Effects of Orthographic Depth on Literacy Performance: Reading Comprehension Difficulties Across Languages

By

Nicola Ann McClung

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Anne Cunningham, Chair
Professor P. David Pearson
Professor Sophia Rabe-Hesketh
Professor Thomas Wickens

Spring 2012

Effects of Orthographic Depth on Literacy Performance:
Reading Comprehension Difficulties Across Languages

Abstract

Effects of Orthographic Depth on Literacy Performance:

Reading Comprehension Difficulties across Languages

by

Nicola Ann McClung

Doctor of Philosophy in Education

University of California, Berkeley

Professor Anne Cunningham, Chair

Orthographic depth, the degree of spelling-to-sound consistency in each language, has been hypothesized to affect the ease and effectiveness with which children learn to read (Frost, Katz, & Bentin, 1987). This linguistic factor has been found to have such a powerful effect on the beginning reading process that readers of English (the deepest alphabetic orthography) are estimated to take two and a half years longer to master basic decoding and word recognition skills than the majority of readers of other alphabetic orthographies (Seymour, Aro, & Erskine, 2003). Although the link between orthographic depth and reading acquisition has been well established, there is little to no research examining this relationship beyond the third grade. As a result, it is widely assumed that by fourth grade, reading is the same in all languages. However, because another body of research has established the ongoing and mutually supportive relationship between word reading and reading comprehension throughout development (e.g., Kintsch, 1998; Perfetti, 2007), the current study sought to challenge this assumption by investigating the long-term role of orthographic depth in reading comprehension performance. Hierarchical linear modeling (HLM) was employed to examine subsets of data that include representative samples of typically developing fourth-grade and 15-year-old readers of alphabetic languages from the 2006 Progress in International Reading and Literacy Study (PIRLS) and the 2009 Program for International Study Assessment (PISA). Findings provide tentative indication that by fourth grade, the role of orthographic depth is diminished for competent and skilled readers. However, the poorest 10 to 25% of readers appear to be continually challenged throughout development when there is a greater degree of ambiguity in their script. In contrast, even when students are below average in their ability, they may enjoy the long-term benefits of reading in a shallow orthography. The implications of this study are wide-ranging and include: accounting for orthographic differences in models of reading comprehension, providing more resources for the lowest performing readers of deep orthographies throughout schooling, and acknowledging that reading comprehension instruction and assessment may not be comparable across languages.

For Karen,
who believed in me, spent countless late nights with my work, and taught me how to write

For Bill,
who taught me how to think and believed in the importance of my work

For University Press Books,
which has always provided a home and inspired a love of reading

and for Lorn(a),
my life

Table of Contents

List of Tables

List of Figures

Acknowledgements

I want to express my deepest gratitude to my committee members. Anne, your unwavering support, poise, and mentorship have shown me how to be a scholar and teacher. David, the combination of your kindness, inclusiveness, trust, intellectual curiosity, and scholarly and professional success has been the greatest inspiration. Sophia, thank you for opening my eyes to an entirely new way of seeing the world and giving me the opportunity to learn about and teach statistics. Tom, thank you for your patience and openness, passion for teaching, creativity, and mostly for your sense of humor; you have challenged and supported me to gain a deeper understanding of quantitative research. Each of you has inspired me in countless unique ways to be a better scholar, teacher, and person.

I would also like to thank my many friends and colleagues, who have believed in me and made my graduate school experience possible in so many ways.

Effects of Orthographic Depth on Literacy Performance:
Reading Comprehension Difficulties Across Languages

Despite educators' and researchers' persistent efforts, many elementary, middle, and high school students continue to struggle with reading grade-level text (Snow, 2001). This problem makes it difficult to access reading materials and information across subject areas and, in turn, prevents students from attaining educational, economic, and life success (NEA, 2007). As a result, the Institute of Education Sciences (IES) asserts that reading comprehension is the most critical issue in the area of literacy research. Furthermore, the need to enhance our understanding of the factors that inhibit and support reading comprehension is pressing, because while the demands for advanced levels of literacy in the workplace increase, there are persistent disparities in reading performance between different demographic groups (Snow, 2001). For example, there is an increasing trend for students in the United States to perform poorly on measures of literacy in reading, mathematics, and science compared to students in other developed countries (Fleischman, Hopstock, Pelczar, & Shelley, 2010), and within the United States there continues to be an achievement gap between different racial/ethnic groups (Hemphill &Vanneman, 2011).

Reading comprehension relies on the ability to access vocabulary and background knowledge, attend to text structure and text coherence, understand the semantic, syntactic, and prosodic aspects of language, draw inferences, make predictions, generate interpretations, integrate information, understand elements of genre, recognize bias, and critically use, evaluate, and analyze text (e.g., Block & Pressley, 2007). Although these skills underlie successful reading comprehension, each skill is dependent on the ability to efficiently recognize words. That is, accurate, fluent, and automatic word recognition frees up cognitive resources and allows the reader to focus on the comprehension process (e.g., Perfetti, 2007; Stanovich, 1980). Thus, word recognition skills serve as the base upon which reading comprehension rests. As a result, uncovering the factors that influence word recognition may provide critical insights regarding our understanding and explanation of students' reading comprehension difficulties. A student's language is one of many such factors that may affect word recognition and that might, in turn, have lasting effects on the reading comprehension process. Although many researchers have focused on the important roles of a student's home language, the language of schooling, and the mismatch between them (Heath, 1982; Reese & Gallimore, 2000), variability in spelling-to-sound consistency across languages may be another source of reading difficulty that stems from the linguistic environment.

Theoretical Framework

Reading is made possible by the fact that every orthography (written language) is a written representation of its phonology (spoken language), and thus the process of converting written words into their speech-based forms is a universal aspect of reading (Sampson, 1985). However, because each language differs in the way the orthography represents the phonology, some aspects of reading are language specific (Frost, 2005).

The connectionist perspective, a prominent theoretical model of reading, explains how word learning occurs in different languages. In this model, cognition operates through simultaneously functioning networks of three types of mental representations: phonological, orthographic, and semantic (e.g., Seidenberg & McClelland, 1989). Although all children have a common cognitive structure, the specifications of these mental representations and the emphasis placed on the connections between them may be different depending on the demands of the language (Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004). Differences in how phonological, orthographic, and semantic processes operate and interact arise in response to the degree of *ambiguity* within a particular word and across words in each language; ambiguity is defined in the relationship between a language's phonology and orthography. Ambiguity can manifest as feedforward (spelling-to-sound) or feedback (sound-to-spelling) irregularity, and can occur at multiple scales of correspondence between spelling and sound: phoneme/grapheme, spelling-body/pronunciation-rime, and whole word levels. According to Van Orden and Kloos (2005), "each language presents a unique compilation of ambiguity that will be uniquely sampled by each reader" (p. 76). Reading, therefore, is affected by the degree of spelling-to-sound and sound-to-spelling ambiguity in the reader's orthography.

The orthographic-depth hypothesis posits that as one moves along the continuum from shallow (unambiguous) to deep (ambiguous) alphabetic orthographies, reading acquisition becomes more difficult (Frost, Katz, & Bentin, 1987). Orthographic depth has also been described as ranging from regular to irregular, consistent to inconsistent, and transparent to opaque. However, orthographic depth (and the shallow-to-deep continuum) will be used for the purposes of this discussion. In shallow orthographies, each letter typically represents a single sound and each sound a single letter. While in deep orthographies some letters and letter combinations have no corresponding sound (e.g., *t* in *listen*), they map onto multiple sounds (e.g., *a* in *cat*, *was*, *car, salt,* and *cake*; and -*ough* in *though*, *through*, *bough*, and *thought*), and there are multiple ways to spell the same sound (e.g., /k/ in *cat*, *trek*, *opaque*, and *check*; Frost, 2005). Although there has been no systematic mathematical classification of languages, there is general agreement that alphabetic orthographies can be classified by their orthographic depth and syllabic complexity (Niessen, Frith, Reitsma, & Öhngren, 2000). To date, the most comprehensive language classification system explaining how language characteristics affect reading acquisition was created by the 2000 European Cooperation in Science and Technology (COST) consortium (Niessen et al., 2000). Table 1 shows how various alphabetic languages can be categorized according to the dual languages dimensions of syllabic complexity and orthographic depth. For example, Finnish is thought to simple and shallow, while English is complex and deep.

Syllabic complexity is another language characteristic that is believed to play a role in the beginning reading process. Specifically, it is thought to affect how readily children become sensitive to the phonological structure of language (Duncan, Colé, Seymour, & Magnan, 2006), a critical prereading skill (Stanovich, 1980). Simple syllabic structure is characterized by the predominance of open consonant-vowel (CV; e.g., *to*), vowel-consonant (VC; e.g., *at*), and vowel (V; e.g., *a*) syllables, where vowels have a relatively limited range. Languages with complex syllabic structures have numerous closed consonant-vowel-consonant syllables with many complex consonant clusters (e.g., CVC, CCVC, and CVCC; *cat*, *that*, and *both*, respectively), including many in the onset or coda positions, and vowels tend to be relatively ambiguous (Frost, 2005). Finnish, Greek, and the Romance languages in general are thought to be simple, while Germanic languages such as Dutch, Swedish, Danish, and English have been classified as complex (Seymour, Aro, & Erskine, 2003). Children who speak a language with a simple syllabic structure have been found to demonstrate more phonological awareness than their complex-language-speaking counterparts prior to any formal instruction (Duncan et al., 2006). However, although it is clearly important to consider the relationship between syllabic complexity and children's early understanding of the structure of spoken language (i.e., phonological awareness), orthographic depth is thought to have a more powerful effect on the reading process when children begin to engage with print (Seymour et al., 2003).

Finnish: An example of a shallow orthography. As an approximately regular language, Finnish demonstrates a shallow orthography. Phoneme/grapheme correspondences are symmetrical and regular (Frost, 2005). There are 16 consonant phonemes (/p/, /t/, /v/, /j/, /h/, /s/, /r/, /l/, /n/, /m/, /k/, /d/, /b/, /g/, /f/, and /ŋ/) and 8 vowel phonemes (/a/, /ä/, /ö/, /o/, /e/, /u/, /y/, and /i/; Lyytinen et al., 2005). These phonemes correspond to single-letter graphemes except for one that is sometimes represented by a diagraph (/ŋ/ is short when it is in front of /k/ and long in *ng*; Lyytinen et al., 2005). Finnish (like Serbo-Croatian and Hungarian) is extremely regular in both directions, being both feedback and feedforward consistent (Lyytinen et al., 2005; Seymour et al., 2003). Thus, once a child has learned the 23 symmetrical phoneme/grapheme correspondence rules, he or she has a tool for reading or spelling any word or nonword in the Finnish language. Additionally, the syllabic structure of Finnish is thought to be quite simple: open syllables are more common than closed syllables, syllable stress patterns are predictable, and words and syllables almost never begin with consonant clusters. Holopainen, Ahonen, Tolvanen, and Lyytinen (2000) found that many Finnish children were already able to read at the time of school entry and most others learned how to read by the end of the first semester of school. Lyytinen et al. (2005) attribute Finnish children's early reading ability to the distinct advantage of learning to read in a language whose orthography has a relatively small number of highly regular phoneme/grapheme correspondences. Notably, international comparison studies have shown that Finnish children have had the highest overall reading achievement in terms of word reading and comprehension (Elley, 1994; OECD, 2010). In sum, learning how to read words in Finnish is a task that is relatively easy to master.

English: An example of a deep orthography. With its extreme inconsistencies, English is classified as deep and complex (Frost, 2005; Seymour et al., 2003) and is thought to be an outlier language (Share, 2008). One salient feature of the English orthography is that morphological information is preserved at the expense of phonemic information, causing bidirectional ambiguities at whole-word, spelling-body/pronunciation-rime, and grapheme/phoneme levels. Indeed, 31% of all English monosyllabic words have been found to be feedforward (i.e., spelling-to-sound) inconsistent (Ziegler, Stone, & Jacobs, 1997). It is estimated that there are over 500 phoneme/grapheme correspondences (Gough, Juel, & Griffith, 1992), which includes 15 vowel sounds that are represented by fewer graphemes (e.g., the letter *a* maps onto at least seven different sounds in *sat*, *was*, *saw*, *far*, *salad*, and *made*; Frost, 2005). At the grapheme/phoneme level, there is both feedforward (e.g., the *t* in whistle) and feedback (i.e., sound-to-spelling) inconsistency (e.g., /i/ in *reef*, *thief*, and *leaf*). At the level of spelling-bodies/pronunciation-rimes, inconsistencies are manifest in words that are spelled similarly (e.g., -*ough*) but pronounced differently (e.g., *though*, *thought*, *through*, *bough*, and *tough*; Glushko, 1979). Ambiguities at the level of whole words are due to the preponderance of English-language homophones (e.g., *break/brake* and *bare/bear)* and homographs (e.g., *wind, bank,* and *tear*). Although there are multiple levels of ambiguity between orthographic and phonologic units in English, Treiman, Mullennix, Bijeljac-Babic, Richmond-Welty (1995) suggest that larger orthographic units tend to be more regular than smaller units. Thus, children are thought to develop both small- and large-unit reading strategies (Goswami, 1988). Zeigler and Goswami (2005) posit that in order to read the 3,000 most frequent monosyllabic words at the level of the rime, a child needs to learn 600 different phonologic/orthographic patterns and 400 phonological rimes. Additionally, the syllabic structure of English is complex. There are 16 syllable structures (Abercrombie, 1967), syllable stress patterns vary (as opposed to being fixed; Schane, 1968), and there are many closed syllables with complex consonant clusters (Frost, 2005). One can thus see how the irregularities of the English writing system put it at the opposite end of the orthographic-depth continuum from Finnish.

Evidence for the Orthographic Depth Hypothesis

To date, the European Concerted Action on Learning Disorders as a Barrier to Human Development (reported in Seymour et al., 2003) has conducted the most comprehensive cross-linguistic study of beginning word reading. This study, which aimed to provide evidence for the orthographic depth hypothesis, compared typically developing first-grade children from 14 European countries on letter knowledge, ability to identify familiar words, and decoding of simple nonwords. Findings indicated that most children were fluent and accurate decoders by the end of first grade—on average they decoded high frequency words with 87% accuracy. However, the Portuguese-, French-, Danish-, and English-speaking children had yet to acquire the level of mastery of these basic reading skills as the children who spoke less deep orthographies. The lowest performing groups were the Danish- and English-speaking children, who read high-frequency words with 71 and 31% accuracy, respectively. Additionally, the design of this study was such that it ruled out the possibility that differences in socioeconomic status (SES) might explain variance in reading performance in different languages. To this end, the English-readers in the study were recruited from a high SES area of Scotland and were performing well ahead of age expectations according to U.K. norms, while the readers of the other European languages came from middle SES backgrounds. Thus, despite their higher levels of SES, the Scottish (English-speaking) children were the lowest performing in the study.

The results from this study suggest that orthographic depth is related to the ease and effectiveness with which children acquire literacy skills. They also suggest that the Finnish children learned how read at least twice as fast as the children learning in English, and that English-speaking children acquired literacy skills more slowly than readers of any other alphabetic language.

Other studies corroborate the Seymour et al. (2003) results. Aro and Wimmer (2003) compared the reading performance of German-, Dutch-, Spanish-, Swedish-, Finnish-, English-, and French-speaking children. They found that on average, most children reading in alphabetic languages read nonwords with 90% accuracy by the end of first grade, except for the English group whose accuracy hovered at 50%. Porpodas (1999) found that children learning to read in Greek also read words and nonwords with 90% accuracy by the end of first grade. Further converging evidence was demonstrated by Landerl (2000) which showed that children reading in Dutch, German, Greek, Italian, Portuguese, and Turkish read nonwords with at least 75% accuracy by the end of first grade. Additionally, Goswami, Porpodas, and Wheelwright (1997) compared learning to read in Greek and English, Öney and Goldman (1984) compared Turkish to English, and Thorstad (1991) compared Italian to English; each observed that relative to other languages, learning to read in English was the most difficult. Looking across the entire range of these cross-linguistic studies, findings have demonstrated that learning to read in a shallow orthography is relatively easy compared to a deep orthography, and English is the most difficult alphabetic language in which to acquire literacy skills.

