Routledge
Taylor & Francis Group

Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency

Karin Landerl[a], H. Harald Freudenthaler[a], Moritz Heene[b], Peter F. De Jong[c], Alain Desrochers[d], George Manolitsis[e], Rauno Parrila [f], and George K. Georgiou[g]

[a]University of Graz; [b]Ludwig Maximilians University Munich; [c]University of Amsterdam; [d]University of Ottawa; [e]University of Crete; [f]Macquarie University; [g]University of Alberta

ABSTRACT

Although phonological awareness (PA) and rapid automatized naming (RAN) are confirmed as early predictors of reading in a large number of orthographies, it is as yet unclear whether the predictive patterns are universal or language specific. This was examined in a longitudinal study across Grades 1 and 2 with 1,120 children acquiring one of five alphabetic orthographies with different degrees of orthographic complexity (English, French, German, Dutch, and Greek). Path analyses revealed that a universal model could not be confirmed. When we specified the best-fitting model separately for each language, RAN was a consistent predictor of reading fluency in all orthographies, whereas the association between PA and reading was complex and mostly interactive. We conclude that RAN taps into a language-universal cognitive mechanism that is involved in reading alphabetic orthographies (independent of complexity), whereas the PA–reading relationship depends on many factors like task characteristics, developmental status, and orthographic complexity.

Phonological awareness (PA) and rapid automatized naming (RAN) have consistently been found to be closely associated with children's reading development explaining unique variance in children's reading skills above and beyond general factors like age and nonverbal IQ (Araújo, Reis, Petersson, & Faísca, 2015; Melby-Lervåg, Lyster, & Hulme, 2012; Norton & Wolf, 2012). PA refers to the ability to identify and manipulate phonological segments in spoken words, and RAN denotes the ability to name serial displays of letters, digits, pictured objects, or colors as quickly as possible.

PA is important as all orthographic systems represent phonological units in one way or another. Children with deficient access to the relevant phonological units will have problems to fully understand the mappings between a certain spoken language and its orthography and training of PA seems to improve reading outcomes (e.g., Suggate, 2016). An important and as yet unresolved issue is whether PA precedes reading acquisition or whether it evolves as a consequence or during the course of learning to read (Bradley & Bryant, 1985; Castles & Coltheart, 2004; Hulme, Snowling, Caravolas, & Carroll, 2005; Morais, Cary, Alegria, & Bertelson, 1979). Although studies with illiterate adults indicate that PA is a consequence of literacy learning (e.g., Morais et al., 1979; for a review, see Dehaene, Cohen, Morais, & Kolinsky, 2015), longitudinal studies have suggested the opposite (e.g., Caravolas et al., 2012). A major methodological limitation of longitudinal studies showing an

association of preschool PA with later reading is that early reading abilities at the first assessment point were typically not assessed. Thus, it is unclear whether the association between early PA and later reading results from children who already have basic reading skills at the first assessment and who perform well on the PA tasks. For example, Wimmer, Landerl, Linortner, and Hummer (1991) showed that at study onset, 16 of 23 early readers (about 70%) scored at ceiling in a phoneme awareness task, whereas only 10% of the 105 nonreaders reached such high scores. Neglecting the fact that a certain percentage of children can read at the onset of schooling may induce faulty interpretations of predictive patterns.

Although RAN is sometimes interpreted as also reflecting phonological processing (Savage, Pillay, & Melidona, 2007), Wolf and Bowers (1999) claimed that it constitutes a separate construct that is related to reading independently. The mechanisms underlying the well-documented RAN–reading relationship are less clear and a subject of debate (Jones, Snowling, & Moll, 2016; Lervåg & Hulme, 2009; Poulsen, Juul, & Elbro, 2015; Protopapas, Altani, & Georgiou, 2013b). The different perspectives largely converge in assuming that sequential naming mimics the timely integration of visual and verbal skills required during efficient word recognition and allows simultaneous processing of multiple stimuli presented in serial fashion, which explains why RAN exerts its strongest effects on reading fluency (Kirby, Georgiou, Martinussen, & Parrila, 2010).

In summary, PA and RAN reflect two distinctive verbal abilities that help children to comprehend and automatize the mappings between their spoken language and the writing system they acquire. Although significant associations between these predictors and reading have been reported in different orthographies (e.g., Dutch: De Jong, 2011; Finnish: Torppa et al., 2013; French: Plaza & Cohen, 2007; German: Moll, Fussenegger, Willburger, & Landerl, 2009; Greek: Protopapas, Altani, & Georgiou, 2013a; Spanish: Rodríguez, van den Boer, Jiménez, & de Jong, 2015; Chinese: Song, Georgiou, Su, & Shu, 2016), the majority of studies have been conducted with children acquiring the English writing system (see the recent meta-analysis by Araújo et al., 2015). These studies have typically investigated the prediction of reading accuracy, whereas in more consistent orthographies, reading fluency is the central dependent measure, as even very young and deficient readers tend to make few reading errors (e.g., Diamanti, Goulandris, Campbell, & Protopapas, 2018; Gangl et al., 2018; Wimmer, 2006; but see van Viersen et al., 2018). English orthography is atypically complex and has been demonstrated to be harder to acquire than other more transparent orthographies (Caravolas, Lervåg, Defior, Seidlová Málková, & Hulme, 2013; Frith, Wimmer, & Landerl, 1998; Seymour, Aro, & Erskine, 2003). Given the unique complexities of English orthography, it needs to be investigated whether the predictive patterns found for English can be generalized to other languages and their writing systems.

The orthographic depth hypothesis postulates that English is hard to learn because it is a deep orthography representing morphological rather than phonological units (Katz & Frost, 1992; see also Venezky, 1970). The fact that English spellings have over time evolved to represent morphology, etymology, and phonology in a complex combination (Bowers & Bowers, 2017) has resulted in a situation where most other alphabetic orthographies represent the phonological structure of spoken language in a more transparent and consistent way. Orthographies that represent the phonemic structure consistently make it relatively easy for young children to work out the mappings between graphemes and phonemes. English represents the phoneme level only inconsistently, whereas more consistent relationships can be found for larger phonological grain sizes like rime units. The psycholinguistic grain-size hypothesis (Ziegler & Goswami, 2005) proposes that identifying and applying these units of variable size and their complex print-to-sound mappings is more difficult than being able to consistently rely on the phonemic level.

