cognitive analysis of reading and spelling]. Amsterdam: Boom Test Publishers.

German:

Moll, K. & Landerl, K. (2010). *SLRT-II – Verfahren zur Differentialdiagnose von Störungen der Teilkomponenten des Lesens und Schreibens*. Bern: Huber.

French:

Jacquier-Roux, M., Valdois, S., & Zorman, M. (2005). *Odédys: Outil de dépistage des dyslexiques (version 2).* Grenoble: Laboratoire Cognisciences.

English:

Elliot, C., Smith, P., & McCulloch, K. (1997). *British Ability Scales II.* Windsor: NFER-Nelson.

Table S1.

Rankings of orthographic complexity according to Seymour et al. (2003) and Borgwaldt et al. (2005)

Languages ranked by increasing orthographic depth [a]	Word reading accuracy (% correct) for familiar content words at the end of 1st grade [b]	Number of word-initial letter-to-phoneme mappings[c]
Finnish	98.17	<=21
Hungarian		30
German	97.25	77
Dutch	92.66	80
French	72.47	94
English	32.59	105

[a] from Seymour et al. (2003), Table 1, except for Hungarian, added here based on similar linguistic considerations.
[b] from Seymour et al. (2003), Table 5.
[c] from Borgwaldt et al. (2005), Table 2, except for Finnish.

Table S2.

Number of Participants per Country, per Grade and per Group.

Number of participants			Group	
Language	Country	Grade	Control	Dyslexic
Low orthographic complexity languages				
Finnish	Finland	2	68	95
		3	131	123
		4	1	2
		Total	200	220
Hungarian	Hungary	2	62	40
		3	58	27
		4	55	25
		Total	175	92
Medium orthographic complexity languages				
German	Germany	3	110	80
		4	92	78
		Total	202	158
	Switzerland	2	7	3
		3	10	7
		4	2	2
		5	25	17
		Total	44	29
	Austria	2	42	31
		3	83	52
		4	62	57
		Total	187	140
Dutch	Netherlands	2	10	27
		3	39	43
		4	34	20
		5	2	8

		Total	85	98
High orthographic complexity languages				
French	France	2	8	6
		3	23	28
		4	26	49
		5	23	36
		6	15	25
		7	7	14
		Total	102	158
English	United Kingdom	3	10	17
		4	25	36
		5	31	43
		6	39	54
		7	38	69
		Total	143	219
			1138	1114

Table S3. *Mean, Standard Deviation, Minimum and Maximum for all Variables, Separately for Each Level of Orthographic Complexity.*

GROUP		AGE[a]	GRADE	BLOCKS[b]	SIMILAR[b]	WDREADFL[c]	DIGSPAN[b]	DIGRAN	PICRAN	PHONDEL
Low orthographic complexity										
Control	Mean	112.46	2.80	11.31	12.19	.29	10.17	.01	.02	.03
(N = 374)	SD	8.08	.68	2.63	3.08	.73	2.67	1.01	1.00	1.01
	Minimum	96	2	7	6	-1.33	3	-2.88	-2.67	-2.73
	Maximum	133	4	19	19	2.86	19	2.92	2.69	2.74
Dyslexic	Mean	112.73	2.64	10.64	10.61	-1.80	8.27	-.79	-.81	-.92
(N = 311)	SD	8.74	.63	2.55	2.75	.44	2.39	1.08	.95	1.14
	Minimum	97	2	7	6	-3.09	1	-3.53	-3.39	-3.81
	Maximum	143	4	19	19	-1.25	18	2.12	2.42	2.38
Medium orthographic complexity										
Control	Mean	114.40	3.35	11.01	12.52	.49	10.32	.02	.04	.02
(N = 515)	SD	10.04	.75	2.52	2.92	.82	2.54	.99	1.01	.98
	Minimum	96	2	7	6	-.83	1	-2.97	-3.01	-3.20
	Maximum	152	5	19	24	3.09	19	2.83	2.75	2.98
Dyslexic	Mean	114.84	3.35	11.08	11.67	-1.79	9.00	-1.22	-1.09	-1.26
(N = 420)	SD	10.78	.80	2.42	2.67	.46	2.48	1.11	1.09	1.14
	Minimum	96	2	7	6	-3.45	2	-4.38	-4.18	-4.48
	Maximum	146	5	18	19	-1.25	19	2.32	2.44	1.79

High orthographic complexity										
Control	Mean	124.22	4.76	11.01	12.62	.98	9.74	.06	.04	.13
(N = 136)	SD	14.88	1.27	2.37	2.86	1.23	3.18	.98	.98	.96
	Minimum	98	3	7	6	-.73	4	-4.26	-4.26	-2.68
	Maximum	155	7	19	19	6.89	19	2.43	2.37	2.16
Dyslexic	Mean	128.24	4.94	10.46	10.71	-1.84	7.53	-1.22	-1.28	-1.38
(N = 253)	SD	15.68	1.28	2.20	2.84	.42	2.69	1.05	1.16	1.01
	Minimum	96	3	7	6	-3.00	2	-4.26	-4.26	-4.21
	Maximum	155	7	19	19	-1.25	15	1.39	2.69	1.13

[a] months

[b] scaled scores (M=10. SD=3)