Criticisms of the cross-linguistic research. One criticism of the cross-linguistic research is that it is impossible to definitively compare readers from different countries due to potentially confounding sociocultural differences. For example, countries may vary in terms of pedagogical practices, the age at which reading instruction begins, curricula, values and beliefs about reading, or demographic characteristics (Zeigler & Goswami, 2005). To address this issue, some researchers have designed studies to compare the reading skills of children from the same country, who speak different languages, yet are matched on many important sociocultural variables, while other researchers have employed within-subject designs to examine the reading skills of bilingual children learning in two languages. For example, Spencer and Hanley (2003) studied similar groups of English- and Welsh-speaking children living in Wales who were learning to read and receiving instruction in their primary language. They found that the students learning in the relatively shallow Welsh orthography were significantly more advanced in word and nonword reading skill than the students learning in English. Similarly, ruling out potential confounds in the relationship between orthographic depth and reading development, Georgiou, Parrila, and Papadopoulos (2008) compared the English and Greek reading skills of bilingual children. The results from their study suggest that there were important differences in how the students processed the two scripts. The children relied more on decoding when reading in Greek and exhibited a larger orthographic-unit approach when reading in English. The results from these studies, in which the researchers were able account for sociocultural differences by making within-country and within-subject comparisons, provide strong additional evidence for the role of orthographic depth in early reading development.

Another criticism of cross-linguistic research is that the stimuli employed in this work may not be comparable across languages. To address this concern, Landerl, Wimmer, and Frith (1997), and Zeigler, Perry, Jacobs, and Braun (2001) designed and conducted studies in German and English using identical stimuli; the words *bank*, *ball*, and *park*, for example, are found in both languages. Landerl et al. (1997) found that young English readers showed reduced reading accuracy compared to German children, which they hypothesized was the result of the greater spelling-to-sound inconsistency in English as compared to German. Similarly, the results from Zeigler et al.'s (2001) study suggest that readers employ different processing strategies (small-, large-, and/or flexible-grain size unit approaches) in response to the constraints of the writing system. The results from these studies indicate that orthographic depth appears to affect the reading process even when identical words are being read across languages.

Overall, this body of work provides supportive evidence for the orthographic-depth hypothesis. It consistently demonstrates that word reading develops more quickly in children learning in shallow orthographies relative to those learning in deep orthographies. The studies also suggest that reading in a deep orthography may draw on a wider range of cognitive skills, which presumably creates more difficulty and variability in the beginning reading process.

The Role of Orthographic Depth in Reading Comprehension

Although there is extensive evidence indicating that orthographic depth has an important role in beginning reading development, little is known about the effect of this language characteristic beyond the third grade. However, it is widely assumed that by fourth grade, readers of all languages have made the transition from *learning to read* to *reading to learn*, and thus it is thought that orthographic factors cease to have an effect on the reading process. Indeed, this is the rationale for implementing large-scale international comparisons of literacy beginning at the fourth grade (Martin, Mullis, & Kennedy, 2007). Despite the assumption that students have mastered word-recognition skills, theoretically, word identification is thought to be supportive of, intertwined with, and in constant interaction with the comprehension process (Kintsch, 1998). For example, the lexical quality hypothesis (Perfetti, 1992) states that orthographic, phonological, and semantic representations and their bindings are critical to the reader's ability to attend to the meaning of print. That is, when the lexical quality of a word is precise, redundant, and well specified, it can be accessed efficiently, without competing with other similar and partly activated candidates. Thus, sublexical and lexical-level knowledge supports the reader's ability to comprehend text (Perfetti, 1992, 2007; Perfetti, Landi, & Oakhill, 2005; Stanovich, 1980).

There is also substantial evidence suggesting that word-level processes continue to have a residual effect on reading performance throughout development (Cunningham & Stanovich, 1997; Perfetti & Bolger, 2004; Vellutino & Scanlon, 1987). For example, when Cunningham and Stanovich (1997) followed a group of first-grade students until they were in 11th grade, they found that first-grade word reading ability was a strong and reliable predictor of 11th grade reading outcomes. Similarly, Cunningham, Stanovich, and Wilson (1990) found that early decoding skill was related to skilled reading at the college level. Accordingly, it is reasonable to hypothesize that early difficulty with word recognition, when related to the depth of the reader's orthography, may lead to later comprehension difficulties.

Although the literature examining the relationship between orthographic depth and reading comprehension ability is limited, it can nonetheless be inferred that spelling-to-sound consistency plays a long-term role in reading performance. Thus, further research is needed to study this possibility in the context of older children. Hanley, Masterson, Spencer, and Evans (2004) provide some initial insights into the lasting consequences of learning how to read in a deep orthography. They found that six- and seven-year-old Welsh-speaking children, who were learning to read in the highly predictable Welsh orthography, performed significantly better on word and nonword reading measures than Welsh-English-speaking children learning to read in English. However, after these students had reached their sixth year of instruction, while the majority of the English-speaking children had caught up to their Welsh-speaking counterparts, the lowest performing 25% of English readers continued to perform significantly below the lowest performing 25% of Welsh readers on all measures of reading achievement. The results from this study provide preliminary evidence that there may be long-term negative effects of reading in a deep orthography—especially for poor readers.

The Current Study

This study sought to determine whether orthographic depth is uniquely related to reading comprehension at the fourth-grade and 15-year-old levels. The following hypotheses were tested:

(1) There are long-term advantages of reading in shallow orthographies compared to deep orthographies.
(2) There are additional long-term positive effects of reading in shallow orthographies compared to deep orthographies for the lowest performing 10 to 25% of readers.
(3) Orthographic depth has the strongest influence on the comprehension skills of the poorest readers, who are presumably still struggling at the word-recognition level.
(4) Reading difficulties related to orthographic depth are exacerbated over time such that the gap between poor readers of relatively deep and shallow orthographies becomes more apparent as students develop from fourth grade to high school.
(5) There is less variance in reading comprehension performance in shallow orthographies than deep orthographies because reading in shallow orthographies is a relatively easier task.

Important control variables that were considered are gender, SES, language spoken at home, age, and the availability of school resources. Gender was considered because prior research has demonstrated that compared to boys, girls tend to have better reading comprehension skills, read more frequently, and have a more positive attitude towards reading and school (Logan & Johnson, 2009). SES is associated with differences in resources, values and beliefs about reading, and the quality and quantity of literacy interactions in the home, and is strongly predictive of reading achievement (Aikens & Barbarin, 2008). First-language learners tend to have better initial reading comprehension and faster rates of growth than second-language learners (Lervåg & Aukrust, 2010). While the findings on school resources are mixed, in some studies they have been positively linked to achievement (Greenwald, Hedges, & Laine, 1996). Finally, it was thought that the above hypotheses would hold true regardless of whether the shallow or deep orthography reflects the dominant language spoken in a country.

Table 2 shows how hypotheses 1, 2, and 5 would play out in the context of the languages included in this study based on COST's hypothetical classification of languages. Figure 1 provides a graphic display of hypotheses 1, 2, 3, 4, and 5: the main effect of orthographic depth and the orthographic depth by reading ability interactions. The gap in performance between readers of shallow and deep orthographies is expected to widen as students decrease in reading ability and increase in age, creating heterogeneity in the variance in reading performance across languages.

Method

Sample and Data

Sample. The participants in this study, who come from Belgium, Canada, Finland, and Switzerland, took part in two large international research projects, the 2006 Progress in International Reading and Literacy Study (PIRLS), and the 2009 Program for International Assessment (PISA). The sample includes 6,180 fourth-grade students from 387 schools and 6,642 15-year-old students from 271 schools in Belgium; 16,176 fourth-grade students from 1,120 schools and 21,531 15-year-old students from 972 schools in Canada; 5,615 15-year-old students from 203 schools in Finland; and 10,359 15-year-old students from 426 schools in Switzerland. Table 3 includes information about the sample and population in each country (language, source of data, age/grade of students, and student- and school-level sample sizes), and Table 4 provides information about the cluster sizes (i.e., students per school) by country.

The rationale for including these particular countries was twofold: first, in order to control for sociocultural, economic, and educational country-level factors, only countries with at least two nationally recognized alphabetic languages (that were also the official languages of instruction and the home languages of the majority of students tested) were considered; and second, countries with more than one national language were included only if at least two of the languages had been classified in different orthographic depth categories by COST. Four countries with a total of seven languages met these criteria: Belgium (Dutch and French), Canada (English and French) in the PIRLS and PISA datasets, and Finland (Swedish and Finnish) and Switzerland (French, Italian, and German) in the PISA dataset. This group of countries was ideal because they alternate in terms of whether the language spoken by the majority of the population has a relatively shallow or deep orthography in each country, and thus it was possible to disambiguate the effect of orthographic depth from the effect of the linguistic makeup of the countries studied. Additionally, this set of countries was interesting because while French is a relatively shallow orthography in Canada, it is a deep orthography in Belgium and Switzerland. However, one limitation of these countries is that they contain languages that are only slightly different in terms of orthographic depth. For example, there may be little meaningful difference in orthographic depth between Swedish and Finnish because they are both shallow orthographies. Thus, a weakness of this study is that it was not possible to make comparisons of languages that were farther apart on the orthographic depth continuum, which could potentially attenuate the magnitude of the results.

Population language characteristics. As mentioned previously, it was important to account for the linguistic makeup of each country in order to determine whether orthographic depth plays a unique long-term role in reading performance, and to rule out the possibility that any linguistic advantage was due to speaking the same language as the majority of the population in that country. In Belgium, approximately 60% of the population speaks Dutch as a mother tongue, while 33% of the population speaks French as a mother tongue, and 7% speaks a different primary language (Languages across Europe, 2011). In Canada, more than 58% of the population speaks English as a mother tongue, approximately 22% of the population speaks French as a mother tongue, and the remaining 20% speaks a different primary language (Statistics Canada, 2008). In Finland, 93% of the population speaks Finnish as a mother tongue, 6% speaks Swedish as a mother tongue, and 1% speaks a different primary language (Languages across Europe, 2011). In Switzerland, approximately 63% of the population speaks German as a mother tongue, 20% speaks French as a mother tongue, and 8% speaks Italian as a mother tongue, while the remaining 9% speaks a different primary language (Languages across Europe, 2011). In sum, the dominant languages in the countries included in this study are: Dutch in Belgium, English in Canada, Finnish in Finland, and German in Switzerland. Table 3 provides the percentage of the population speaking each language by country.

Data. Occurring every five years at the fourth-grade level (or its national equivalent), PIRLS aims to provide information about literacy development and education on an international level. The fourth year of formal schooling is chosen specifically because students are thought to have made the transition from *learning to read* to *reading to learn* at this grade level—a critical period in reading development (Joncas, 2007, p. 36). PIRLS is conducted by the International Association for the Evaluation of Educational Achievement (IEA), and is funded by the participating countries with support from the World Bank and the U.S. Department of Education's National Center for Educational Statistics (NCES; Mullis, Kennedy, Martin, & Sainsbury, 2006). PISA focuses on the capabilities of 15-year-olds in regard to reading literacy. This project is coordinated by the Organization for Economic Cooperation and Development (OECD), is carried out by the PISA consortium, and is led by the Australian Council for Educational Research (ACER; Fleischman et al., 2010). The goal of PISA is to measure academic skills that are pertinent to life success as students begin to transition into adulthood.

Sampling methodologies. In both international projects, all countries used a uniform sampling approach that followed international guidelines and specifications to ensure that differences in national achievement outcomes could not be attributed to the use of different sampling methodologies. Two-stage stratified sample designs were used in both the PIRLS and PISA assessments, and probability samples were drawn from target populations in each country.

In PIRLS, schools were the first stage, while intact classrooms were the second (Joncas, 2007). Each country typically sampled 150 schools in order to yield a representative sample of at least 4,000 students (Mullis, Marin, Kennedy, & Foy, 2007). All of the countries included in this study met the PIRLS sampling requirement that the participants reflected at least 85% of both the students and schools that were sampled, or a combined rate of 75% (Joncas, 2007).

The first-stage sampling units in PISA consisted of schools that were systematically sampled from a list of all PISA-eligible schools in each country (full- and part-time educational institutions, vocational training programs, and foreign schools containing 15-year-old students). At the second stage, probability samples of approximately 35 students were selected from each school containing the target cluster size for each country (typically more than 35 students). In schools with less than the target cluster size, all students were selected (Hopstock & Pelczar, 2011).

Inclusion criteria. The PIRLS participants were representative samples of typically developing fourth-graders and the PISA sample included 15-year-old students (i.e., between 15 years and three months to 16 years and two months at the beginning of the testing period). The participants were not intellectually or functionally disabled, nor non-native language speakers (i.e., unable to read or speak the language of the test; Fleischman et al., 2010; Martin et al, 2007). Additionally, PIRLS excluded schools that were geographically remote, had very few students, or had a curriculum that differed from the mainstream education system (Joncas, 2007).

Assessment design. PIRLS 2006 was a paper-and-pencil test designed to provide internationally comparable reading-related data at the national level. The assessment consisted of 10 authentic text passages (i.e., materials commonly used in schools), five literary and five informational. The passages were rotated among 13 test booklets containing 126 multiple choice and constructed response items, designed to be completed in two 40-minute sessions. Any given student was asked to read one of 13 booklets, containing a unique configuration of the 10 passages. Following the testing, students were given at least 20 minutes to complete a student survey. Additionally, a home survey, given to the primary care giver, gathered information about family demographics, and a school questionnaire provided information from the school principal, such as the availability of school resources (Martin et al., 2007).

The 2009 PISA assessment was paper-and-pencil or computer-based. The reading assessment consisted of 102 cognitive items containing multiple-choice, short-response, and constructed-response questions that were arranged in seven units around a common stimulus. On average, there were 15 items per unit. Four of the seven possible units were allocated to 13 test booklets in a rotated design. Each student was given one test booklet containing the 102 reading items, which took approximately two hours to complete. Some booklets also included mathematics and/or science items. In addition to the test, students were given a 30-minute questionnaire in which questions were answered about their background, attitude, and experience in school, and principals answered a 30-minute questionnaire providing information at the school level (Fleischman et al., 2010).

Translation. In any cross-linguistic study, it is critical that the measures are reliable and contain comparable information across languages. The development of both PIRLS and PISA included exhaustive procedures to verify that the translation of each assessment corresponded to international standards, and to ensure equality across languages. Translation was provided for the test directions, passages, and items, student, home, and school questionnaires, directions for preparing and administering the assessment at schools, and scoring guides for students' open response questions (Martin et al., 2007).

The PIRLS instruments were created in English and then translated into 45 languages. The translation process included: detailed standards for cultural adaptation and translation, comparisons of at least two independent translations at national centers, translation verification by the IEA Secretariat based on professional translations from an independent company, differential item functioning (DIF) analyses to ensure item performance was comparable across languages, and reviews and corrections to any items that were considered problematic based on the aforementioned criteria (Mullis et al., 2007).

Efforts to establish linguistic equivalence of the PISA instrument included a double translation design such that a source version of the test (typically English) was translated into the national languages of each country. Comparisons were then made after the tests were translated back into the source language. Additionally, there were established translation and adaptation standards, training for national staff, and the equivalence of each version was confirmed by international verifiers (Hopstock & Pelczar, 2011).

Overall, groups of international PIRLS and PISA experts, along with test developers, worked to ensure that the items were (a) comparable across languages, (b) culturally unbiased, and (c) appropriate in terms of the interests and reading levels of fourth-grade students and 15-year-olds (Fleischman et al., 2010; Martin et al., 2007).

Measures

The PIRLS comprehension measures were based on item response theory (IRT), and were developed in workshops for country representatives who reviewed the items and passages extensively. The PIRLS 2006 assessment framework defined and assessed two purposes for reading: (1) for literary experience and (2) to acquire and use information. Four major reading comprehension processes were also specified, which were assessed within the two purposes for reading: (1) focus on and retrieve explicitly stated information, (2) make straightforward inferences, (3) interpret and integrate ideas and information, and (4) examine and evaluate content, language, and textual elements (Mullis et al., 2007, p. 17).

The final version of the PIRLS reading assessment included texts that spanned many genres, including five literary texts (e.g., short stories or episodes with illustrations), five informational texts (e.g., biographies), and narratives and expositions (e.g., scientific, geographical, and procedural texts that included text boxes, photographs, maps, or diagrams). The literary texts had two main characters, a plot with a few central events, and took place in a variety of settings. The informational texts were organized by topic or sequence (Mullis et al., 2007).

The PISA reading assessment measured a similar array of reading and thinking skills as did PIRLS, and was moderately aligned with the National Assessment of Education Progress (NAEP; Fleischman et al., 2010). PISA measured three cognitive processes, called “aspect categories,” relating to students’ abilities to respond to text: (1) access and retrieve, (2) integrate and interpret, and (3) reflect and evaluate. However, whereas NAEP aimed to measure school performance, the content of PISA was developed internationally and focused on real-world contexts. The goal of PISA was to determine whether students near the end of high school had developed the reading skills that are essential for full participation in society (i.e., prerequisite reading levels for post-secondary education and career success).