Although a universal view of reading development would let us expect similar predictive patterns across orthographies, these cross-linguistic theories of reading development would predict that the relevance of, and interplay between, cognitive predictors may vary depending on the structure of the particular orthography. In particular, the variability in children's ability to access phonological units of the spoken language might account for more variance in deep orthographies with large grain sizes

than in shallow orthographies with small grain sizes. More specifically, early deficits in PA might be less detrimental in orthographies that present the phonological structure of language in a simple and transparent way. Indeed, Wimmer, Mayringer, and Landerl (2000) identified a sample of German-speaking children with a marked and specific deficit in PA at school entry who did not develop any reading problems later on. Thus, at least in transparent orthographies, PA may not be a necessary precursor of reading but may evolve mostly during the process of learning to read.

Cross-linguistic predictions are less clear for RAN. Proficient integration of visual and verbal processes is relevant for fluent reading in all orthographies, so similar predictive patterns might be expected. It has been proposed that RAN might be more relevant in transparent orthographies where reading accuracy is often at ceiling early on and reading fluency tasks are used to capture individual differences in reading (Kirby et al., 2010; Mann & Wimmer, 2002). However, reading fluency is important in English as well, although it has not always been assessed in earlier studies.

Empirical evidence on the predictors of reading acquisition in different orthographies comes from a range of languages. However, comparing findings across studies is often difficult because of differences in study design (cross-sectional or longitudinal), sample characteristics (random vs. dyslexic readers, age and educational background), predictor measures included, dependent measures included (word or nonword reading accuracy or speed, reading comprehension, spelling), and—last but not least—the language and writing system to be acquired. To examine whether the relationship of PA and RAN with reading are universal or language specific, cross-linguistic studies are of particular relevance.

Cross-sectional studies comparing several alphabetic orthographies provided empirical support for the hypothesis that PA may indeed be less relevant in consistent than in inconsistent orthographies (Landerl et al., 2013; Vaessen et al., 2010; Ziegler et al., 2010), whereas the RAN–reading relationship was either not modulated by orthography (Vaessen et al., 2010; Ziegler et al., 2010) or found to be stronger in inconsistent compared to more consistent orthographies (Georgiou, Aro, Liao, & Parrila, 2015; Landerl et al., 2013; Moll et al., 2014; see also the meta-analysis by Araùjo et al., 2015).

It is possible that the reduced relationship of PA with reading results from the higher consistency of an orthography: Intensive practice of systematic and highly reliable decoding procedures may induce sufficient PA even among children with serious reading difficulties. This is less of a concern for findings on the RAN–reading relationship, as RAN does not seem to be influenced by reading development (Lervåg & Hulme, 2009; Scalisi, Desimoni, & Di Vito Curmini, 2013; Wei, Georgiou, & Deng, 2015; but see Compton, 2003; Wolff, 2014).

Only a few cross-linguistic studies assessed PA and RAN before the onset of schooling and investigated their impact on later reading skills using longitudinal designs. In one of these studies, Georgiou, Torppa, Manolitsis, Lyytinen, and Parrila (2012) followed children's reading acquisition from preschool to end of Grade 2 in three languages (English, Greek, and Finnish). Whereas nonword decoding was predicted by preschool PA and RAN skills in English, early letter knowledge was the only significant predictor in the two more transparent orthographies Greek and Finnish.

Caravolas et al. (2012) also followed a cross-linguistic sample including English and more consistent alphabetic orthographies (Spanish, Czech, and Slovak). PA and RAN were significant predictors of first-grade reading in addition to preschool reading and preschool letter knowledge (whereas verbal short-term memory did not account for any unique variance). Caravolas et al. (2013) further assessed reading six times between kindergarten and the end of Grade 2. Their analysis confirmed earlier findings that reading development progresses more slowly in English than in the more consistent orthographies. In addition, predictive patterns of reading growth were similar across three orthographies (Slovak was no longer included), with the exception that early letter knowledge was less important in English than in Spanish and Czech. The authors argued that the developmental trajectories would progress faster in consistent orthographies due to their high reliability of letter-sound correspondences, whereas the underlying mechanisms would remain mostly universal, at least for alphabetic orthographies. A major limitation of this study was that cognitive predictors were

assessed only once at study onset, so that potential interactive relations with reading could not be identified.

Finally, a comprehensive cross-linguistic longitudinal assessment was based on the International Longitudinal Twin Sample contrasting reading acquisition in English with two more consistent Scandinavian alphabetic orthographies (Swedish and Norwegian). In their early reports (Furnes & Samuelsson, 2009, 2010), preschool PA predicted reading in the consistent orthographies only at the end of Grade 1, whereas it continued to be a significant predictor for the English language sample in Grade 2, and RAN predicted reading across orthographies. In spite of these early differences, Peterson et al. (2018) recently reported a language-universal model for cross-lagged relations between PA, RAN, and reading at prekindergarten (age 5), kindergarten (age 6), Grade 1, and Grade 4 (Grade 2 data were not included). Nevertheless, there were two significant language differences: the effect of prekindergarten literacy (i.e., letter knowledge and environmental print knowledge) on kindergarten PA was stronger in the Scandinavian group, whereas Grade 1 reading exerted a stronger influence on Grade 4 PA in the English language group. Overall, the cross-lagged correlations revealed an interactive relation of PA with reading throughout the study period but relatively low coefficients from PA to the two reading assessments in Grades 1 and 4. RAN was consistently related to reading at the next assessment point. Somewhat surprisingly, prekindergarten print knowledge predicted kindergarten RAN. The fact that prekindergarten RAN was the only RAN assessment that did not include a letter condition may explain why it did not fully explain kindergarten RAN so that tasks measuring early knowledge about print (including letters) showed an impact. There was also a (just about significant) prediction from kindergarten RAN to first-grade reading, whereas the very same coefficient for first-grade RAN to fourth-grade reading was not significant. So at least for reading development during school, Peterson et al. (2018) confirmed a unidirectional prediction from RAN to reading.