[c] z-scores based on national norms

All other variables: z-scores based on control groups

Table S4

Multilevel Logistic Model for Age Limited Sample (Grades 3, 4, and 5, n = 1486)

	odds	Confidence Interval				
	ratio	lower	upper	t-value	Pr(>|t|)	het
Grade	0.979	0.789	1.214	-0.195	.845	.048
Gender	0.701	0.542	0.908	-2.694	.007	.218
Block Design	1.055	0.999	1.115	1.934	.053	.957
Similarities	0.894	0.852	0.938	-4.602	<.001	.173
Phoneme Deletion	0.337	0.285	0.398	-12.851	<.001	.027
RAN Digits	0.343	0.293	0.402	-13.288	<.001	<.001
Digit Span	0.708	0.615	0.814	-4.848	<.001	.838

Note: het = *p*-value of heterogeneity

Table S5

Multilevel Logistic Models for Age-Limited Sample (Grades 3, 4, and 5), Separately for Each Level of Orthographic Complexity

	Odds	Confidence Interval			
	ratio	lower	upper	z_value	Pr(>|z|)
Low Complexity (*n* = 416)					
Grade	0.378	0.189	0.755	-2.755	.006
Gender	0.685	0.426	1.102	-1.560	.119
Block Design	1.064	0.964	1.175	1.239	.215
Similarities	0.914	0.841	0.994	-2.101	.036
Phoneme Deletion	0.414	0.309	0.554	-5.934	<.001
RAN Digits	0.537	0.411	0.702	-4.546	<.001
Digit Span	0.677	0.514	0.891	-2.783	.005
Medium Complexity (*n* = 815)					
Grade	1.090	0.807	1.472	0.561	.575
Gender	0.845	0.578	1.236	-0.868	.386
Block Design	1.071	0.989	1.161	1.682	.093
Similarities	0.870	0.808	0.938	-3.634	<.001
Phoneme Deletion	0.315	0.245	0.405	-9.005	<.001
RAN Digits	0.267	0.209	0.341	-10.621	<.001
Digit Span	0.741	0.606	0.906	-2.927	.003
High Complexity (*n*=255)					
Grade	1.361	0.774	2.393	1.070	.285
Gender	0.250	0.107	0.583	-3.208	.001

Block Design	1.020	0.845	1.232	0.211	.833
Similarities	0.917	0.796	1.055	-1.214	.225
Phoneme Deletion	0.178	0.096	0.330	-5.478	<.001
RAN Digits	0.243	0.143	0.414	-5.205	<.001
Digit Span	0.742	0.479	1.150	-1.335	.182

Note: Cox & Snell R^2 = 0.287, 0.409, and 0.498 for low, medium, and high complexity; Nagelkerke R^2 = 0.387, 0.548, and 0.681 for low, medium, and high complexity; AUC = 0.814, 0.881, and 0.930 for low, medium, and high complexity

Table S6

Multilevel Logistic Model Including RAN Pictures (n = 1998)

	Odds	Confidence Interval				
	ratio	lower	upper	t-value	Pr(>|t|)	het
Grade	0.968	0,850	1.101	-0.499	.618	<.001
Gender	0.858	0.687	1.071	-1.356	.175	.867
Block Design	1.069	1,020	1.121	2.785	.005	.294
Similarities	0,900	0.863	0.937	-5,070	<.001	.755
Phoneme Deletion	0.334	0.291	0.385	-15.309	<.001	<.001
RAN Pictures	0.382	0.335	0.436	-14.287	<.001	.038
Digit Span	0.718	0.637	0.809	-5.446	<.001	.095

Note: het = *p*-value of heterogeneity

Table S7

Multilevel Logistic Models Including RAN Pictures, Separately for Each Level of Orthographic Complexity

	Odds ratio	Confidence Interval		z_value	Pr(>|z|)
		lower	upper		
Low Complexity (*n* = 682)					
Grade	0.631	0.471	0.845	-3.092	.002
Gender	0,880	0.607	1.276	-0.672	.501
Block Design	1.059	0,980	1.143	1.455	.146
Similarities	0.919	0.858	0.984	-2.413	.016
Phoneme Deletion	0.469	0.377	0.584	-6.802	<.001
RAN Pictures	0.467	0.377	0.578	-6.997	<.001
Digit Span	0.605	0.488	0,750	-4.574	<.001
Medium Complexity (*n* = 932)					
Grade	0.905	0.728	1.126	-0.894	.371
Gender	0.996	0.706	1.405	-0.021	.984
Block Design	1.122	1.043	1.206	3.095	.002
Similarities	0.905	0.848	0.966	-2.987	.003
Phoneme Deletion	0.289	0.232	0.361	-10.946	<.001
RAN Pictures	0.335	0.271	0.413	-10.198	<.001
Digit Span	0.809	0.674	0,970	-2.288	.022
High Complexity (*n* = 384)					
Grade	1.421	1.081	1.868	2.518	.012
Gender	0.476	0.244	0.931	-2,170	.030

Block Design	0.966	0.834	1.119	-0.466	.641
Similarities	0,870	0.776	0.976	-2.369	.018
Phoneme Deletion	0.169	0.105	0.272	-7,320	<.001
RAN Pictures	0.308	0.205	0.462	-5.697	<.001
Digit Span	0.658	0.464	0.931	-2.361	.018

Note: Cox & Snell R^2 = 0.298, 0.380, and 0.477 for low, medium, and high complexity; Nagelkerke R^2 = 0.399, 0.508, and 0.656 for low, medium, and high complexity; AUC = 0.824, 0.866, and 0.925 for low, medium, and high complexity