Item response theory was employed to develop the PISA items and literacy subscales. Each item was designed to measure one of the cognitive process categories: access and retrieve, integrate and interpret, and reflect and evaluate (Fleischman et al., 2010). Item difficulty was based on the: (1) number of pieces of information needed to locate/consider, (2) amount of inference required, (3) amount and prominence of competing information, (4) length and complexity of text, (5) type of interpretation required, (6) familiarity with structure and genre, (7) nature of knowledge needed to bring to item (narrow vs. broad), (8) depth of understanding required, and (9) type of information (Fleischman et al., 2010, p. A-16). Although the students within each subsample were given different items, the IRT approach made it possible to describe reading performance on a simple set of continuous scales.

In both PIRLS and PISA, plausible values (i.e., estimates of student ability) were used to address issues of biased statistical inferencing and to allow the use of standard statistical tools to estimate population characteristics. The procedure of using plausible values, initially developed for the analysis of NAEP data, and now common to large-scale assessments including PIRLS and PISA, facilitates the unbiased estimation of structural parameters (compared to the use of direct ability estimates, in which population variance is either under- or over-estimated depending on the estimation method; Wu, 2005).

Reading comprehension. The outcome variables in this study were chosen to reflect the highest level of comprehension ability measured on the PIRLS and PISA assessments. The reader had to move beyond word-, phrase-, and sentence-level processing and basic levels of drawing meaning from print, to show proficiency on these measures.

The PIRLS "Interpret and Integrate Ideas and Information" subsection required the reader to integrate background knowledge and ideas and information contained in the text, and to make implicit connections based on the reader's perspective (Martin et al., 2007). For example, the student could utilize background knowledge to infer the protagonist's motive, visualize the setting or characters, compare and contrast information contained in the text, or identify the author's tone. Overall, this subsection evaluated the reader's ability to consider and reflect on the text's structure and language conventions, and the author's purpose, positionality, and point of view (Martin et al., 2007).

The PISA "Reflect and Evaluate" subsection required the reader to critically evaluate and analyze the text. The reader was asked to recognize the author's purpose, stance, bias, and devices, to take a position on the text's representation, and to identify elements of genre and text structure. To this end, the reader had to relate information contained in the text to his or her own background knowledge, experience, and perspectives beyond the text (Fleischman et al., 2010). This subsection strongly reflected PISA's definition of reading comprehension: "Reading literacy is understanding, using, reflecting on, and engaging with written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD 2010, p. 23).

Poor reader and very poor reader. As previously discussed, higher levels of reading comprehension rely on the ability to attend to text structure and text coherence, draw inferences, make predictions, generate interpretations, integrate information, understand elements of genre, recognize bias, and critically use, evaluate, integrate, and analyze text (e.g., Block & Pressley, 2007). However, although these skills underlie successful reading comprehension, they are, themselves, dependent on the ability to efficiently recognize words (Perfetti, 2007; Stanovich, 1980). Thus, basic reading skills such as recognizing words and identifying meaning at the lexical level are core to the overall reading process (Kintsch, 1998). In the current study, it is hypothesized that deep orthographies exacerbate lexical-level reading difficulty, which, in turn, is associated with text-based comprehension weaknesses. As a result, basic reading and higher-level text comprehension are key constructs in this study, and are thought to be separate but interrelated processes.

Drawing from and extending the work of Hanley et al. (2004), who defined poor readers as students performing below the 25th percentile, the categories of poor and very poor readers were chosen in order to determine what percentage of the population may be affected by orthographic depth and therefore, in practice, might become the target of supplementary instruction designed to ameliorate reading difficulties. Parenthetically, these students may struggle to read because of a variety of factors (e.g., poor instruction, a poor home-literacy environment, speaking a nonstandard dialect, second language learning, and/or an undiagnosed learning disability), any of which may interact with the characteristics of the reader's orthography.

Poor and very poor readers performed between the 25th and 10th percentiles or below the 10th percentile, respectively, within his or her country and language group on relatively simple measures of reading achievement: the PIRLS' "Focus and Retrieve Explicitly Stated Information" and PISA's "Access and Retrieve" subsections. These scales measured the reader's ability to answer questions about information explicitly stated in the text, a skill that largely relies on efficient word recognition. These basic reading skills are thought to measure a separate construct than the higher-level reading and thinking skills measured in the outcome variable. Specifically, students had to read the text, access meaning on a basic level, and retrieve information contained directly in the text (Fleischman et al., 2010; Martin et al., 2007).

Language. This variable denotes language of the test and the classroom instruction.

Student background characteristics. Age, gender, language spoken at home, and SES are important control variables in this study. In PIRLS, students answered survey questions such as, "When were you born?" and "Are you a boy or a girl?" Parents or primary caregivers responded to questions on the home survey such as, "What language do most of the activities in Question 8 take place?" (Question 8 reported six literacy-related home activities), and "What is the highest level of formal education attained by the child's mother?" In PISA, students responded to, "Are you male or female?" "What language do you speak at home most of the time?" and "What is the highest level of schooling completed by your mother?"

Socioeconomic status. The mother's education level was used as a proxy for SES. The 1997 International Standard Classification of Education (ISCED) was used to report the mother's highest level of educational attainment. Level 0 is pre-primary education, Level 1 is primary education or the first stage of basic education, Level 2 is lower secondary or the second stage of basic education, Level 3 is upper secondary education, Level 4 is post-secondary non-tertiary education, Level 5 is the first stage of tertiary education, and Level 6 is the second stage of tertiary education (UNESCO, 2006).

Age. There was a grade/age tradeoff in terms of which inclusion criterion was the focus of each international assessment. Age was used as a control variable in the PIRLS analyses, because age varied greatly in the fourth grade sample. However, age was not included in any of the PISA analyses because the sample was limited to 15-year-old students.

School resources. In the PIRLS dataset, an index of available school resources was based on each principal's responses related to how much the school's capacity to provide instruction was affected by a shortage of resources (e.g., qualified teaching staff, reading and second-language teachers, materials, supplies, facilities, computers, and library books). The index was created by taking the average of each principal's responses, which were reported on a 4-point scale (1 = a lot, 2 = some, 3 = a little, and 4 = not at all) to the school resource questions. The index reflected a reliable scale across all countries participating in PIRLS (Cronbach's Alpha was .85; Trong & Kennedy, 2007). In the PISA dataset, the principal reported whether there was shortage of instructional materials by responding to the question: "Is your school's capacity to provide instruction hindered by a shortage or inadequacy of instructional materials?"

School identification. In order to capture unobserved school-level factors and account within school correlations in reading scores, a unique identification number was used as an indicator of a child's enrollment at a specific school.

Tables 5 and 6 contain the PIRLS and PISA questions and responses, respectively. Table 7 provides the coding scheme for all of the variables included in this study.

Statistical Techniques

As mentioned previously, in order to account for country-level characteristics that may largely affect student outcomes (e.g., values and beliefs about reading, income per capita, public expenditure on education, instructional method, and starting age of instruction), the analyses included only within-country comparisons. Although some of these factors may vary within countries as a function of language, comparing students from the same country makes it more likely that they are matched on many cultural, economic, and educational variables (Hanley et al., 2004). Furthermore, the within-country comparison approach relaxes the assumption that the regression coefficients for language are the same across countries, which may be unrealistic given the variance in dialects within languages and the degree to which they are represented by the orthography (i.e., forms of oral language are constantly changing even though written language does not; Craig & Washington, 2006). For example, although Canadian-, Dutch-, and Swiss-French share approximately the same writing system, spelling-to-sound consistency may vary as a result of differing linguistic and historical influences on the development of oral dialects, especially those with highly variant vowel production patterns (Frost, 2005).

Because the PIRLS and PISA data were collected from students within specific schools, they presumably have a multilevel structure. Thus, likelihood-ratio tests were conducted, comparing ordinary regression to hierarchical linear regression (HLM), to investigate whether a random intercept was needed for schools in the models for each country. Because all of the tests were significant, random intercepts for schools were included in all models. At the fourth-grade level, where students were nested in classrooms, and classrooms were nested in schools, additional likelihood-ratio tests indicated that random intercepts at the classroom level were not required in any country. As a result, two-level models emerged as the best fitting to the data in all analyses. The HLM approach is also advantageous in terms of controlling for unobserved school-level factors (Rabe-Hesketh & Skrondal, 2005; Raudenbush & Byrk, 2002).

Further likelihood-ratio and Wald tests suggested that models which included a variable describing the school's resources were not significantly better in any country. Thus, the school resources variable was dropped in all subsequent analyses, and was not included in the final models. Finally, homogeneity of variance tests were employed to determine whether there was greater variance in reading scores in deep orthographies compared to shallow orthographies. Appendix A provides the for model specification for the final models.

Results

Descriptive Statistics

PIRLS Belgium. In Belgium, there were 6,180 students from 387 schools with complete data for all of the variables of interest in this study. Dutch comprehension scores fell into the high range of possible scores for the entire PIRLS sample (M = 550.99, SD = 52.28). The students who took the test in Dutch (n = 3,968) were approximately 10 years old (M = 9.96, SD = .01). Approximately 9% (n = 342) were considered very poor readers and 15% (n = 577) poor readers. This subsample comprised 51% (n = 2,032) males and 49% (n = 1,936) females, who came from various SES backgrounds: 3% (n = 135) ISCED Level 1 or did not go to school, 13% (n = 523) ISCED Level 2, 33% (n = 1,310) ISCED Level 3, and 29% (n = 1,166) ISCED Level 5A or 1st degree. Of these students, approximately 95% (n = 3,757) spoke Dutch at home and 5% (n = 211) spoke another language.

French comprehension scores in Belgium fell into the lowest range of possible scores for the entire PIRLS sample included in this study (M = 507.48, SD = 65.99). The students who took the test in French (n = 2,212) were approximately 10 years old (M = 9.87, SD = .01).

Approximately 8% (*n* = 178) were considered very poor readers and 14% (*n* = 315) poor readers. This subsample included 51% (*n* = 1,119) males and 49% (*n* = 1,093) females, who came from various SES backgrounds: 7% (*n* = 148) ISCED Level 1 or did not go to school, 18% (*n* = 399) ISCED Level 2, 24% (*n* = 535) ISCED Level 3, 8% (*n* = 184) ISCED Level 4, 37% (*n* = 814) ISCED Level 5B, and 21% (*n* = 834) beyond ISCED Level 5A, 1st degree. Of these students, approximately 94% (*n* = 2,088) spoke French at home and 6% (*n* = 124) spoke another language. Table 8 provides the descriptive statistics for the PIRLS variables.

PIRLS Canada. There were 16,176 Canadian students from 1,120 schools with complete data for all of the variables this study. French comprehension scores fell into the mid range of possible scores for the entire PIRLS sample (*M* = 525.46, *SD* = 62.76). At the time of the data collection, the average age of participating students of interest who took the test in French (*n* = 3,333) was approximately 10 years old (*M* = 9.96, *SD* = .38). Approximately 9% (*n* = 304) were considered very poor readers and 14% (*n* = 462) poor readers. This subsample comprised 49% (*n* = 1,647) males and 51% (*n* = 1,686) females, who came from various SES backgrounds: 3% (*n* = 105) ISCED Level 1 or did not go to school, 5% (*n* = 154) ISCED Level 2, 16% (*n* = 531) ISCED Level 3, 40% (*n* = 1,330) ISCED Level 4, 22% (*n* = 720) ISCED Level 5B, 14% (*n* = 477) ISCED Level 5A, 1st degree, and 0% (*n* = 16) were coded as "n/a" (but were not considered missing). Of these students, approximately 80% (*n* = 2,659) spoke French at home and 20% (*n* = 674) spoke another language.

English comprehension scores in Canada fell into the highest range of possible scores for the entire PIRLS sample in this study (*M* = 562.72, *SD* = 64.91). The Canadian students who took the test in English (*n* = 12,843) were approximately 10 years old (*M* = 9.89, *SD* =.34). Approximately 9% (*n* = 1,113) were considered very poor readers and 14% (*n* = 1,798) poor readers. This subsample comprised 49% (*n* = 6,328) males and 51% (*n* = 6,515) females, who came from different SES backgrounds: 3% (*n* = 400) ISCED Level 1 or did not go to school, 5% (*n* = 657) ISCED Level 2, 25% (*n* = 3,242) ISCED Level 3, 35% (*n* = 4,558) ISCED Level 4, 18% (*n* = 2,375) ISCED Level 5B or 1st degree, 10% (*n* = 1,304) ISCED Level 5A, 1st degree, and 2% (*n* = 307) were coded as "n/a" (but were not considered missing). Of these students, approximately 91% (*n* = 11,663) spoke the English at home and 9% (*n* = 1,180) spoke another language.

PISA Belgium. In Belgium, there were 6,642 participating 15-year-old students from 271 schools. Dutch comprehension scores were in the high range for the overall PISA sample (*M* = 532.72, *SD* = 93.50). Of the students who took the test in Dutch, 8% (*n* = 292) were classified as very poor readers and 13% (*n* = 515) as poor readers. This subsample comprised 50% (*n* = 1,951) males and 50% (*n* = 1,923) females, who came from various SES backgrounds: 1% (*n* = 48) did not complete ISCED Level 1, 2% (*n* = 91) ISCED Level 1, 6% (*n* = 220) ISCED Level 2, 17% (*n* = 650) ISCED Level 3B or 3C, 74% (*n* = 2,865) ISCED Level 3A. Of these students, approximately 76% (*n* = 2,932) spoke Dutch at home and 24% (*n* = 942) spoke another language.

In Belgium, 2,786 students took the test in French and comprehension scores were in the low range for the overall PISA sample (*M* = 505.34, *SD* = 109.66). Of the students that took the test in French (*n* = 196), 7% (*n* = 196) were classified as very poor readers and 13% (*n* = 372) as poor readers. This subsample consisted of 51% (*n* = 1,407) males and 49% (*n* = 1,361) females, who came from different SES backgrounds: 2% (*n* = 58) did not complete ISCED Level 1, 4% (*n* = 108) ISCED Level 1, 9% (*n* = 236) ISCED Level 2, 15% (*n* = 426) ISCED Level 3B or 3C, 70% (*n* = 1,940) ISCED Level 3A. Of these students, approximately 83% (*n* = 2,295) spoke

French at home and 17% (*n* = 473) spoke another language. Tables 9 and 10 provide the descriptive statistics for the PISA variables.

PISA Canada. In Canada, there were 21,531 15-year-old students from 972 schools with complete data for all of the variables of interest in this study. Of these students, 5,169 took the test in French, demonstrating average comprehension ability in the low range for the overall PISA sample (*M* = 511.31, *SD* = 85.72). Eight percent (*n* = 423) of this group was classified as very poor readers and 14% (*n* = 718) as poor readers. This subsample comprised 47% (*n* = 2,419) males and 53% (*n* = 2,750) females, who came from various SES backgrounds: less than 1% (*n* = 23) did not complete ISCED Level 1, 1% (*n* = 64) ISCED Level 1, 7% (*n* = 355) ISCED Level 2, 91% (*n* = 4,727) ISCED Level 3A. Of these students, approximately 70% (*n* = 3,629) spoke French at home and 30% (*n* = 1,540) spoke another language.

In English, average reading comprehension scores were in the high range (*M* = 532.58, *SD* = 92.18). Of the students who took the test in English (*n* = 16,362), 9% (*n* = 1,416) were classified as very poor readers and 14% (*n* = 2,367) as poor readers. This subsample included 49% (*n* = 8,076) males and 51% (*n* = 8,286) females, who came from various SES backgrounds: 1% (*n* = 95) did not complete ISCED Level 1, 1% (*n* = 136) ISCED Level 1, 6% (*n* = 979) ISCED Level 2, 93% (*n* = 15,152) ISCED Level 3A. Of these students, approximately 90% (*n* = 14,698) spoke English at home and 10% (*n* = 1,664) spoke another language.

PISA Finland. In Finland, there were 5,615 15-year-old students from 203 schools with complete data for all of the variables of interest in this study. Finnish comprehension scores were in the highest range of the PISA scores in the overall sample included in this study (*M* = 539.76, *SD* = 85.73). Of the students who took the test in Finnish (*n* = 4,278), 10% (*n* = 411) were classified as very poor readers and 15% (*n* = 629) as poor readers. This subsample comprised 49% (*n* = 2,102) males and 51% (*n* = 2,176) females, who came from various SES backgrounds: 1% (*n* = 22) did not complete ISCED Level 1, 4% (*n* = 183) ISCED Level 1, 6% (*n* = 259) ISCED Level 2, 44% (*n* = 1,884) ISCED Level 3B or 3C, 45% (*n* = 1,930) ISCED Level 3A. Of these students, approximately 97% (*n* = 4,170) spoke Finnish at home and 3% (*n* = 108) spoke another language.