The present study

Findings on the predictive role of PA and RAN across orthographies are as yet inconclusive. Cross-sectional analyses mostly suggest differences between languages, particularly for PA, whereas longitudinal studies are more in line with a universal account of predictive patterns. So far, only one study (Peterson et al., 2018) has investigated interactive patterns in terms of cross-lagged relationships between these important predictors and reading during development. The present study aimed to contribute to this discussion by following children learning to read in five languages (English, French, German, Dutch, and Greek) longitudinally during the first stages of their reading development.

The five languages are written by alphabets with a large variability of orthographic consistency: In a comparative analysis of orthographic complexity based on the dual-route-cascaded model (Schmalz, Marinus, Coltheart, & Castles, 2015, Table 1), English was found to be characterized by a high number of multiletter rules and many irregular words. In French, the number of multiletter rules is even higher than in English, whereas the percentage of irregular word spellings is lower. Schmalz et al.'s (2015) analysis revealed that German and Dutch are clearly less complex than English and French, as they have much lower numbers of multiletter rules and lower percentages of irregular words than English. The main difference between German and Dutch is that the former has more context-sensitive letter-sound rules. Greek orthography was not included in this analysis, but its complexity is clearly more similar to German and Dutch than to English and French: Irregular words are rather exceptional and consistency of letter-sound rules is generally high (according to Seymour et al., 2003, higher than in German and Dutch), whereas sound-letter correspondences are less consistent, mostly for morphological reasons. Among the languages included in our study, Greek is the only one that is not written in Roman alphabet (see Verhoeven & Perfetti, 2017, for detailed descriptions of each of the five languages and orthographies).

In each language, reasonably large samples were followed during a critical period in their reading development: From the beginning of Grade 1, when formal reading instruction starts, to the end of Grade 2, when basic processes of decoding and word recognition are established (with an

Table 1. Means, Standard Deviations, and Reliability (Omega) for Reading, PA, and RAN Across Languages and Assessment Points

	English			French			German			Dutch			Greek		
	M	SD	ω	M	SD	ω	M	SD	ω	M	SD	ω	M	SD	ω
Reading[a]															
T1	16.31	14.49	0.86	13.61	10.89	0.96	2.90	6.66	0.94	8.02	6.50	0.96	6.44	9.96	0.96
T2	33.38	15.96	0.90	43.12	14.08	0.93	28.12	14.70	0.99	18.95	9.97	0.85	29.98	9.52	0.89
T3	48.46	14.22	0.93	63.37	13.29	0.86	43.95	15.53	0.84	33.88	11.37	0.87	43.74	11.63	0.87
PA[b]															
T1	9.91	4.35	0.86	10.76	5.93	0.85	6.66	4.47	0.85	9.91	4.51	0.84	4.81	5.36	0.91
T2	11.79	4.37	0.86	15.19	5.86	0.86	15.10	4.73	0.88	14.10	4.82	0.89	11.82	6.03	0.82
T3	15.05	5.00	0.88	18.63	4.13	0.79	17.85	4.03	0.84	18.75	3.86	0.74	15.74	5.40	0.83
RAN[c]															
T1	24.07	7.24	0.91	23.56	5.19	0.86	31.64	8.86	0.90	26.52	7.10	0.89	29.80	8.80	0.89
T2	21.78	5.54	0.89	21.47	4.10	0.83	23.03	5.38	0.86	23.53	5.69	0.89	22.02	4.76	0.69

Note. PA = phonological awareness; RAN = rapid automatized naming; T1 = beginning of Grade 1; T2 = end of Grade 1; T3 = end of Grade 2.
[a]Number of items read correctly in 1 min per condition (words, nonwords). [b]Number of correct responses per condition (words, nonwords; max. = 24). [c]Time in seconds per slide.

intermediate assessment at the end of Grade 1). This study period was chosen because we wanted to apply the very same measures of PA, RAN, and reading across all assessment points. Expanding the study period to younger (or older) children was likely to induce floor (or ceiling) effects.

The aim of the study was to test two models of the language-universal view of predictive patterns: First, we wanted to test a universal account of unidirectional prediction of PA and RAN for later reading. If this model could not be confirmed, we planned to test a more interactive model allowing for cross-lagged predictions for PA (but not RAN) and reading. In case we would fail to find any clear evidence for a universal account, we planned to specify the best-fitting models separately for each language in order to inspect whether we could confirm earlier evidence on stronger relations of PA and RAN in the more complex orthographies.

Method

Participants

The data used in the present study are part of a longitudinal project on reading and spelling development in five Indo-European alphabetic orthographies. The present sample consisted of 1,120 Grade 1 children followed until Grade 2. As in some of the participating countries (particularly Austria and Greece) usually no reading instruction or preparation is provided in preschool, the onset of primary school was the earliest possible point to start assessments. Of the participants, 172 children were native speakers of English recruited from six public elementary schools in Alberta, Canada (82 girls; M_{age} = 79.12 months at first assessment); 262 children were native speakers of French recruited from eight public elementary schools in Ottawa, Canada (136 girls; M_{age} = 78.12 months at first assessment); 343 children were native speakers of German recruited from nine public schools in Graz, Austria (177 girls; M_{age} = 79.11 months at first assessment); 114 children were native speakers of Dutch recruited from five public schools in the vicinity of Amsterdam (63 girls; M_{age} = 78.52 months at first assessment); and 229 were native speakers of Greek recruited from six public elementary schools in Crete, Greece (120 girls; M_{age} = 76.10 months at first assessment). Our participants were recruited on a voluntary basis (convenience sampling) and were assessed at the beginning (October/November) and end (April/May) of Grades 1 and 2 (only the German language sample was not assessed at the beginning of Grade 2). By the end of Grade 2, our sample consisted of 148 English-speaking (9% attrition), 240 French-speaking (9% attrition), 330 German-speaking (4% attrition), 106 Dutch-speaking (7% attrition), and 219 Greek-speaking (8% attrition) children. We also ran a

test for data missing completely at random (MCAR) using the R package "MissMech" (Jamshidian, Jalal, & Jansen, 2014). There was no sufficient evidence to reject MCAR at .05 significance level with a p value of .07 of the nonparametric test for MCAR. The children in each site came mostly from families of middle socioeconomic background (based on the location of the schools and on parents' education). None of these children were identified as having learning, emotional, or sensory disabilities. Informed consent was obtained from care-takers and schools before data collection and children gave their verbal consent. The study was conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki.

Measures

To make measures comparable across the five languages, special attention was paid to the task selection and development as well as test administration procedures (e.g., instructions, application of discontinuation rules). For several of these tasks, a similar laptop computer and shared program application (run by Empirisoft DirectRT, v. 2012) were used to present the stimuli and record children's responses.

Phonemic awareness

The widely used paradigm of Phoneme Elision was used to assess phonemic awareness: We applied Phoneme Elision with real words and with nonwords. Each task was designed so as to match items phonologically across languages. Both tasks included four practice items and 24 experimental items, all prerecorded by a native speaker of each language and presented through speakers plugged into a laptop computer. Children were presented with one item at a time, asked to repeat it, and then asked to remove a sound from it and say what was left. The items were presented in four blocks of six items. The blocks were ordered in increasing levels of difficulty: deletion of (a) a full syllable (e.g., car[pet]), (b) a simple onset consonant (e.g., [s]it), (c) a consonant from an onset cluster (e.g., [b]lend), and (d) a consonant from a cluster in the middle of a two-syllabic (non)word (e.g., wi[n]dow). A discontinuation rule of four errors in the block of six items was applied. A participant's score was the total number of correct responses.

Rapid automatized naming

Color and Digit Naming were assessed. Children were asked to name as fast as possible four colors (blue, red, green, and yellow) or digits (2, 4, 5, and 7) arranged in semirandom order in four rows of six. Each task was preceded by a practice trial to ensure that children knew the names of the colors/digits and then administered twice with the items arranged in a different order. A participant's score was the average time to name both cards in each task.

Reading fluency

The word reading efficiency and phonemic decoding efficiency subtests from the Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999) and similar forms that exist in the other languages (Brus & Voeten, 1979; Georgiou, Papadopoulos, Fella, & Parrila, 2012; Moll & Landerl, 2010; Van Den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994) were used to assess reading fluency. Children were asked to read lists of real words and pseudowords as fast and accurately as possible within a 60-s limit. In both tasks, the child's score was the total number of items read correctly within the specified time limit.

Procedure

In the current analysis, data from three assessment points are reported: beginning of Grade 1 (T1), end of Grade 1 (T2), and end of Grade 2 (T3). As the German language sample did not participate in the assessment beginning of Grade 2, it was not included in the current analysis. The same task battery was used across assessments.

All tasks were administered in a quiet room in the child's school by trained research assistants. The tests were administered in two sessions of about 30 min each. All tests were given in the same order, and all administration and scoring procedures were standardized across all children and all languages.

Statistical analysis

In a first step, modeling the relationships among the variables of interest within a latent variable framework was thoroughly explored but turned out not to be feasible because the covariance matrices of the latent variables were not positive definite in the German sample. Thus, it was decided to test path models based on manifest variables for each construct.

To examine the cross-lagged relations between PA, RAN, and reading we performed path analysis in lavaan (Rosseel, 2012) using full information maximum likelihood estimation. First, we tested a baseline model representing unidirectional effects from both PA and RAN to later reading. If this model could not be confirmed, we planned to test an extended model including cross-lagged relations between PA, RAN, and reading. In case we would fail to find clear evidence for a universal account, we planned to identify the best-fitting model separately for each language in order to inspect similarities and differences across languages.

To examine the fit of each model, we used a set of fit indices recommended by Hu and Bentler (1999) and Beauducel and Wittmann (2005), whereupon the comparative fit index (CFI) value should be higher than .95, and the root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) should be less than .05. Because previous research has shown that the suggested cutoff values cannot be generalized because their sensitivity to detect misspecifications also depends on parameters that are unrelated to model misspecification (see Greiff & Heene, 2017, for an overview), we also took the result of the chi-square model test into account when judging the overall model fit (see also Ropovik, 2015).

Results

Descriptive statistics

Correlations between the two measures of PA, RAN, and reading were significant in each language and at each assessment point (PA: rs between .614 and .833; RAN: rs between .471 and .787; reading: rs between .843 and .989, all ps < .001). Thus, they were combined into one average score per construct. Table 1 presents the descriptive statistics for the combined scores of PA, RAN, and reading at each assessment point across the five languages. Table 1 also presents reliability estimates according to Raykov's (2001) omega, which had a mean of .87 and a standard deviation of .06, and no major differences became evident between orthographies. All measures were sensitive to developmental changes, and there were no floor or ceiling effects. Note that at T1, clear differences in reading skills emerged: Although the English and French samples already showed relatively advanced reading skills, the children learning to read the more consistent orthographies showed much lower scores, reflecting cultural differences in reading preparation in preschool.

Results of path analysis

To test the universal account of predictive patterns, we specified a multigroup model in which each variable (PA, RAN, reading) predicted itself at the following assessment point and PA and RAN predicted reading at the following assessment point. The fit of this model was not satisfactory, $\chi^2(75) = 348.591$, $p < .001$, CFI = .948, SRMR = .099 and RMSEA = .127, and did not support the proposed universal account. Testing this model separately for each language also yielded unsatisfactory fit results as indicated by the fit indices and the chi-square model test (for the corresponding fit indices, see Table 2).

Table 2. FIT Indices for the Baseline Model, the Extended Model, and the Empirically Modified Models per Language

	χ^2 (df)	RMSEA	CFI	SRMR
Baseline model				
English	61.405 (15), $p < .001$	.134	.956	.100
French	56.424 (15), $p < .001$	.103	.968	.085
German	100.116 (15), $p < .001$	.127	.938	.093
Dutch	48.229 (15), $p < .001$	.139	.930	.105
Greek	82.417 (15), $p < .001$	.168	.934	.119
Extended model				
English	40.113 (13), $p < .001$	.110	.974	.062
French	26.716 (13), $p < .050$	.063	.989	.034
German	86.080 (13), $p < .001$	.126	.947	.081
Dutch	41.425 (13), $p < .001$	.138	.940	.073
Greek	39.378 (13), $p < .001$	.093	.974	.062
Modified models				
English	29.804 (13), $p < .050$	.087	.984	.054
French	20.935 (12), $p = .051$	.053	.993	.032
German	49.526 (12), $p < .010$	.094	.973	.051
Dutch	23.022 (15), $p = .084$	.068	.983	.087
Greek	31.588 (12), $p < .050$	.083	.981	.051

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Next, we tested an extended multigroup model in which we added predictions from reading to PA and RAN[1] at the following assessment point. Even though the fit of this model, $\chi^2(65) = 233.712$, $p < .001$, CFI = .968, SRMR = .063 and RMSEA = .107, was somewhat better than that of the first model, it was again not satisfactory. Testing this extended model separately for each language again yielded no satisfactory fit results as indicated by the fit indices and the chi-square model test (see Table 2).