Swedish comprehension scores were in the mid range (*M* = 515.15, *SD* = 86.46). Of the students who took the test in Swedish (*n* = 1,337), 9% (*n* = 121) were classified as very poor readers and 15% (*n* = 199) as poor readers. This subsample consisted of 47% (*n* = 624) males and 53% (*n* = 713) females, who came from different SES backgrounds: less than 1% (*n* = 1) did not complete ISCED Level 1, 4% (*n* = 50) ISCED Level 1, 6% (*n* = 75) ISCED Level 2, 47% (*n* = 635) ISCED Level 3B or 3C, 43% (*n* = 576) ISCED Level 3A. Of these students, approximately 78% (*n* = 1,049) spoke Swedish at home and 22% (*n* = 288) spoke another language.

PISA Switzerland. In Switzerland, 10,359 15-year-old students from 426 schools participated in this study. Italian comprehension scores were in the lowest range of possible scores for the overall PISA sample included in this study (*M* = 462.12, *SD* = 78.26). Of the students who took the test in Italian (*n* = 442), 8% (*n* = 36) were classified as very poor readers and 16% (*n* = 72) as poor readers. This subsample comprised 57% (*n* = 251) males and 43% (*n* = 191) females, who came from various SES backgrounds: 1% (*n* = 6) did not complete ISCED Level 1, 2% (*n* = 10) ISCED Level 1, 25% (*n* = 109) ISCED Level 2, 30% (*n* = 134) ISCED Level 3B or 3C, 41% (*n* = 183) ISCED Level 3A. Of these students, approximately 68% (*n* = 302) spoke Italian at home and 32% (*n* = 140) spoke another language.

German comprehension scores were in the low range (*M* = 503.19, *SD* = 90.82). Of the students who took the test in German (*n* = 6,031), 8% (*n* = 469) were classified as very poor readers and 13% (*n* = 785) as poor readers. This subsample included 50% (*n* = 3,028) males and 50% (*n* = 3,003) females, who came from different SES backgrounds: 1% (*n* = 68) did not complete ISCED Level 1, 2% (*n* = 125) ISCED Level 1, 24% (*n* = 1,462) ISCED Level 2, 56% (*n* = 3,358) ISCED Level 3B or 3C, 17% (*n* = 1,018) ISCED Level 3A. Of these students, approximately 85% (*n* = 5,137) spoke German at home and 15% (*n* = 894) spoke another language.

French comprehension scores were also in the low range relative to the other linguistic groups included in this study (*M* = 500.42, *SD* = 87.92). Of the students who took the test in French (*n* = 3,886), 9% (*n* = 338) were classified as very poor readers and 14% (*n* = 550) as poor readers. This subsample comprised 50% (*n* = 1,949) males and 50% (*n* = 1,937) females, who came from various SES backgrounds: 3% (*n* = 107) did not complete ISCED Level 1, 4% (*n* = 160) ISCED Level 1, 20% (*n* = 763) ISCED Level 2, 46% (*n* = 1,787) ISCED Level 3B or 3C, 28% (*n* = 1,069) ISCED Level 3A. Of these students, approximately 83% (*n* = 3,234) spoke French at home and 17% (*n* = 652) spoke another language.

Missing Data

Listwise deletion was used as a method for handling missing data. Independent sample *t*-tests were conducted to address the possibility that there might be patterns of bias in this approach. In all countries, missing data were due primarily to lower response rates to questions on the home surveys, indicating that students with incomplete data were, or had parents who were, less willing or able to fill out the home survey. Overall, there were consistent patterns across languages such that students with missing data had significantly and consistently lower reading comprehension scores than students with complete data in all countries, datasets, and language groups. The one exception was in Belgium where the difference in performance between 15-year-old French-speaking students with missing and complete data (1.3 standard deviations) was much greater than this difference in the Dutch-speaking sample (approximately .83 standard deviations). This suggests that the poor readers excluded from the French sample were performing considerably lower than the poor readers excluded from the Dutch sample. Generally, these results indicated that although missing data could not be considered random, they did not introduce bias because each of the compared language groups in each country was similarly affected. Thus, it is argued that that the consistent pattern of missingness allows us to draw conclusions about the data.

The missing data did present a limitation, however, in that a significant portion of poor readers was excluded in all language subgroups (and especially from the 15-year-old French sample in Belgium), thus probably compromising the power of the analysis to detect differences across orthographies for precisely the subpopulation of most interest. Table 11 provides the percentages of missing data and the mean differences in reading comprehension scores between students with missing and complete data by language subgroup.

Hierarchical Linear Modeling Results

Main effects of language on reading comprehension. The first set of models addressed the first hypothesis and sought to determine whether there was a main effect of orthographic depth on reading comprehension ability across countries. Two-level, hierarchical regression models (with random intercepts for schools) were specified because likelihood-ratio tests indicated that there was significant between-school variability in all countries, after controlling for the covariates. Residual diagnostics suggested that the assumptions (normality and homoscedasticity) were met for the level-one and level-two residuals in all of the final models. Tables 12, 13, and 14 provide the results from the multilevel regression analyses and report the unstandardized beta coefficients. Table 15 shows the effect size results (for the main effect of orthography and interactions between orthographic depth and reading ability) in terms of standard deviation differences.

Across countries and datasets, a consistent pattern of main-effect results emerged: Students who spoke the majority language in their country performed significantly and considerably above students who spoke a minority language, when controlling for reading ability, gender, SES, language spoken at home, and age—regardless of the depth of the orthography. This result indicated that there did not appear to be a lasting effect of orthographic depth on reading ability for competent and skilled readers beyond the third grade. In Belgium, scores for fourth-grade Dutch-speaking students were estimated to be 42.81 points higher, on average, than French speaking students ($p < .001$, $d = .67$), and 15-year-old Dutch-speaking students' comprehension skills were estimated to be 34.7 points higher, on average, than 15-year-old French-speaking students' ($p < .001$, $d = .29$). In Canada, English-speaking fourth-grade students were estimated as 37.79 points higher, on average, than French-speaking students ($p < .001$, $d = .56$), and on average, English-speaking 15-year-old students' scores were estimated as 18.79 points above French-speaking 15-year-olds' scores ($p < .001$, $d = .20$). In Finland, Finnish-speaking students were estimated to outperform Swedish-speaking students, on average, by 25.31 points ($p < .001$, $d = .29$). In Switzerland, although there was no difference between German- and French-speaking students, French- and German-speaking students were estimated to perform 40.47 and 36.16 points, respectively, above Italian-speaking students ($p < .001$, $d = .44$ and $p < .001$, $d = .39$).

These results suggest that, while there appears to be no long-term role of orthographic depth in reading comprehension performance for competent and skilled readers across counties, there appears to be a beneficial effect of reading in the language spoken by the majority of the population, even after controlling for those variables that serve as proxies for socioeconomic status, such as parental education. This benefit may be a result of students having increased exposure to spoken and written language in everyday life (Gathercole & Thomas, 2009), curricular materials may be more advanced or carefully developed, educational systems more consistent and established (Wiley & Wright, 2004), and/or underlying dominant language ideologies related to linguistic diversity (Wiley & Lukes, 1996). Overall, the results from the first set of multilevel models that addressed Hypothesis 1 appear to indicate that reading in the majority language (regardless of its orthographic depth) is a considerable advantage for the reader.

Interaction effects: Reading ability and orthographic depth. Although reading in the majority language may provide an additional level of support, orthographic depth appears to play an enduring role in reading comprehension for the poorest of readers. The set of multilevel models exploring Hypothesis 2 suggest that even beyond the third grade, many of the poorest 10 to 25% of readers of relatively deep orthographies appear to be continually challenged by the greater degree of ambiguity in their script. Conversely, there may be long-term beneficial effects of reading in relatively shallow orthographies for poor readers. Findings indicate that there were significant orthographic depth by reading ability interaction effects in Belgium, Canada, and Switzerland at both age levels, when controlling for gender, SES, language spoken at home, and age. In Belgium, fourth-grade Dutch-speaking poor readers scored 20 points higher, on average, than French-speaking poor readers ($p < .001$, $p < .001$, $d = .32$), and Dutch-speaking very poor readers scored, on average, 31.08 points higher than their French-speaking counterparts ($p < .001$, $p < .001$, $d = .49$). At the 15-year-old level, Dutch-speaking students performing between the 10th and 25th percentile scored 13.8 points higher than French-speaking students performing between the 10th and 25th percentile ($p < .001$, $p < .001$, $d = .12$), and very poor readers of Dutch (below the 10th percentile) scored 9.36 points higher than very poor readers of French, however, the difference was not significant. In Canada, although poor fourth-grade French readers were only slightly above poor fourth-grade English readers ($\beta = .87$), and the difference was not significant, very poor French readers were significantly and considerably above very poor English readers ($\beta = 11.07$, $p < .001$, $d = .17$). At the 15-year-old level, both poor ($\beta = 8.95$, $p < .01$, $d = .10$) and very poor readers ($\beta = 17.11$, $p < .001$, $d = .18$) of French scored significantly above English-speaking counterparts. In Finland, there were no significant differences between poor and very poor readers of Finnish and Swedish—possibly because there may be little meaningful difference in orthographic depth between the two languages. In Switzerland, although there was no difference between poor and very poor readers of French and German, very poor readers of Italian scored 33.86 points higher, on average, than readers of French ($p < .01$, $d = .44$) and 31.28 points higher than their German-speaking counterparts ($p < .001$, $d = .39$). These results indicate that the majority of struggling readers of shallow orthographies were at a considerable advantage over struggling readers of deep orthographies, providing cautious support for the proposition that there are long-term adverse effects of reading in a deep orthography for poor readers—regardless of whether it is the majority or minority language in the country.

Support for Hypothesis 3 is evidenced in the effect size results, which tend to indicate that the gap in reading performance between readers of shallow and deep orthographies is the largest between very poor readers, followed by poor readers, yet appears to attenuate in the competent/skilled reader range. Additionally, while struggling readers appear to be at a double disadvantage when the minority language has the relatively deep orthography in the country (e.g., Belgium), orthographic depth is associated with a much narrower gap when the minority language is represented by the more shallow orthography (e.g., Canada). However, there did not appear to be support for Hypothesis 4—that the gap between poor readers of different orthographies would widen as students developed from fourth-grade to the 15-year-old level—as effect sizes were not larger for older students. Figures 2 through 7 are graphic displays of the interaction effects in each country. Note that these figures provide examples of the relationship between orthographic depth and reading ability for female students, from the same SES category, who spoke the language of the test at home, and had the mean age in their country.

Homogeneity of Variance Test Results

Homogeneity of variance tests were conducted (without controlling for the covariates) to address Hypothesis 5, and suggest that there may be greater variance in reading scores in deep orthographies compared to shallow orthographies in fourth grade and at the 15-year-old level. Across languages and datasets (except for the French/German comparison in Switzerland), readers of languages with deep orthographies exhibited more variance in comprehension ability than readers of languages with shallow orthographies. This finding extends previous research that has demonstrated the existence of this relationship during the beginning reading process (Hutzler et al., 2004; Seymour et al., 2003; Zeigler & Goswami, 2005). These results suggest that the added challenges of reading in a deep orthography (specifically for poorer readers) may create more variability in how students perform during the reading comprehension task throughout development. Table 16 provides the results from the homogeneity of variance tests in each country.

In sum, hierarchical linear modeling and variance tests were conducted to address the research questions. The results from this study provide modest evidence to support the conclusion that after the third grade: (a) the role of orthographic depth is diminished for competent and skilled readers, (b) the critical linguistic factor in reading comprehension for this group is whether or not the student speaks the language spoken by the majority of the population in his or her country, and (c) as students decrease in reading ability, they may be increasingly and continually challenged by reading in a language with a deep orthography. Moreover, it should be noted that struggling readers of a relatively deep orthography, that is also the minority language in the country, may be at a double disadvantage when it come to the comprehension process. Finally, (d) in relatively deep orthographies, there also appears to be more variance in reading comprehension skill, presumably because there is more inherent difficulty in the reading task.

Discussion

The relationship between orthographic depth and reading performance is more complex than previously considered. Because beginning word reading is the process of converting spelling to sound, and each phonology differs in the way it is represented by its orthography, it has been argued that a comprehensive model of reading acquisition must account for such differences across languages (i.e., the orthographic depth hypothesis; Frost et al., 1987). There is a preponderance of evidence that supports the orthographic depth hypothesis and demonstrates that the degree of ambiguity in spelling-to-sound consistency in a language affects the ease and effectiveness with which children learn to read. Furthermore, English, with its extreme degree of irregularity, is purported to be the most difficult alphabetic language in which to acquire beginning reading skills (e.g., Aro & Wimmer, 2003; Goswami et al., 1997; Landerl, 2000; Porpodas, 1999; Seymour et al., 2003; Öney & Goldman, 1984). Seymour et al. (2003), Seymour and Evans (1999), and Duncan and Seymour (2000) demonstrate this relationship, and their findings point to

> 7 years as the reading age at which foundation literacy acquisition [is] normally complete … [suggesting] … that readers of English require 2 1/2 or more years of literacy learning to achieve a mastery of familiar word recognition and simple decoding which is approached within the first year of learning in a majority of European languages (Seymour, et al., 2003, p. 167).

Indeed, Seymour et al. (2003) estimate that the acquisition of fluent and accurate word identification in English is slower than in other alphabetic languages by a ratio of 2.5:1.

Looking across the cross-linguistic reading research, it is clear that orthographic differences should be taken into account during the beginning reading process; however, little is known about whether such differences continue to have an influence on reading comprehension achievement as children advance throughout schooling. Yet there is wide agreement that by fourth grade the majority of children have transitioned from *learning to read* to *reading to learn* in all languages (i.e., they have mastered the word-level skills necessary for the comprehension of text; e.g., Joncas, 2007; Seymour et al., 2003). Thus, it is assumed that orthographic factors cease to play a role in the literacy process. However, because there is evidence of a long-term relationship between word reading and text comprehension (Cunningham & Stanovich, 1997; Perfetti & Bolger, 2004; Vellutino & Scanlon, 1987), this study sought to challenge the notion that reading is the same in all languages after third grade by investigating the continued role of orthographic depth in reading comprehension performance at the fourth-grade and 15-year-old levels. Thus, the possibility that our understanding of skilled reading and reading difficulties throughout development must also include consideration of the differences across orthographies was as the forefront of this study.

Drawing from the orthographic depth hypothesis (Frost et al., 1987), Kintsch's (1998) work on the interactive and mutually supportive relationship between word reading and the overall comprehension process, and Perfetti's lexical quality hypothesis (1992, 2007), the specific goals of this study were to investigate whether spelling-to-sound inconsistency is associated with (1) lower average literacy achievement, (2) additional long-term reading difficulties in poor readers, (3) even greater weaknesses in reading comprehension performance for the poorest readers, (4) added reading difficulties for older students versus younger students, and (5) greater variability in reading performance. Overall, the results from this study indicate that by fourth grade, orthographic depth may cease to play a role in reading comprehension achievement for competent and skilled readers, and that the critical linguistic factor for this group is whether or not they speak the language spoken by the majority of the population in their country. But for poor readers, this study demonstrates a different relationship and provides cautious support for the continued association between orthographic depth and poor reading achievement beyond the third grade. Furthermore, the effect of orthographic depth on text comprehension appears to increase as students decrease in reading ability. Thus, because poor readers have not fully mastered sound-to-spelling relationships in deep orthographies, it is argued that orthographic depth may be related to greater variability in reading comprehension performance throughout development.

Contributions to Research

A number of meaningful relationships between language and reading comprehension were observed in this study. Notably, the results from this study build on and extend the cross-linguistic reading research, which has sought to address the level of specificity regarding the orthographic depth hypothesis. One major contribution is that, while effect sizes tended to be small, the results from this study provide cautious support for the continued role of orthographic depth in poor reading ability at the fourth-grade and 15-year-old levels. An explanation for the small effect sizes is that, despite the fact that many cross-linguistic studies have sampled intact classrooms, they have typically not employed hierarchical designs, which take into account the clustered nature of many educational data. Thus, it may be the case that the smaller effect sizes actually reveal a more accurate representation of the relationship between orthographic depth and reading achievement, where a significant portion of the variance in reading performance can be explained at the classroom level. Further cross-linguistic research, especially studies that employ experimental designs and take the hierarchical nature of the measurement into account, will enhance the generalizability of these findings.

Main Effects of Language on Reading Performance

Competent and skilled comprehension. Across countries, and at all age levels, a consistent pattern of main-effect results emerged: Students who spoke the majority language in their country performed significantly and considerably above students who spoke a minority language, regardless of the level of ambiguity in the reader's script, and when controlling for SES. This result indicated that orthographic depth did not appear to have a lasting role in competent and skilled reading achievement beyond the third grade.