Because none of the models just presented (except the extended model for French) provided satisfactory fit indices and the chi-square tests were always significant, we further carried out a search for possible misspecifications separately for each language starting from the extended universal model. For each language, the following model modification strategy was applied: (a) Nonsignificant paths were omitted from the model specification, and (b) a misspecification search according to Saris, Satorra, and Van Der Veld (2009) was conducted to detect possibly missing model parameters. semTools (semTools contributors, 2016) was used to conduct the misspecification search:

(a) A standardized path coefficient of .1 or greater and residual correlations of .10 or greater were considered as critical misspecifications. Power to detect such misspecifications was set to .75 as suggested by Saris et al.
(b) A parameter that (a) was flagged as misspecified and (b) was causally possible (i.e., only directional paths from *consecutive* measurement occasions besides residual correlations were allowed) was freely estimated in a second step and the model parameters were then reestimated.
(c) Step b was repeated until a satisfactory model fit according to the three-index strategy was obtained or no reasonable model modifications could be implemented according to the rationale outlined in Step b.
(d) Possibly misspecified path coefficients were given preference over residual correlations in the misspecification search because the former directly point to causal misspecifications among the variables used in our model, whereas the latter may point to unknown omitted variables not contained in the model.

The best-fitting models that resulted from this procedure for each of the five languages are presented in Figure 1. Consistently across all five languages, the best predictor for each of the three

---

[1]As there was no reason to assume that reading might have an impact on RAN, RAN was not included as a criterion measure at T3.
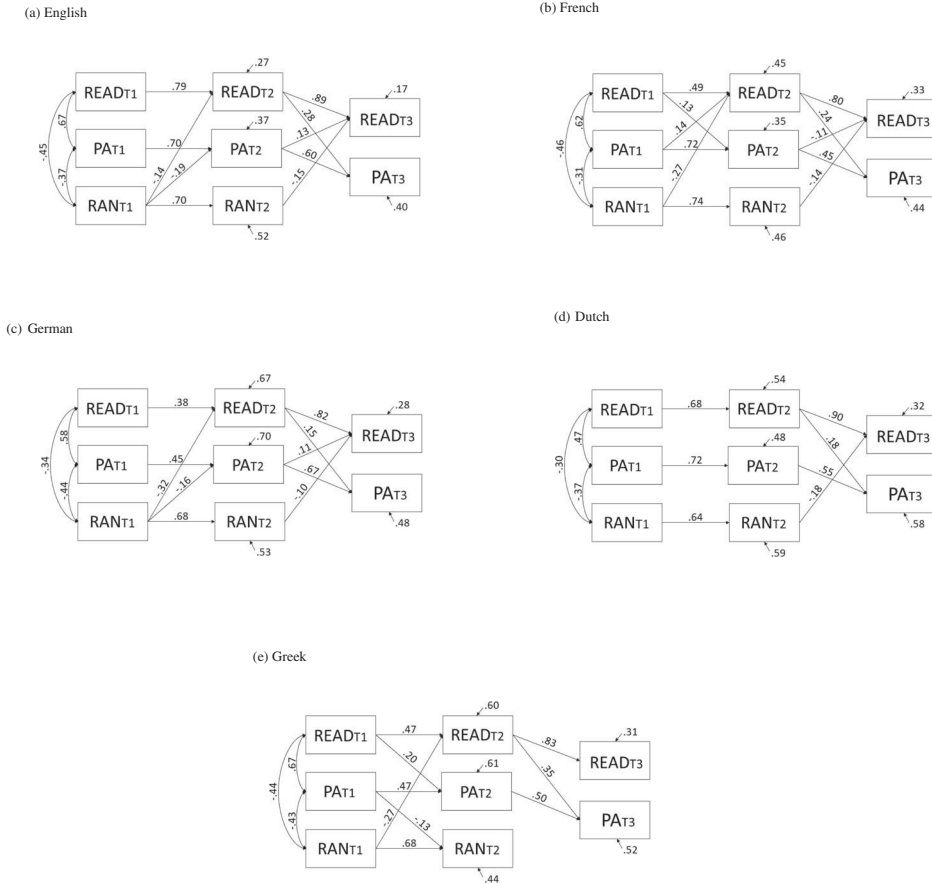
Figure 1. Best-fitting modified model separately for each language.

constructs was the same construct at the preceding assessment point, indicating stability in PA, RAN, and reading during the first two school years.

Surprisingly, the only language in which PA at T1 predicted reading at T2 was French, which shows an interactive pattern with PA predicting reading and at the same time reading predicting PA across all three assessment points. Greek was the only language in which reading at T1 unidirectionally predicted PA at T2. The other three orthographies (English, German, and Dutch) showed largely parallel development of PA and reading across Grade 1 (T1 to T2). From the end of Grade 1 to the end of Grade 2, reading predicted PA in all five language groups and PA predicted reading in English, German, and French but not in Dutch and Greek. Note that in French, PA at T2 showed a negative standardized regression weight. The correlational structure between PA and reading at T2, and between PA and reading at T3 revealed the existence of a negative suppression (the condition of $r_{\text{reading T2, PA T2}} > r_{\text{reading T3, PA T2}} / r_{\text{reading T3, reading T2}}$ was fulfilled with .55 > .36 / .79; see Friedman & Wall, 2005). As a consequence, PA acted as a variable suppressing variance in reading at T2 that was unrelated to the variance in reading at T3.

The pattern for RAN was more consistent: RAN predicted reading at the following assessment point in all orthographies, with only two exceptions: In Dutch, RAN at T1 was not significantly related to reading at T2, and in Greek, RAN at T2 was not significantly related to reading at T3. In addition, RAN at T1 exerted a predictive influence on PA at T2 in English and German and PA at T1 predicted RAN at T2 in Greek.