However, one limitation of this study is that the measures utilized in large-scale databases like PIRLS and PISA may lack the sensitivity needed to fully test the hypotheses, and any lingering effects of orthographic depth on competent or skilled reading might be too small or varied to make solid inferences about the individuals in the sample. It is also important to acknowledge that although the question of whether variation in spelling-to-sound consistency is related to long-term reading difficulties was central to this study, it is possible that there could be a continued impact of orthographic depth on reading performance, yet differences across languages do not actually cause reading difficulties in competent and skilled readers. For example, when Ziegler et al. (2001) compared adult readers in German and English on measures that were extremely sensitive to detecting differences in adult skilled reading and perfectly matched across languages (i.e., identical words and nonwords such as *zoo* and *sand*), the authors found that readers employed different processing strategies (small-, large-, and/or flexible-grain size unit approaches) in response to the constraints of the writing system. Similarly, in their psycholinguistic grain size theory, Ziegler and Goswami (2005) posit that differences across languages in terms of which units are most regular in the writing system (i.e., whether the orthography is regular at the individual letter level or if larger units lead to more regularity) leave developmental "footprints" that can be observed in adult skilled reading (p. 3). Thus, it may be the case that even after readers reach some threshold of reading proficiency, orthographic factors continue to influence the reading process (i.e., via differential strategy use), however, they do not affect skilled reading comprehension and thus cannot be detected by the PIRLS and PISA measures.

Lasting Effects of Orthographic Depth on Poor Reading Comprehension

Although the results from this study suggest that spelling-to-sound consistency may cease to play a role in competent and skilled reading, they also suggest that poor readers may continue to struggle when there is a greater degree of ambiguity in their script. In the present study, poor readers (between the 10th and 25th percentile) at the fourth-grade level were at a considerable advantage reading in Dutch, which is relatively shallow compared to the French orthography, and very poor fourth-grade readers (below the 10th percentile) of shallow orthographies had significantly higher reading comprehension scores than their counterparts reading in deep orthographies in both Belgium and Canada. At the 15-year-old level, shallow orthographies were advantageous for poor readers in Belgium and Canada, and very poor readers in Canada and Switzerland. At the country level, this orthographic depth by reading ability interaction functioned by closing the majority/minority achievement gap when the majority language had a deep orthography (i.e., Canada and Switzerland), and by widening the gap when the majority language had the relatively shallow orthography (i.e., Belgium).

The results from this study did not indicate that reading difficulties related to orthographic depth are exacerbated over time such that the gap between poor readers of relatively deep and shallow orthographies becomes more apparent as students get older. However, it is possible that this hypothesis could not be adequately tested if the PIRLS and PISA measures were assessing substantively different constructs. Further research, using vertically aligned instruments, is needed to clarify the longitudinal relationship between orthographic depth and reading comprehension performance.

Although it remains unclear whether reading difficulties become more severe over time in languages with deep orthographies, it does appear that throughout development, the association between orthographic depth and reading ability is the strongest in students with the poorest reading ability. This relationship is observed in the increasing pattern of effect sizes.

There appears to be little to no difference between competent/skilled readers of different orthographies, however, small but meaningful effect sizes were observed in the poor-reader range, and the performance differences between shallow and deep orthographies became even more visible when students were performing below the 10th percentile in reading ability. Furthermore, these results held even when controlling for home language (i.e., if the student spoke a different language at home than the test/school language), SES, and regardless of whether reading occurred in the majority or minority language of the country.

One explanation for this result is that deep orthographies bring about added challenges in the reading process simply because there are more phoneme representations per grapheme to learn. Specifically, it is possible that by fourth grade, given the greater number of sound-symbol relationships, poor readers of deep orthographies have yet to form the level of precision in lexical quality—orthographic representations and their bindings to words' other constituent identities—that is necessary for efficient reading comprehension (Perfetti, 1992). In contrast, the reliable, transparent, and predictable nature of shallow orthographies may more readily facilitate poor readers' ability to form "fully specified orthographic representation[s] … [that are] … bonded to … phonemic representation[s]" so that they can be accessed efficiently, without competing with other similar and partly activated word candidates (Perfetti, 1992, p. 160). By this argument, in shallow orthographies it is easier to develop fast and automatic word identification, allowing the process to become subconscious. This advantage supports the reader's ability to allocate cognitive resources to higher-level reading comprehension processes such as adopting a high standard of coherence, attending to the syntactic and semantic structure of language, and evaluating and integrating text (Perfetti, 1992, 2007; Perfetti et al., 2005; Stanovich, 1980). In sum, poor readers may continue to be challenged or confused when there is a greater degree of spelling-to-sound inconsistency in a writing system, relative to poor readers of orthographies in which this relationship is more limited and straightforward.

Dyslexia in different languages. One question that arises in the context of the current study is determining the initial cause of students' reading difficulties. Although students with documented disabilities were specifically excluded from the PIRLS and PISA databases, it is possible that experiential or biological factors did limit the poor, and especially very poor, reading behavior observed in this study. For example, some students with undiagnosed reading disabilities may have participated in the PIRLS and PISA studies. Nevertheless, it has been argued that the primary cause of reading difficulties is limited language and literacy experience, and that experiential factors such as poor teaching can lead to the same reading difficulties as dyslexia. Furthermore, reading-related cognitive skills (e.g., phonological awareness) are thought to be continuous in the population, and thus while the poor readers in this study did not officially meet the "cutoff" criteria for disabilities, they might be performing in the lower range of the distributions of these important skills (Torgesen et al., 2001; Vellutino, Fletcher, Snowling, & Scanlon, 2004). It therefore makes sense to consider how the results from this study might be interpreted within the context of the research on dyslexia in different languages.

It has been well established in the English-language research that the reading profiles of students with dyslexia generally include difficulty detecting and manipulating the sound structure of oral language (i.e., phonological awareness), which leads to difficulty decoding, and which further disrupts fluent and accurate word recognition and spelling (e.g., Snowling, 2001; Vellutino et al., 2004). However, an emerging body of cross-linguistic literature has contested the notion that phonological skills are universally core to reading disability (e.g., de Jong & van der Leij, 1999, 2003; Wimmer, Landerl, & Frith, 1999). One hypothesized difference across languages is that readers of shallow orthographies need to rely on their phonological skills only during the first two years of reading, after which point they can rely on relatively automatic word recognition. Specifically, it is thought that the transparent nature of shallow orthographies more readily facilitates phonological recoding (Share, 1995) and the amalgamation among the phonological, orthographic, and semantic forms of language (Ehri, 2005), which bring about the large body of immediately recognizable sight words that are necessary for the fluid comprehension of text. That is, decoding and automatic word recognition become established with relatively less practice in shallow orthographies compared to deep orthographies, independent of the reader's initial level of phonological awareness (de Jong & van der Leij, 1999, 2003; Wimmer, Landerl, & Frith, 1999).

Although the English language research has demonstrated that reading difficulties related to phonological weaknesses persist into adulthood (Bruck, 1992; Pennington et al., 1999), this longitudinal relationship has not been observed in readers of shallow orthographies. For example, van Daal and van der Leij (1999) found that Dutch adolescents with dyslexia performed at the same level as their typical peers on measures of phonological awareness (e.g., phoneme segmentation), and Landerl and Wimmer (2000) found that normal and disabled German nonword spelling performance was comparable after only a few months of reading and spelling instruction. These researchers have also pointed out that instructional methods—which can be influenced by the constraints of the writing system—may also vary in the extent to which they facilitate phonological skills (de Jong & van der Leij, 2003). Specifically, the straightforward phonics approach, typical in shallow orthographies, strongly supports phonological recoding relative to instruction that includes larger-unit and whole-word methods that pertain more to English.

Although there have been some consistently different results regarding dyslexia across languages, it should also be noted that cross-linguistic researchers suggest that their work does not imply that phonological weaknesses cease to exist in shallow orthographies (de Jong & van der Leij, 2003). Instead, they argue that phonologically-related manifestations of dyslexia, such as the extent to which underlying weaknesses can be observed in reading performance, disappear after the second year of literacy learning (Landerl, 2000).

When considering the inconsistent results from the research on dyslexia in different languages, one possible explanation for the observed differences in reading comprehension outcomes between the poor readers of different orthographies in this study is that getting to the automatic level of word recognition necessary to support reading comprehension in shallow orthographies requires less reliance on the phonological system. It is also possible that a related factor is whether letter/sound or spelling-body/pronunciation-rime correspondences are more regular in the orthography (Ziegler & Goswami, 2005). Readers with limited reading-related cognitive skills may continue to be challenged when reading requires the adoption of a more complex and flexible set of decoding strategies, which places more demands on the phonological system. It is also possible that instructional methods that, may out of necessity in a deep orthography, focus on larger sublexical units and whole words are less effective at promoting students' phonological recoding skills.

The cross-linguistic dyslexia research has also contributed to the debate about rapid automatized naming (RAN)—the efficiency of retrieval of phonological codes from permanent memory—and its role in word recognition (Norton & Wolf, 2012). While findings are mixed as to whether RAN is a component of English-language dyslexia, it has been argued that RAN is more highly predictive of reading difficulties in shallow orthographies than are phonological skills, an argument that appears to be compatible with the behavioral data (e.g., de Jong & van der Leij, 1999).

One observed difference across languages is that English-speaking children with dyslexia have been found to have persistent weaknesses in reading accuracy—error rates have been estimated to be between 50 and 70% (Rack, Snowling, & Olson, 1992)—while readers of relatively shallow orthographies tend to have consistently lower rates of inaccuracy; e.g., 6% in Dutch (de Jong & van der Leij, 2003) and 25% in French (Sprenger-Charolles, Siegel, Béchennec, & Serniclaes, 2003). However, poor reading fluency appears to be common to dyslexia in all of the alphabetic languages, which may be related to phonological (and possibly RAN) weaknesses in English, yet only RAN deficits in shallow orthographies (Caravolas, 2005). Thus, while English-language researchers have continued to debate whether, in addition to phonological skills, RAN is a unique and important construct related to word reading, a less well-established body of cross-linguistic research suggests that both of these cognitive skills vary in readers, yet reliance on such skills is dependent on the demands of the orthography. Nevertheless, regardless of which cognitive skills actually underlie the reading process, it is clear that very poor readers exhibit difficulties in both fluency and accuracy in deep orthographies, while they typically struggle with fluency only in shallow orthographies. These differing fluency and accuracy weaknesses may be related to disparities in reading comprehension performance across languages throughout development.

To summarize, there are several possible explanations for the poor reading comprehension differences observed in this study: (1) there is simply a larger quantity of letter-sound relationships to learn in deep orthographies (Frost et al., 1987); (2) in shallow orthographies, reading words (and thus reading instruction) is mostly a straightforward, serial, letter-by-letter decoding process, while deep orthographies require the reader to operate flexibly at multiple levels of correspondence between spelling and sound (Ziegler & Goswami, 2005); and/or (3) readers must draw on a wider range of cognitive skills in deep orthographies (i.e., phonology and RAN) compared to shallow orthographies (i.e., RAN; Caravolas, 2005). Nevertheless, regardless of which of these explanations best portrays the differing relationship between word reading and comprehension across languages, it appears that orthographic depth has a lasting role in poor reading achievement.

Implications

One important way to address the persistent comprehension difficulties faced by poor readers of deep orthographies is by improving reading instruction. This group of readers might benefit from long-term and explicit teaching about the regularities of written language, and from extensive practice translating print into speech. For example, it might be especially helpful to teach students that the English orthography is more regular at the level of the spelling-body/pronunciation rime, and students may benefit from increased instruction in phonograms, morphological patterns, and word families (Venezky, 1999). Additionally, it is important that students understand regular orthographic patterns at the level of graphemes. For example, in English, while some letters almost always make the same sound (e.g., *t*), others change their pronunciation depending on their proximity to other letters (e.g., *r* and *l* controlled vowels in words such as *car* and *salad*; Moats, 2011). It is especially important that teachers understand and are prepared to teach students about the structure of a language with a deep orthography.

While there is a considerable amount of research that suggests morphology instruction directly improves vocabulary and comprehension (Nagy & Anderson, 1984; Nagy, 2007), learning to read words in the context of their morphological and etymological patterns might also indirectly improve comprehension by facilitating decoding and orthographic learning. Venezky (1999) called this approach to word reading the morphographemic level of analysis. For example, many words with Latin origins contain the root -*gress*, "to step," and these words (e.g., *progress*, *regress*, *congress*, *digress*, *transgress*, *egress*) are usually spelled and pronounced the same way. Similarly, teachers might point out common patterns contained in words with Germanic roots such as common vowel teams (e.g., *sixteen*, *road*, *meat*, and *soon*). When decoding and word reading are taught in the context of morphology and etymology, students are more likely to recognize that most words follow some orthographic-phonological pattern and that very few are rule breakers (e.g., *sugar*, *orange*, *gnome*; Venezky, 1999). Additionally, educational policies should account for the possibility that poor readers of deep orthographies may need more time and practice to form the well-specified orthographic representations that are necessary to comprehend text, by continuing word-level instruction into the later grades.

While teachers should strengthen, expand, and possibly lengthen their word-level instruction into the later grades, they should also place a consistent emphasis on comprehension instruction—both in spoken and written language—throughout schooling. For example, instruction that focuses on improving students' oral vocabularies and their comprehension of complex forms of spoken language can occur simultaneously with instruction that provides students with texts at appropriate readability levels, increasing the likelihood that poor readers can both read and access the meaning of print.

Overall, teachers can do more of all of the things we already know work to improve reading: teach phonological awareness, decoding, vocabulary, morphology, and comprehension, and foster students' interest in reading and their self-efficacy as readers (Moats, 2011). In fact, teachers should be skilled at providing students with a multitude of different types of reading experiences so that students continue to develop and refine both their word reading and comprehension skills into adulthood. With increased practice and success making the connections among the phonological, orthographic, semantic, syntactic, prosodic, and discourse forms of language, we can increase the likelihood that readers of all orthographies will successfully comprehend text. Finally, because getting to a level of reading comprehension proficiency appears to take longer and requires more work for poor readers of deep orthographies compared to shallow orthographies, it should be acknowledged that large-scale international reading assessments such as PIRLS and PISA might not be comparable across languages.

Limitations and Future Research

One limitation of this study is that due to the natural linguistic make-up of the countries and the need to control for sociocultural differences across countries—e.g., educational systems, public expenditure on education, and starting age of instruction—it was not possible to make comparisons of languages that are dramatically different in terms of orthographic depth. This limitation may be highlighted by the nonsignificant orthographic depth by reading ability interaction in Finland, where Finnish and Swedish are only slightly different in orthographic depth. It is also possible that if future research can account for differences at the country level and compare reading in languages at the opposite end of the orthographic depth continuum (e.g., Finnish and English), larger differences might actually be found in reading comprehension performance than are shown in the present study.

Another limitation of this study is that the observed relationships were correlational; future research employing experimental designs is needed to elucidate the orthographic aspects of reading. Furthermore, the measures available in the PIRLS and PISA databases may not have been the most ideal in answering the research questions. For example, it would have been informative to have measures of students' phonological processing and RAN abilities, and vertically aligned measures may have better enabled comparisons between the fourth-grade and 15-year-old levels. However, these limitations are balanced by the fact that standardized large-scale educational surveys such as PIRLS and PISA permit investigations of the structural invariance of the orthographic predictors of literacy achievement at a scale inaccessible to most individual researchers. Further research that employs more sensitive and vertically aligned measures of word reading and comprehension ability, measures a wider range of reading-related cognitive skills, and is experimental in design will improve the generalizability of this study's results.

Conclusion

Like beginning readers, older students with poor reading ability may continue to experience the added challenges of reading in a deep orthography. Specifically, writing systems in which there is a greater degree of spelling-to-sound ambiguity appear to draw on a wider range of cognitive skills than transparent orthographies, which makes achieving the level of fluent and accurate word recognition necessary to successfully comprehend text a relatively difficult task. As such, a comprehensive model of reading development must include an understanding that both learning to read *and* reading during the later grades may be specific to the demands of the orthography. Additionally, educational policy and practice should account for the considerable number of normal readers of deep orthographies (approximately 10 to 25%) that may continue to struggle with print throughout development. On a grand scale, educational systems within which students are reading in deep orthographies, should be prepared to expect and accommodate a greater degree of diversity in students' reading abilities. This study underscores the importance of teacher training in the area of reading, especially in deep orthographies such as English. Teachers need thorough knowledge of the factors that support reading development, and they should understand and be able to teach students about the structure of such languages.

In sum, the relationship between orthographic depth and reading achievement is more complex than previously understood. With replication and varied research designs, we will further deepen our knowledge of the linguistic aspects of reading, knowledge that can then be used to improve our ability to support all students on the path to becoming literate.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Aikens, N. L., & Barbarin, O. (2008). SES differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*(2), 235–251. doi:10.1037/0022-0663.100.2.235

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*, 621–635. doi:10.1017/S0142716403000316

Block, C. C., & Pressley, M. (2007). Best practices in teaching comprehension. In L. Gambrell, L. Mandel-Morrow, & M. Pressley (Eds.), *Best practices in literacy instruction* (pp. 220–242). New York: Guilford Press.

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Developmental Psychology, 28*, 874–886. doi:10.1037/0012-1649.28.5.874

Caravolas, M. (2005). The nature and causes of dyslexia in different languages. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 336–335). doi:10.1002/9780470757642.ch18

Craig, H., & Washington, J. (2006). *Malik goes to school*. Mahwah, NJ: Lawrence Erlbaum.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. doi:10.1037/0012-1649.33.6.934

Cunningham, A. E., Stanovich, K. E., & Wilson, M. (1990). Cognitive variation in adult college students differing in academic ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 129–159). New York: Academic Press.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91*(3), 450–476. doi:10.1037/0022-0663.91.3.450

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95*(1), 22–40. doi:10.1037/0022-0663.95.1.22

Duncan, L. G., Colé, P., Seymour, P. H. K., & Magnan, A. (2006). Differing sequences of metaphonological development in French and English. *Journal of Child Language, 33*(2), 369–99. doi:10.1017/S030500090600732X

Duncan, L. G., & Seymour, P. H. K. (2000). Socio-economic differences in foundation-level literacy. *British Journal of Psychology, 91*, 145–166. doi:10.1348/000712600161736

Ehri, L. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188.

Elley, W. (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Pergamon.

Fleischman, H., Hopstock, P., Pelczar, M., & Shelley, B. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context* (NECS 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.

Frost, R. (2005). Orthographic systems and skilled word recognition processes in reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 272–295). doi:10.1002/9780470757642.ch15

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104–115. doi:10.1037/0096-1523.13.1.104

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition, 12*, 213–237. doi:10.1017/S1366728909004015

Georgiou, G., Parrila, R., & Papadopoulos, T. (2008). Predictors of word decoding and reading fluency in English and Greek: A cross-linguistic comparison. *Journal of Educational Psychology, 100*, 566–580. doi:10.1037/0022-0663.100.3.566

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674–691. doi:10.1037/0096-1523.5.4.674

Goswami, U. (1988). Orthographic analogies and reading development. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40A*(2), 239–268. doi:10.1080/02724988843000113

Goswami, U., Porpodas, C. D., & Wheelwright, S. (1997). Children's orthographic representations in English and Greek. *European Journal of Psychology of Education, 12*, 273–292.

Gough, P. B., Juel, C., & Griffith, P. L. (1992). Reading, spelling, and the orthographic cipher. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 35–48). Hillsdale, NJ: Erlbaum.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3), 361–396. doi:10.3102/00346543066003361

Hanley, R., Masterson, J., Spencer, L., & Evans, D. (2004). How long do the advantages of learning to read a transparent orthography last? An investigation of the reading skills and reading impairment of Welsh children at 10 years of age. *The Quarterly Journal of Experimental Psychology, 57*(8), 1393–1410. doi:10.1080/02724980343000819

Heath, S. B. (1982). What no bedtime story means: Narrative skills at home and school. *Language in Society, 11*(1), 49–76.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Holopainen, L., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2000). Two alternative ways to model the relation between reading accuracy and phonological awareness at preschool age. *Scientific Studies of Reading, 4*, 77–100.

Hopstock, P., & Pelczar, M. (2011). *Technical report and user's guide for the Program for International Student Assessment (PISA): 2009 data files and database with U.S. specific* *variables* (NCES 2011-025). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*(3), 273–296.

Joncas, M. (2007). PIRLS 2006 sample design. In M. Martin, I. Mullis, & A. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 35–48). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education, 7*, 239–257. doi:10.1007/BF03173177

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics, 21*, 243–263.

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

Languages across Europe. (2011). BBC News Online country profile: Belgium. Retrieved from http://www.bbc.co.uk/languages/european_languages/countries/belgium.shtml

Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry, 51*(5), 612–620. doi:10.1111/j.1469-7610.2009.02185.x

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading, 32*(2), 199–214. doi:10.1111/j.1467-9817.2008.01389.x

Lyytinen, H., Aro, M., Holopainen, L., Leiwo, M., Lyytinen, P., & Tolvanen, A. (2005). Children's language development and reading acquisition in a highly transparent orthography. In M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 47–62). Mahwah, NJ: Lawrence Erlbaum.

Martin, M., Mullis, I., & Kennedy, A. (Eds.). (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Moats, L. C. (2011). *Speech to print: Language essentials for teachers*. Baltimore, MD: Brookes Publishing.

Mullis, I., Kennedy, A., Martin, M., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications, 2nd edition*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I., Martin, M., Kennedy, A., & Foy, P. (2007). *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries.* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). New York: Guilford Press.

Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304–330.

National Endowment for the Arts (NEA). (2007). *To read or not to read: A question of national consequence* (Research Report #47). Retrieved from the NEA website: http://www.nea.gov/research/toread.pdf

Niessen, M., Frith, U., Reitsma, P., & Öhngren, B. (2000). *Learning disorders as a barrier to human development 1995–1999.* Evaluation report. Technical Committee COST Social Sciences.

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

Organization for Economic Cooperation and Development (OECD). (2010). *PISA 2009 Results: What students know and can do; Student performance in reading, mathematics and science*. OECD Publishing. doi:10.1787/9789264091450-en

Pennington B. F., Filipek P. A., Lefly, D., Churchwell, J., Kennedy, D. N., Simon, J. H., … DeFries, J. C. (1999). Brain morphometry in reading-disabled twins. *Neurology, 53*, 723–9.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383. doi:10.1080/10888430701530730

Perfetti, C. A., & Bolger, D. J. (2004). The brain might read that way. *Scientific Studies of Reading, 8*, 293–304

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). doi: 10.1002/9780470757642.ch13

Porpodas, C. D. (1999). Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities, 32*, 406–416. doi:10.1177/002221949903200506

Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29–53.

Raudenbush, S., & Byrk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* New York: Sage.

Reese, L., & Gallimore, R. (2000). Immigrant Latinos' cultural model of literacy development: An alternative perspective on home-school discontinuities. *American Journal of Education, 108*(2), 103–13.

Sampson, G. (1985). *Writing systems.* Stanford, CA: Stanford University Press.

Schane, S. A. (1968). *French phonology and morphology*. Cambridge, MA: MIT Press.

Seidenberg, M., & McClelland, J. (1989). More words but still no lexicon: Reply to Besner et al. *Psychological Review, 97*, 477–452.

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

Seymour, P. H. K., & Evans, H. M. (1999). Foundation-level dyslexias: Assessment and treatment. *Journal of Learning Disabilities, 32*, 394–405.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218.

Share, D. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584–615. doi:10.1037/0033-2909.134.4.584

Snow, C. E. (2001). *Reading for understanding.* Santa Monica, CA: RAND Education and the Science and Technology Policy Institute.

Snowling, M. J. (2001). From language to reading and dyslexia. *Dyslexia, 7*(1), 37–46.

Spencer, L. H., & Hanley, R. J. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology, 94*, 1–28. doi:10.1348/000712603762842075

Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., & Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading and in spelling: A four year longitudinal study. *Journal of Experimental Child Psychology, 84*, 194–217. doi:10.1016/S0022-0965(03)00024-9

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32–71.

Statistics Canada. (2008). *Canadian demographics at a glance* (Catalogue No. 91-003-X). Retrieved from http://www.statcan.gc.ca/pub/91-003-x/91-003-x2007001-eng.pdf

Thorstad, G. (1991). The effect of orthography and the acquisition of literacy skills. *British Journal of Psychology, 82*, 527–537. doi:10.1111/j.2044-8295.1991.tb02418.x

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K., Conway, T., & Rose, E. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33–58. doi:10.1177/002221940103400104

Treiman, R., & Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*, 741–749. doi:10.1037/0096-3445.124.2.107

Trong, K. L., & Kennedy, A. M. (2007). Reporting PIRLS 2006 Questionnaire Data. In M. Martin, I. Mullis, & A. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

UNESCO. (2006). *International Standard Classification of Education, ISCED 1997.* Paris: UNESCO Institute of Statistics.

van Daal, V. H. P., & van der Leij, A. (1999). Developmental dyslexia: Related to specific or general deficits? *Annals of Dyslexia, 49*, 71–103. doi:10.1007/s11881-999-0020-2

Van Orden, G., & Kloos, K. (2005). The question of phonology and reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 61–78). doi:10.1002/9780470757642.ch4

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321–363.

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 Psychiatry, 45*(1), 2–40.

Venezky, R. (1999). *The American way of spelling*. New York: Guilford.

Wiley, T. G., & Lukes, M. (1996). English-only and standard English ideologies in the U.S. *TESOL Quarterly, 30*(3), 511–535. doi:10.2307/3587696

Wiley, T. G., & Wright, W. E. (2004). Against the undertow: Language-minority education policy and politics in the "age of accountability." *Educational Policy, 18*(1), 142–168. doi:10.1177/0895904803260030.

Wimmer, H., Landerl, K., & Frith, U. (1999). *Learning to read German: Normal and impaired acquisition.* New York, NY: Cambridge University Press.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*, 114–128. doi:10.1016/j.stueduc.2005.05.005

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. doi:10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*, 379–384. doi:10.1111/1467-9280.00370

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation for *_ough* and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers, 29*(4), 600–618.

Öney, B., & Goldman, S. (1984). Decoding and comprehension skills in Turkish and English: Effects of the regularity of grapheme-phoneme correspondences. *Journal of Educational Psychology, 76*, 557–567. doi:10.1037/0022-0663.76.4.557

Table 1

European Commission COST Action A8 Hypothetical Classification of Languages

	Orthographic depth				
Syllabic complexity	Shallow	→	→ →	→ →	Deep
Simple	Finnish	Greek	Portuguese	French	
		Italian			
		Spanish			
Complex		German	Dutch	Danish	English
		Norwegian	Swedish		
		Icelandic			

Table 2

Hypotheses Based on the European Commission COST Action A8 Hypothetical Classification of Languages

	Reading difficulty							
Reading difficulty	Shallow	→	→	→	→	→	→	Deep
Simple	Finnish (Finland)	Italian (Switzerland)				French (Canada/ Belgium/ Switzerland)		
↓								
↓		German (Switzerland)		Dutch (Belgium)				English (Canada)
↓								
Complex				Swedish (Finland)				

Table 3

Sample and Population Characteristics by Country: Language, Source of Data, Age/Grade of Students, Student- and School-Level Sample Sizes

Country	Language (% of population)	Dataset	Age or grade	No. of students	No. of schools
Belgium	Dutch (60%) French (33%)	PIRLS	4th grade	6,180	387
		PISA	15-year-olds	6,642	271
Canada	English (58%) French (22%)	PIRLS	4th grade	16,176	1,120
		PISA	15-year-olds	21,531	972
Finland	Finnish (93%) Swedish (6%)	PISA	15-year-olds	5,615	203
Switzerland	German (63%) French (20%) Italian (8%)	PISA	15-year-olds	10,359	426

Table 4

Cluster Size by Country

	No. of Schools	Minimum	Average	Maximum
PIRLS Belgium	387	1	16	29
PIRLS Canada	1,120	1	14.4	42
PISA Belgium	271	1	24.5	35
PISA Canada	972	1	22.2	215
PISA Finland	203	1	27.7	35
PISA Switzerland	426	1	24.3	155

Table 5

PIRLS Questions and Responses

Variable	Source	Question	Response
Sex	Student questionnaire	Are you a boy or a girl?	Girl; boy
Socioeconomic status	Home survey	What is the highest level of education completed by the child's father (or stepfather or male guardian) and mother (or stepmother or female guardian)?	Some ISCED level 1 or 2 or did not go to school; ISCED level 2; ISCED level 3 or 4; ISCED level 5A, 1st degree; ISCED Level 5B; beyond ISCED level 5A
Home language	Home questionnaire	In what language do most of the activities in Question 8 take place?	Language of test; another language
Age	Student questionnaire	When were you born?	Fill the circles next to the month and year you were born.
Instructional resources	School questionnaire	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?	Qualified teaching staff; teachers with specialization in reading; second language teachers; instructional materials (e.g. textbooks); supplies (e.g., papers, pencils); school buildings and grounds; heating/cooling and lighting systems; instructional space (e.g., classrooms; special equipment for physically disabled students; computers for instructional purposes; computer software for instructional purposes; computer support staff; library books; audio-visual resources

Note: The instructional resources index was created by taking the average of each principal's responses, which were reported on a 4-point scale (1 = a lot; 2 = some; 3 = a little; and 4 = not at all) to the school resource questions.

Table 6

PISA Questions and Responses

Variable	Source	Question	Response
Sex	Student questionnaire	Are you female or male?	Female; male
Socioeconomic status	Student questionnaire	What is the <highest level of schooling> completed by your mother?	ISCED level 3A; ISCED level 3B, 3C; ISCED level 2; ISCED level 1; did not complete ISCED level 1
Home language*	Student questionnaire	What language do you speak at home most of the time?	<Language 1>; <Language 2>; <Language 3>; < ... etc. >; other language
Instructional resources	School questionnaire	Is your school's capacity to provide instruction hindered by a shortage or inadequacy of instructional materials?	Not at all; a little; to some extent; a lot

*Country specific.

Table 7

Coding Scheme for PISA and PIRLS Variables

Variable	Name	Coded response	
		PIRLS	PISA
Comprehension ability	comprehend	Plausible value for student's PIRLS fourth grade reading score on the Interpret and Integrate Ideas and Information subsection	Plausible value for student's Reflect and Evaluate subsection of PISA
Poor reader	poor	1 if between the 10th and 25th percentiles within language subgroup on PIRLS Focus on and Retrieve Explicitly Stated Information subsection; 0 if otherwise	1 if between the 10th and 25th within language subgroup percentiles on Access and Retrieve subsection of PISA; 0 if otherwise
Very poor reader	verypoor	1 if below the 10th percentile within language subgroup on PIRLS Focus on and Retrieve Explicitly Stated Information subsection; 0 if otherwise	1 if below the 10th percentile within language subgroup on Access and Retrieve subsection of PISA; 0 if otherwise
French/English (Canada)	french	1 if French; 0 if English	1 if French; 0 if English
Dutch/French (Belgium)	dutch	1 if Dutch; 0 if French	1 if Dutch; 0 if French
Finnish/Swedish (Finland)	finnish	-	1 if Finnish; 0 if Swedish
French/ German/ Italian (Switzerland)	italian german	-	Dummy variables for language (French is the reference group)
Sex	female	1 if female; 0 if male	1 if female; 0 if male
Socioeconomic status	SES	Dummy variables for the following categories of mother's education: some ISCED level 1 or 2 or did not go to school (reference group); ISCED level 2; ISCED level 3 or 4; ISCED level 5A; ISCED level 5B, 1st degree; beyond ISCED level 5A	Dummy variables for the following categories of mother's education: ISCED level 3A; ISCED level 3B, 3C; ISCED level 2; ISCED level 1; did not complete ISCED level 1 (reference group)
Home language	homelang	1 if not test language; 0 if test language	1 if not test language; 0 if test language
Age	age	Student's age	-
Shortage of instructional resources	resource	Dummy variables for the following categories: a lot (reference group); some; a little; not at all	Dummy variables for the following categories: a lot (reference group); to some extent; very little; not at all
School identifier	schoolid	Each school's unique numerical ID	Each school's unique numerical ID