Discussion

The current study contributes to the research literature on predictive patterns of reading development across orthographies. Note that this is the first study that applied the very same predictor and reading measures in five orthographies over a 2-year period, allowing the investigation of cross-lagged relationships between measures. The obvious advantage of this approach was that children learning to read in different languages were assessed within the same study design. However, although tasks were matched as closely as possible, they may not always be fully comparable across languages. Furthermore, it is difficult to control for cultural differences in reading preparation and reading instruction between languages. In the current study, language differences were observable for all tasks at all assessment points. In particular, at the first assessment point, most of the German-speaking children were still nonreaders, whereas French- and English-speaking children had comparably advanced reading skills (with Dutch and Greek children somewhere in between). Differences between consistent and inconsistent orthographies in amount of reading preparation have been reported before (Caravolas et al., 2012; Furnes & Samuelsson, 2009; Georgiou, Parrila, & Papadopoulos, 2008; Mann & Wimmer, 2002). In transparent orthographies, there is no need to start early, as typically developing children quickly acquire the basics of reading during first grade. Accordingly, language differences were quite marked for PA as well as reading, whereas for RAN, mean scores were more similar across languages (particularly for T2 and T3). Obviously, these differences between language groups must be kept in mind when we turn to interpreting predictive patterns.

Our study clearly contributes to the increasing research literature indicating high stability of reading and reading-related skills during development (Caravolas et al., 2013; Hulslander, Olson, Willcutt, & Wadsworth, 2010; Landerl & Wimmer, 2008; Peterson et al., 2018; Peterson, Pennington, Olson, & Wadsworth, 2014). By far the best predictor of reading was always reading at the previous assessment point. PA and RAN also turned out to be constructs that were stable over time.

The critical question of the current study was whether we would find similar predictive patterns across orthographies. Important to note, two universal models based on standard views on the prediction of reading development were not confirmed by our cross-linguistic data set, suggesting that there were significant differences in predictive patterns between our five language groups. Note that age at the first assessment of predictor measures was clearly higher in our study than in the earlier studies (Caravolas et al., 2013; Caravolas et al., 2012; Peterson et al., 2018), in which the authors reported no language differences in the prediction of reading. It could therefore be argued that the current design may have been less sensitive to identify the very early patterns of prediction, as our participants were comparably advanced in their reading development at study onset. Although this is indeed an issue for the English and French subsamples, the initial reading skills of the German and Greek sample indicate that they were at the very start of reading development.

Arguably, our finding of significant language differences is influenced by the obvious differences in task performance. When we specified the best-fitting models separately for each language, predictive patterns turned out to be similar across languages for RAN but quite different for PA. The differential patterns of prediction of PA and RAN provide further evidence for the distinctness of their contribution to reading as proposed by Wolf and Bowers (1999).

RAN was a consistent predictor of reading in all orthographies (with two exceptions: In Dutch, RAN predicted reading from T2 to T3 but not between T1 and T2, and in Greek RAN predicted reading from T1 to T2, but not between T2 and T3). An influence of reading on RAN was not observed between T1 and T2, which confirms our decision to not include RAN as a criterion measure at T3 based on earlier assumptions of a unidirectional RAN–reading relationship (Lervåg & Hulme, 2009; Wei et al., 2015). Across the two first-grade assessments, we also observed a number of interactions between RAN and PA (RAN at T1 predicted PA at T2 in English and German and PA at T1 predicted RAN at T2 in Greek), showing that the two constructs are not completely

independent from each other (Wolf & Bowers, 1999). Although the current findings are not informative with respect to the mechanisms underlying the RAN–reading relationship, this consistent pattern across orthographies indicates that RAN taps a universal mechanism that is of similar relevance in learning to read across alphabetic orthographies, irrespective of differences in their complexity (see Georgiou et al., 2015, for a similar finding). The results of the current study also confirm earlier evidence that RAN is the best predictor of reading fluency across orthographies (e.g., Furnes & Samuelsson, 2009; Georgiou et al., 2008; Moll et al., 2014). It will be highly important to further investigate the underlying causal mechanisms.

No consistent pattern appeared for the PA–reading relationship: Only one language group (French) provided empirical support for the classic view that early PA is a specific predictor of reading; however, even in this language the relationship between PA and reading was clearly interactive, with associations going in both directions across the full study period. A similar interactive pattern emerged from the end of Grade 1 to the end of Grade 2 for English and German, whereas in Dutch and Greek, the prediction was unidirectional from reading to PA.

The findings in the German-speaking subsample seem particularly interesting, as these children were virtually nonreaders at the onset of our study. Of interest, the empirical findings of this subsample reflect a developmental perspective of the PA–reading relation that was presented by Wimmer et al. (1991): Most children seem to acquire PA in the course of learning to read. Studies that did not assess reading at the onset of data collection would have missed this important influence of early reading skills. In the long run, those children who do not develop PA quickly when practicing reading in an alphabetic writing system may then fall behind their classmates in reading.

It is possible that the prediction from PA at T1 to reading at T2 in the French-speaking sample simply reflects the very same pattern as the German sample between T2 and T3. French-speaking children already showed quite good reading skills at T1 and might better line up with the German-language sample at T2. English-speaking children, on the other hand, also had quite good reading skills at T1, but their PA–reading relation was more similar to German than to French. English and French are the two most complex orthographies in our language comparison, and it is interesting to see that there are nevertheless critical differences between the language-specific patterns of prediction. In Dutch, reading and PA also showed parallel development between T1 and T2, and only later on, from T2 to T3, reading impacted on PA. The Dutch sample was relatively small, and it is possible that a larger sample might have revealed a modest influence of T2 PA on T3 reading as was observed in German and English. Greek, the third consistent orthography, showed the clearest pattern with predictions only from reading to PA but not the other direction. Overall, we found little evidence for the "classic" view that PA predicts reading (Adams, 1990; Liberman, Shankweiler, & Liberman, 1989). Our findings rather support an "interactive" view of the PA–reading relationship suggesting that most children develop their phonological awareness in the context of reading acquisition, and not before (Castles & Coltheart, 2004; Wimmer et al., 1991).