Table 8

Descriptive Statistics for the PIRLS Variables

Variable	Belgium		Canada	
	French	Dutch	French	English
Total n	2,212 (36%)	3968 (64%)	3,333 (21%)	12,843 (79%)
Reading comprehension	(*M* = 507.48, *SD* = 65.99)	(*M* = 550.99, *SD* = 52.28)	(*M* = 525.46, *SD* = 62.76)	(*M* = 562.72, *SD* = 64.91)
Age	(*M* = 9.87, *SD* = .01)	(*M* = 9.96, *SD* =.01)	(*M* = 9.96, *SD* = .38)	(*M* = 9.89, *SD* = .34)
Poor reader	315 (14%)	577 (15%)	462 (14%)	1,798 (14%)
Very poor reader	178 (8%)	342 (9%)	304 (9%)	1,113 (9%)
Male	1,119 (51%)	2,032 (51%)	1,647 (49%)	6,328 (49%)
Female	1,093 (49%)	1,936 (49%)	1,686 (51%)	6,515 (51%)
SES: Some Level 1 or 2 or did not go to school	148 (7%)	135 (3%)	105 (3%)	400 (3%)
SES: Level 2	399 (18%)	523 (13%)	154 (5%)	657 (5%)
SES: Level 3	535 (24%)	1,310 (33%)	531 (16%)	3,242 (25%)
SES: Level 4	184 (8%)	0 (0%)	1,330 (40%)	4,558 (35%)
SES: Level 5B	814 (37%)	1,166 (29%)	720 (22%)	2,375 (18%)
SES: Level 5A, 1st degree	0 (0%)	834 (21%)	477 (14%)	1,304 (10%)
SES: Beyond Level 5A, 1st degree*	132 (6%)	0 (0%)	16 (0%)	307 (2%)
Home language is not the test language	124 (6%)	211 (5%)	674 (20%)	1,180 (9%)
Home language is the test language	2,088 (94%)	3,757 (95%)	2,659 (80%)	11,663 (91%)

*n/a in Canada

Table 9

Descriptive Statistics for the Dutch and Canadian PISA Variables

Variable	Belgium		Canada	
	Dutch	French	French	English
Total n	3,874	2,768	5,169	16,362
Reading comprehension	(*M* = 532.72, *SD* = 93.50)	(*M* = 505.34, *SD* =109.66)	(*M* = 511.31, *SD* = 85.72)	(*M* = 532.58, *SD* = 92.18)
Poor reader	515 (13%)	372 (13%)	718 (14%)	2,367 (14%)
Very poor reader	292 (8%)	196 (7%)	423 (8%)	1,416 (9%)
Male	1,951 (50%)	1,407 (51%)	2,419 (47%)	8,076 (49%)
Female	1,923 (50%)	1,361 (49%)	2,750 (53%)	8,286 (51%)
SES: Did not complete Level 1	48 (1%)	58 (2%)	23 (0%)	95 (1%)
SES: Level 1	91 (2%)	108 (4%)	64 (1%)	136 (1%)
SES: Level 2	220 (6%)	236 (9%)	355 (7%)	979 (6%)
SES: Level 3B or 3C	650 (17%)	426 (15%)	-	-
SES: Level 5A, 1st degree	2,865 (74%)	1,940 (70%)	4,727 (91%)	15,152 (93%)
Home language is not the test language	942 (24%)	473 (17%)	1,540 (30%)	1,664 (10%)
Home language is the test language	2,932 (76%)	2,295 (83%)	3,629 (70%)	14,698 (90%)

Table 10

Descriptive Statistics for the Finnish and Swiss PISA Variables

Variable	Finland		Switzerland		
	Finnish	Swedish	Italian	German	French
Total n	4,278	1,337	442	6,031	3,886
Reading comprehension	(M = 539.76, SD = 85.73)	(M = 515.15, SD = 86.46)	(M = 462.12, SD = 78.26)	(M = 503.19, SD = 90.82)	(M = 500.42, SD = 87.92)
Poor reader	629 (15%)	199 (15%)	72 (16%)	785 (13%)	550 (14%)
Very poor reader	411 (10%)	121 (9%)	36 (8%)	469 (8%)	338 (9%)
Male	2,102 (49%)	624 (47%)	251 (57%)	3,028 (50%)	1,949 (50%)
Female	2,176 (51%)	713 (53%)	191 (43%)	3,003 (50%)	1,937 (50%)
SES: Did not complete Level 1	22 (1%)	1 (0%)	6 (1%)	68 (1%)	107 (3%)
SES: Level 1	183 (4%)	50 (4%)	10 (2%)	125 (2%)	160 (4%)
SES: Level 2	259 (6%)	75 (6%)	109 (25%)	1,462 (24%)	763 (20%)
SES: Level 3B or 3C	1,884 (44%)	635 (47%)	134 (30%)	3,358 (56%)	1,787 (46%)
SES: Level 5A, 1st degree	1,930 (45%)	576 (43%)	183 (41%)	1,018 (17%)	1,069 (28%)
Home language is not the test language	108 (3%)	288 (22%)	140 (32%)	894 (15%)	652 (17%)
Home language is the test language	4,170 (97%)	1,049 (78%)	302 (68%)	5,137 (85%)	3,234 (83%)

Table 11

Percentages of Missing Data and Mean Differences in Reading Comprehension Scores Between Students with Missing and Complete Data by Language Subgroup

Country	Dataset	Language	Percentage of Missing Data	Mean Comprehension Difference	Standard Deviation Difference
Belgium	PIRLS	Dutch	11%	28.78***	.50
		French	11%	31.89***	.51
	PISA	Dutch	16%	77.95***	.83
		French	12%	124.10***	1.3
Canada	PIRLS	English	22%	23.56***	.36
		French	19%	22.90***	.36
	PISA	English	7%	71.65***	.77
		French	8%	71.11***	.82
Finland	PISA	Finnish	3%	64.77***	.75
		Swedish	5%	59.05***	.68
Switzerland	PISA	German	15%	68.50***	.75
		French	9%	65.88***	.74
		Italian	9%	38.34**	.48

Note: 1 = sample with complete data, 0 = sample with missing data.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 12

Fixed Effects Estimates and Variance-Covariance Estimates for Models of the Predictors of Fourth-Grade Reading Comprehension (Interpret and Integrate Ideas and Information) on the PIRLS 2006 Assessment in Canada and Belgium

	Belgium		Canada	
Parameter	Model 1	Model 2	Model 1	Model 2
		Fixed effects		
Shallow orthography	42.81*** (1.58)	36.77*** (1.67)	-37.79*** (1.37)	-38.93*** (1.46)
Poor reader	-64.97*** (1.50)	-78.11*** (2.50)	-76.25*** (1.03)	-76.39*** (1.15)
Very poor reader	-104.15*** (1.94)	-124.80*** (3.24)	-126.95*** (1.29)	-129.23*** (1.44)
Female	2.30* (1.02)	2.45* (1.01)	8.04*** (0.70)	8.05*** (0.70)
SES: Level 2	6.09* (2.74)	5 (2.73)	10.38*** (2.52)	10.39*** (2.52)
SES: Level 3	10.41*** (2.64)	9.95*** (2.62)	9.22*** (2.11)	9.17*** (2.11)
SES: Level 4	2.73 (3.87)	3.44 (3.84)	13.93*** (2.08)	13.91*** (2.08)
SES: Level 5B	22.11*** (2.67)	21.57*** (2.65)	21.50*** (2.16)	21.55*** (2.16)
SES: Level 5A, 1st degree	30.72*** (2.94)	31.42*** (2.92)	27.83*** (2.28)	27.86*** (2.28)
SES: Beyond level 5A, 1st degree	37.00*** (4.35)	34.16*** (4.34)	5.44 (3.18)	5.66 (3.18)
Home language is not test language	-3.26 (2.33)	-3.44 (2.31)	-4.40*** (1.22)	-4.51*** (1.22)
Age	-4.13*** (1.16)	-4.27*** (1.15)	0.15 (1.04)	0.17 (1.04)
Poor reader × shallow orthography		20.00*** (3.09)		0.87 (2.55)
Very poor reader × shallow orthography		31.08*** (3.95)		11.07*** (3.15)
Intercept	549.76*** (11.96)	555.42*** (11.89)	564.15*** (10.55)	564.19*** (10.54)
		Variances		
Level 2				
ψ	2.46*** (0.07)	2.40*** (0.07)	2.62*** (0.04)	2.62*** (0.04)
Θ	3.67*** (0.01)	3.66*** (0.01)	3.77*** (0.01)	3.77*** (0.01)
n	6,180	6,180	16,176	16,176

Note: Standard errors in parentheses. Shallow orthographies are French in Canada and Dutch in Belgium. SES = socioeconomic status. ψ = between school variance and Θ = within school variance. Reference groups: French (Belgium) and English (Canada), competent/skilled readers, males, SES: some ISCED level 1 or 2 or did not go to school, home language is the test language.
**p* < .05. ***p* < .01. ****p* < .001.

Table 13

Fixed Effects Estimates and Variance-Covariance Estimates for Models of the Predictors of Fourth-Grade Reading Comprehension (Reflect and Evaluate) on the PISA 2009 Assessment in Canada and Belgium

	Belgium		Canada	
Parameter	Model 1	Model 2	Model 1	Model 2
		Fixed effects		
Shallow orthography	34.17*** (6.27)	31.04*** (6.32)	-18.79*** (2.31)	-21.34*** (2.36)
Poor reader	-93.65*** (2.28)	-101.57*** (3.58)	-105.40*** (1.22)	-107.46*** (1.38)
Very poor reader	-152.18*** (3.23)	-158.11*** (5.10)	-172.08*** (1.57)	-176.03*** (1.78)
Female	7.78*** (1.53)	7.88*** (1.53)	16.28*** (0.85)	16.33*** (0.85)
SES: Level 5A, 1st degree	36.61*** (5.98)	36.73*** (5.98)	17.17** (5.64)	17.39** (5.64)
SES: Level 3B or 3C	19.29** (6.12)	19.35** (6.12)	-4.79 (5.85)	-4.58 (5.85)
SES: Level 2	19.33** (6.42)	19.38** (6.42)	0.44 (7.03)	0.44 (7.03)
SES: Level 1	24.12*** (7.05)	24.29*** (7.05)	-	-
Home language is not the test language	-9.61*** (1.87)	-9.25*** (1.87)	-10.72*** (1.37)	-10.77*** (1.37)
Poor reader × shallow orthography		13.08** (4.63)		8.95** (2.85)
Very poor reader × shallow orthography		9.36 (6.54)		17.11*** (3.65)
Intercept	481.69*** (7.62)	483.44*** (7.63)	538.39*** (5.77)	539.58*** (7.99)
		Variances		
Level 2				
ψ	3.89*** (0.05)	3.89*** (0.05)	3.30*** (0.03)	3.30*** (0.03)
Θ	4.03*** (0.01)	4.03*** (0.01)	4.08*** (0.01)	4.08*** (0.01)
n	6,642	6,642	21,531	21,531

Note: Standard errors in parentheses. Shallow orthography is French in Canada and Dutch in Belgium. SES = socioeconomic status. ψ = between school variance and Θ = within school variance. Reference groups: French (Belgium) and English (Canada), competent/skilled readers, males, SES: did not complete Level 1, home language is the test language.
**p* < .05. ***p* < .01. ****p* < .001.

Table 14

Fixed Effects Estimates and Variance-Covariance Estimates for Models of the Predictors of Fourth-Grade Reading Comprehension (Reflect and Evaluate) on the PISA 2009 Assessment in Finland and Switzerland

	Finland		Switzerland	
Parameter	Model 1	Model 2	Model 1	Model 2
		Fixed effects		
Shallow orthography	25.31*** (3.46)	25.83*** (3.61)	-4.3 (3.53)	-4.85 (3.60)
Italian			-40.47*** (6.08)	-45.60*** (6.38)
Poor reader	-92.05*** (2.25)	-87.74*** (4.52)	-99.12*** (1.63)	-100.78*** (2.55)
Very poor reader	-159.18*** (2.77)	-162.09*** (5.65)	-163.68*** (2.08)	-166.54*** (3.19)
Female	32.74*** (1.59)	32.74*** (1.59)	21.96*** (1.08)	21.96*** (1.08)
SES: Level 5A or 1st degree	55.06*** (12.22)	55.04*** (12.23)	25.71*** (4.27)	25.41*** (4.27)
SES: Level 3B or 3C	44.01*** (12.22)	44.01*** (12.22)	17.07*** (4.20)	16.76*** (4.20)
SES: Level 2	27.71* (12.55)	27.65* (12.55)	8.06 (4.22)	7.66 (4.22)
SES: Level 1	23.15 (12.73)	23.1 (12.73)	0.49 (5.09)	0.1 (5.09)
Home language is not the test language	-15.54*** (3.40)	-15.64*** (3.40)	-15.37*** (1.57)	-15.25*** (1.57)
Poor reader × shallow orthography		-5.67 (5.16)		2.18 (3.34)
Very poor reader × shallow orthography		3.79 (6.40)		2.57 (4.21)
Poor reader × Italian				9.89 (7.81)
Very poor reader × Italian				33.86** (10.29)
Intercept	480.25*** (12.70)	479.88*** (12.71)	503.15*** (5.01)	503.96*** (5.04)
		Variances		
Level 2				
ψ	2.87*** (0.07)	2.87*** (0.07)	3.43*** (0.04)	3.43*** (0.04)
Θ	4.04*** (0.01)	4.04*** (0.01)	3.97*** (0.01)	3.97*** (0.01)
n	5,615	5,615	10,359	10,359

Note: Standard errors in parentheses. Shallow orthography is Finnish in Finland and German in Switzerland. SES = socioeconomic status. ψ = between school variance and Θ = within school variance. Reference groups: Swedish (Finland) and French (Switzerland), competent/skilled readers, males, SES: did not complete Level 1, home language is the test language.
**p* < .05. ***p* < .01. ****p* < .001.

Table 15

Effect Sizes: Standard Deviation Differences

Dataset	Shallow orthography	Deep orthography	Majority language(s)	Minority language	Main effects of orthography	Poor × shallow interaction	Very poor × shallow interaction
Belgium (PIRLS)	Dutch	French	Dutch	French	.67***	.32***	.49***
Canada (PIRLS)	French	English	English	French	-.56***	.01	.17***
Belgium (PISA)	Dutch	French	Dutch	French	.29***	.12***	.09
Canada (PISA)	French	English	English	French	-.20***	.10**	.18***
Finland (PISA)	Finnish	Swedish	Finnish	Swedish	.29***	-.06	-.04
Switzerland (PISA)	German	French	German/ French	Italian	-.05	.02	.03
	Italian	French	German/ French	Italian	-.44***	.11	.37***
	Italian	German	German/ French	Italian	-.39***	.08	.34**

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. And the orthography variables are coded: 1 = shallow orthography and 0 = deep orthography.

Table 16

Variability in Comprehension Scores (Standard Deviations) and Homogeneity of Variance Test, Results by Country

Country	Dataset	Shallow orthography	Deep orthography
Belgium	PIRLS	52.28	65.99***
	PISA	93.50	109.66***
Canada	PIRLS	62.76	64.91*
	PISA	85.72	92.18***
Finland	PISA	85.73	86.47
Switzerland	PISA	78.26 (Italian)	90.82*** (German)
		90.82* (German)	87.92 (French)
		78.26 (Italian)	87.92*** (French)

*$p < .05$. **$p < .01$. ***$p < .001$.

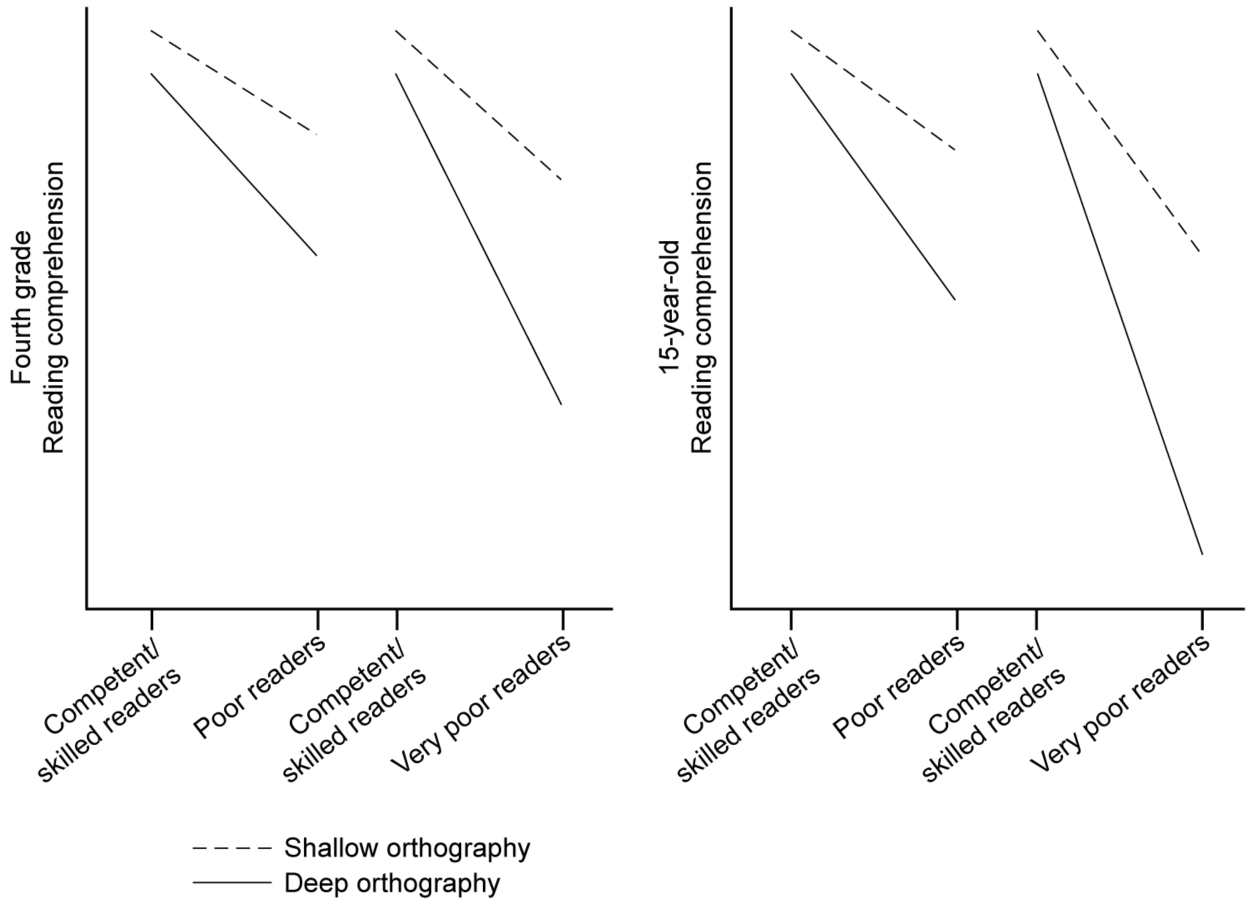

Figure 1. Graphic display of the hypotheses: (1) There are long-term advantages of reading in shallow orthographies compared to deep orthographies; (2) There are additional long-term positive effects of reading in shallow orthographies compared to deep orthographies for the lowest performing 10 to 25% of readers; (3) Orthographic depth has the strongest influence on the comprehension skills of the poorest readers; (4) Reading difficulties related to orthographic depth are exacerbated over time; (5) There is less variance in reading comprehension performance in shallow orthographies than deep orthographies.