By and large, the language differences observed in our study support the view that the relevance of early phonology increases with the complexity of the orthography that is acquired (see also Duncan et al., 2013). In Greek and Dutch, the most transparent orthographies in our study, PA did not predict reading, whereas reading at T2 predicted PA at T3. In German, an interactive pattern became apparent in Grade 2, and in French, this interactive pattern was evident right from the start. The most surprising finding of our study is perhaps that in English—arguably the most complex orthography in our study—PA at T1 did not predict reading at T2, when differences in early reading skills were controlled for. We cannot rule out that findings might have been different if we could have assessed this sample earlier in development. However, Peterson et al. (2018) also found only modest prediction from kindergarten and first-grade PA to the subsequent reading assessments. Thus, current cross-lagged designs reveal a different picture of the causal mechanisms underlying the PA–reading relationship than earlier, unidirectional models. Based on these two large-scale assessments (both studies are based on samples of more than 1,000 children), we conclude that the predictive power of PA for reading

may have been overestimated. Although PA skills may be an excellent and highly useful diagnostic indicator of problems in reading, its direct contribution to reading development might be less causal than is generally assumed (see Peterson et al., 2018, for a similar argument). We speculate that instead of being a prerequisite for learning to read, PA may function as a corequisite skill for typical reading development.

In summary, the findings of our longitudinal cross-linguistic study show that RAN is a universal predictor of reading in five alphabetic orthographies varying in consistency. In turn, the relationship between PA and reading appears to be complex and interactive (Castles & Coltheart, 2004). We readily acknowledge that cross-linguistic studies like ours have inherent drawbacks, most important the matching of samples and stimuli across so many languages. Our findings are specific to the age group assessed and the outcome measures used. In studies of this type, control of extraneous measures (e.g., socioeconomic status) and differences in reading preparation and reading instruction are also difficult to control. Despite these obvious limitations, we are strongly convinced that such a cross-linguistic approach can be highly valuable and that more parallel studies in different (not only alphabetic) orthographies will help to further specify universal and language-specific cognitive mechanisms underlying reading acquisition across the languages of the world (Daniels & Share, 2018; Verhoeven & Perfetti, 2017).

Conflict of interest

The authors report that there are no conflicts of interest.

ORCID

Rauno Parrila ⓘ http://orcid.org/0000-0003-4250-8980

References

Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA: The MIT Press.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*, 868–883. doi:10.1037/edu0000006

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*, 41–75. doi:10.1207/s15328007sem1201_3

Bowers, J. S., & Bowers, P. N. (2017). Beyond phonics: The case for teaching children the logic of the English spelling system. *Educational Psychologist*, *52*, 124–141. doi:10.1080/00461520.2017.1288571

Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor, MI: University of Michigan Press.

Brus, B. T., & Voeten, M. J. M. (1979). *Een-Minuut-Test, vorm A en B: Verantwoording en handleiding. [One-minute-test, version A and B: Justification and manual]*. Lisse, the Netherlands: Swets & Zeitlinger.

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*, 1398–1407. doi:10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C. L. M., Onochie-Quintanilla, E., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*, 678–686. doi:10.1177/0956797611434536

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*, 77–111. doi:10.1016/S0010-0277(03)00164-1

Compton, D. L. (2003). Modeling the relationship between growth in rapid naming speed and growth in decoding skill in first-grade children. *Journal of Educational Psychology*, *95*, 225–239. doi:10.1037/0022-0663.95.2.225

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*, 234–244. doi:10.1038/nrn3924

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading*, *22*, 101–116. doi:10.1080/10888438.2017.1379082

De Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading*, *15*, 314–337. doi:10.1080/10888438.2010.485624

Diamanti, V., Goulandris, N., Campbell, R., & Protopapas, A. (2018). Dyslexia profiles across orthographies differing in transparency: An evaluation of theoretical predictions contrasting English and Greek. *Scientific Studies of Reading*, 22, 55–69. doi:10.1080/10888438.2017.1338291

Duncan, L. G., Casto, S. L., Defior, S., Seymour, P. H. K., Baillie, S., & Serrano, F. (2013). Phonological development in relation to language and literacy: Variations on a theme in six alphabetic orthographies. *Cognition*, 127, 398–419. doi:10.1016/j.cognition.2013.02.009

Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, 59(2), 127–136. doi:10.1198/000313005X41337

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2, 31–54. doi:10.1207/s1532799xssr0201_2

Furnes, B., & Samuelsson, S. (2009). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, 21, 85–95. doi:10.1016/j.lindif.2010.10.005

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. doi:10.1002/dys.401

Gangl, M., Moll, K., Jones, M. W., Banfi, C., Schulte-Körne, G., & Landerl, K. (2018). Lexical reading in dysfluent readers of German. *Scientific Studies of Reading*, 22, 24–40. doi:10.1080/10888438.2017.1339709

Georgiou, G., Aro, M., Liao, C.-H., & Parrila, R. (2015). The contribution of RAN pause time and articulation time to reading across languages: Evidence from a more representative sample of children. *Scientific Studies of Reading*, 19, 135–144. doi:10.1080/10888438.2014.956927

Georgiou, G., Papadopoulos, T. C., Fella, A., & Parrila, R. (2012). Rapid naming speed components and reading development in a consistent orthography. *Journal of Experimental Child Psychology*, 112, 1–17. doi:10.1016/j.jecp.2011.11.006

Georgiou, G., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing: an Interdisciplinary Journal*, 25, 321–346. doi:10.1007/s11145-010-9271-x

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. doi:10.1037/0022-0663.100.3.566

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, 33, 313–317. doi:10.1027/1015-5759/a000450

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. doi:10.1080/10705519909540118

Hulme, C., Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading*, 9, 351–365. doi:10.1207/s1532799xssr0904_2

Hulslander, J., Olson, R. K., Willcutt, E. G., & Wadsworth, S. J. (2010). Longitudinal stability of reading-related skills and their prediction of reading development. *Scientific Studies of Reading*, 14, 111–136. doi:10.1080/10888431003604058

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software*, 56, 1–31. doi:10.18637/jss.v056.i06

Jones, M., Snowling, M., & Moll, K. (2016). What automaticity deficit? Activation of lexical information by readers with dyslexia in a RAN-Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 465–474. doi:10.1037/xlm0000186

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam, the Netherlands: Elsevier Science Publishers.

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. doi:10.1598/RRQ.45.3.4

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., & Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, 54, 686–694. doi:10.1111/jcpp.12029

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and orthographic spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, 100, 150–161. doi:10.1037/0022-0663.100.1.150

Lervåg, A., & Hulme, C. (2009). Rapid Automatized Naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, 20, 1040–1048. doi:10.1111/j.1467-9280.2009.02405.x

Liberman, L. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & L. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–34). Ann Arbor, MI: the University of Michigan Press.

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing: an Interdisciplinary Journal*, 15, 653–682. doi:10.1023/A:1020984704781

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138, 322–352. doi:10.1037/a0026744

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading*, 13, 1–25. doi:10.1080/10888430802631684

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest [Reading and Spelling Test]*. Bern, Switzerland: Hans Huber.

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., & Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies: Is English an outlier orthography? *Learning and Instruction*, 29, 65–77. doi:10.1016/j.learninstruc.2013.09.003

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–331. doi:10.1016/0010-0277(79)90020-9

Norton, E. S., & Wolf, M. (2012). Rapid Automatized Naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, 63, 427–452. doi:10.1146/annurev-psych-120710-100431

Peterson, R. L., Arnett, A. B., Pennington, B. F., Byrne, B., Samuelsson, S., & Olson, R. K. (2018). Literacy acquisition influences children´s rapid automatized naming. *Developmental Science*, 21, e12589. doi:10.1111/desc.12589

Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. (2014). Longitudinal stability of phonological and surface subtypes of developmental dyslexia. *Scientific Studies of Reading*, 18, 347–362. doi:10.1080/10888438.2014.904870

Plaza, M., & Cohen, H. (2007). The contribution of phonological awareness and visual-attntion in early reading and spelling. *Dyslexia*, 13, 67–76.

Poulsen, M., Juul, H., & Elbro, C. (2015). Multiple mediation analysis of the relationship between rapid naming and reading. *Journal of Research in Reading*, 38, 124–140. doi:10.1111/j.1467-9817.2012.01547.x

Protopapas, A., Altani, A., & Georgiou, G. (2013a). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, 116, 914–929. doi:10.1016/j.jecp.2013.08.004

Protopapas, A., Altani, A., & Georgiou, G. K. (2013b). RAN backward: A test of the visual scanning hypothesis. *Scientific Studies of Reading*, 17, 453–461. doi:10.1080/10888438.2013.769556

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, 54, 315–323.

Rodríguez, C., Van Den Boer, M., Jiménez, J. E., & De Jong, P. F. (2015). Developmental changes in the relations between RAN, phonological awareness, and reading in Spanish children. *Scientific Studies of Reading*, 19, 273–288. doi:10.1080/10888438.2015.1025271

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, 6, 1715. doi:10.3389/fpsyg.2015.01715

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, Articles*, 48 (2), 1–36. Retrieved from. doi:10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & Van Der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications. *Structural Equation Modeling*, 16, 561–568. doi:10.1080/10705510903203433

Savage, R., Pillay, V., & Melidona, S. (2007). Deconstructing rapid automatized naming: Component processes and the prediction of reading difficulties. *Learning and Individual Differences*, 17, 129–146. doi:10.1016/j.lindif.2007.04.001

Scalisi, T. G., Desimoni, M., & Di Vito Curmini, L. (2013). Age and schooling effects on the development of phonological awareness (PA) and rapid automatized naming (RAN) in Italian children aged 5-7 years. *Journal of Applied Psycholinguistics*, 13, 45–60.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, 22, 1614–1629. doi:10.3758/s13423-015-0835-2

semTools contributors (2016). *semTools: Useful tools for structural equation modeling*. Retrieved from https://CRAN.R-project.org/package=semTools

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174. doi:10.1348/000712603321661859

Song, S., Georgiou, G. K., Su, M., & Shu, H. (2016). How well do phonological awareness and rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-analysis. *Scientific Studies of Reading*, 20, 99–123. doi:10.1080/10888438.2015.1088543

Suggate, S. P. (2016). A meta-analysis of the long-term effect of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49, 77–96. doi:10.1177/0022219414528540

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.

Torppa, M., Parrila, R., Niemi, P., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2013). The double deficit hypothesis in the transparent Finnish orthography: A longitudinal study from Kindergarten to Grade 2. *Reading and Writing: an Interdisciplinary Journal*, 26, 1353–1380. doi:10.1007/s11145-012-9423-2

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. doi:10.1037/a0019465

Van Den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). *De Klepel – verantwoording, diagnostiek en behandeling [Nonword Reading Test]*. Nijmegen, The Netherlands: Berkhout.

Van Viersen, S., De Bree, E. H., Zee, M., Maasssen, B., Van Der Leij, A., & De Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science*, 29, 418–428. doi:10.1177/0956797617736886

Venezky, R. (1970). *The structure of English orthography*. The Hague, the Netherlands: Mouton.

Verhoeven, L., & Perfetti, C. (Eds.). (2017). *Learning to read across languages and writing systems*. Cambridge, UK: Cambridge University Press.

Wei, W., Georgiou, G. K., & Deng, D. (2015). Examining the cross-lagged relationships between RAN and word reading in Chinese. *Scientific Studies of Reading*, 19, 446–455. doi:10.1080/10888438.2015.1077447

Wimmer, H. (2006). Don´t neglect reading fluency! *Developmental Science*, 9, 447–448. doi:10.1111/j.1467-7687.2006.00527.x

Wimmer, H., Landerl, K., Linortner, R., & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. *Cognition*, 40, 219–249. doi:10.1016/0010-0277(91)90026-Z

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, 92, 668–680. doi:10.1037/0022-0663.92.4.668

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91, 415–438. doi:10.1037/0022-0663.91.3.415

Wolff, U. (2014). RAN as a predictor of reading skills, and vice versa: Results from a randomised reading intervention. *Annals of Dyslexia*, 64, 151–165. doi:10.1007/s11881-014-0091-6

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, 21, 551–559. doi:10.1177/0956797610363406

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages. A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29. doi:10.1037/0033-2909.131.1.3