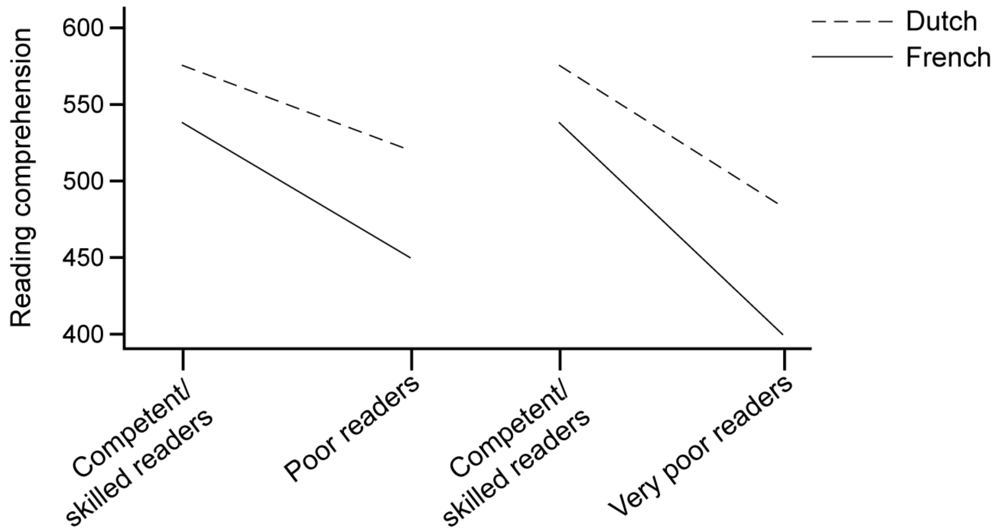

Figure 2. Interactions between orthographic depth and fourth grade reading ability in Belgium (PIRLS).

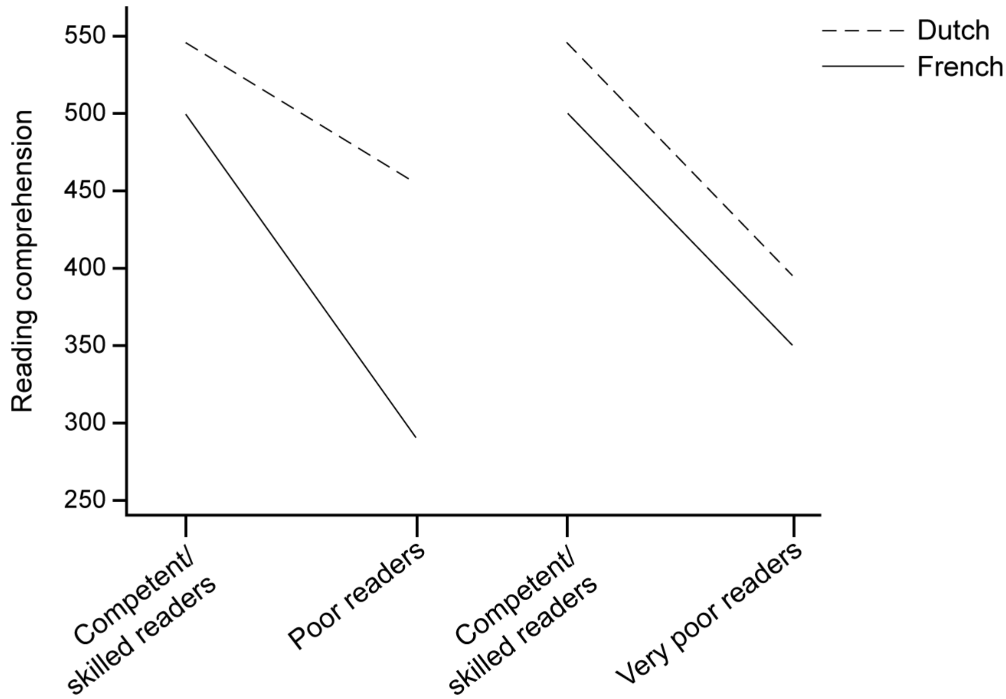

Figure 3. Interactions between orthographic depth and 15-year-old reading ability in Belgium (PISA).

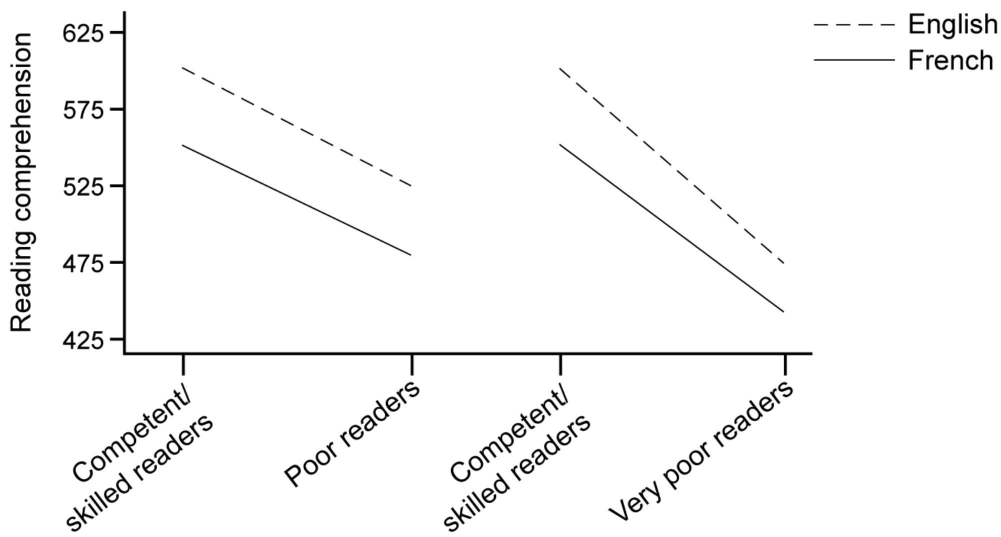

Figure 4. Interactions between orthographic depth and fourth grade reading ability in Canada (PIRLS).

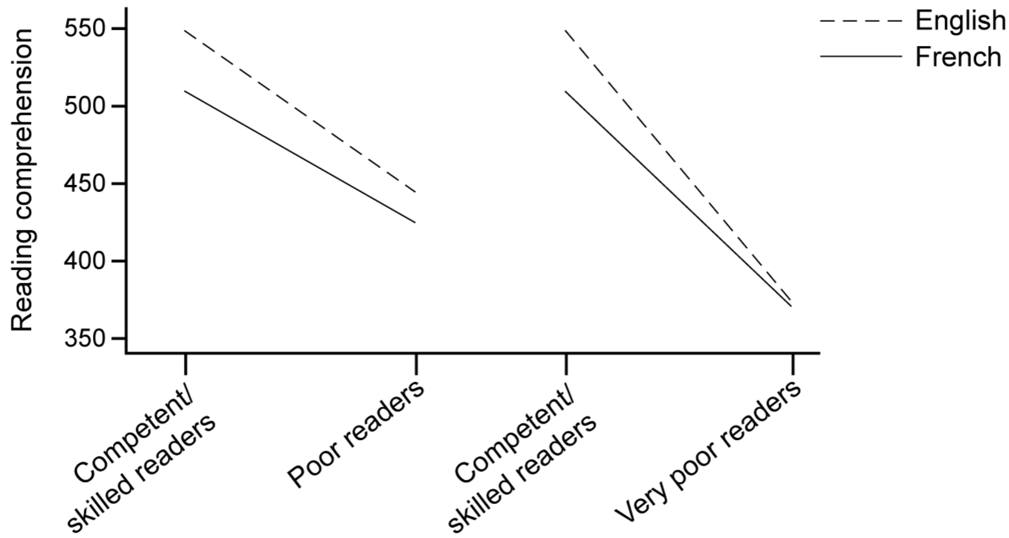

Figure 5. Interactions between orthographic depth and 15-year-old reading ability in Canada (PISA).

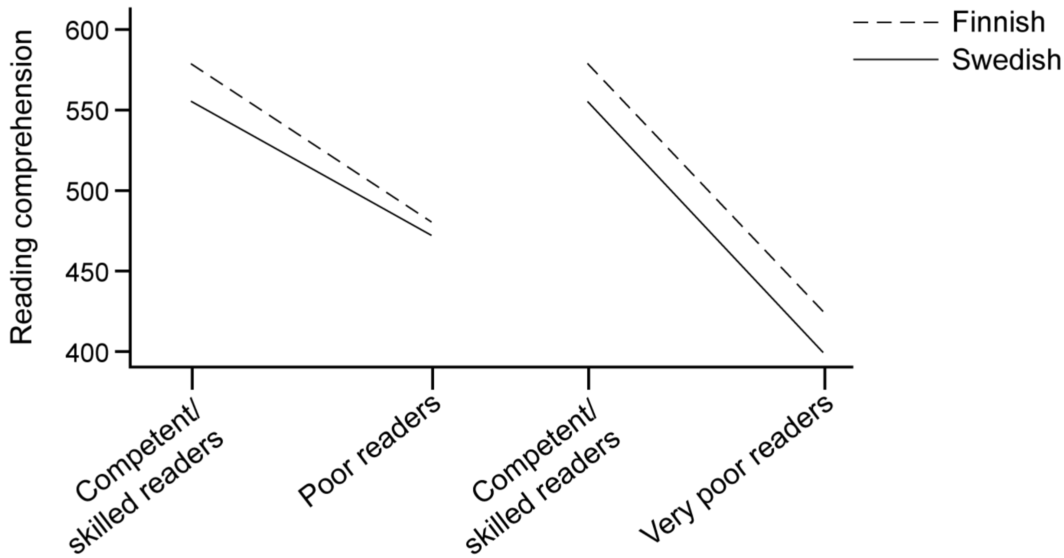

Figure 6. Predictors of 15-year-old reading comprehension in Finland (PISA).

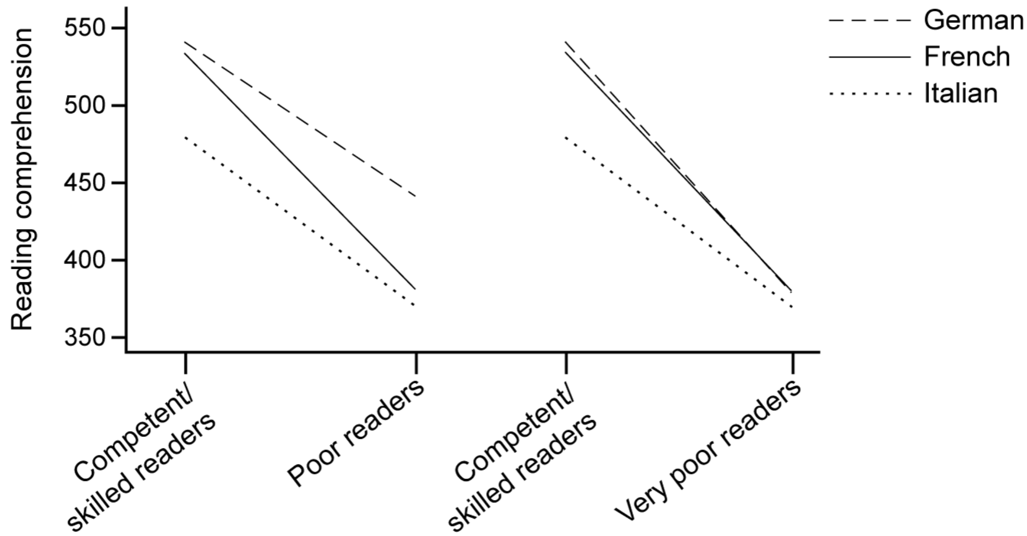

Figure 7. Interactions between orthographic depth and 15-year-old reading ability in Switzerland (PISA).

Appendix A

Equations for Multilevel Models

Hypothesis 1: There are long-term advantages of reading in shallow orthographies compared to deep orthographies, when controlling for reading ability, gender, SES, language spoken at home, and age (PIRLS only).

Belgium and Canada (PIRLS):

$$Y_{ij} = \beta_o + \beta_1 shalloworthography_{ij} + \beta_2 poor_{ij} + \beta_3 verypoor_{ij} + \beta_4 female_{ij} + \boldsymbol{\beta_5 SES_{ij}} + \beta_6 homelang_{ij} + \beta_7 age_{ij} + \zeta_j + \varepsilon_{ij}$$

Belgium, Canada, and Finland (PISA):

$$Y_{ij} = \beta_o + \beta_1 shalloworthography_{ij} + \beta_2 poor_{ij} + \beta_3 verypoor_{ij} + \beta_4 female_{ij} + \boldsymbol{\beta_5 SES_{ij}} + \beta_6 homelang_{ij} + \zeta_j + \varepsilon_{ij}$$

Switzerland (PISA):

$$Y_{ij} = \beta_o + \beta_1 italian_{ij} + \beta_2 german_{ij} + \beta_3 poor_{ij} + \beta_4 verypoor_{ij} + \beta_5 female_{ij} + \boldsymbol{\beta_6 SES_{ij}} + \beta_7 homelang_{ij} + \zeta_j + \varepsilon_{ij}$$

Where Y_{ij} (for all the above models) is the reading comprehension score for the i^{th} student in the j^{th} school and $\zeta_j | x_{ij} \sim N(0, \psi)$ and $\varepsilon_{ij} | x_{ij}, \zeta_j \sim N(0, \theta)$ and $\mathbf{SES_{ij}}$ is a vector of dummy variables for SES (reference group: some ISCED level 1 or 2 or did not go to school). In the models other than Switzerland, orthography is specific to each country: Belgium (1 = Dutch, 0 = French), Canada (1 = French, 0 = English), and Finland (1 = Finnish, 0 = Swedish).

Hypotheses 2 and 3: There are additional long-term positive effects of reading in shallow orthographies compared to deep orthographies for the lowest performing 10 to 25% of readers, and orthographic depth has the strongest influence on the comprehension skills of the poorest readers, when controlling for reading ability, gender, SES, language spoken at home, and age (PIRLS only).

Hypothesis 4: Reading difficulties related to orthographic depth are exacerbated over time such that the gap between poor readers of relatively deep and shallow orthographies becomes more apparent as students develop from fourth grade to high school.

Belgium and Canada (PIRLS):

$$Y_{ij} = \beta_o + \beta_1 shalloworthography_{ij} + \beta_2 poor_{ij} + \beta_3 verypoor_{ij} + \beta_4 female_{ij} + \boldsymbol{\beta_5 SES_{ij}} + \beta_6 homelang_{ij} + \beta_7 age_{ij} + \beta_8 poor_{ij} \times orthography_{ij} + \beta_9 verypoor_{ij} \times orthography_{ij} + \zeta_j + \varepsilon_{ij}$$

Belgium, Canada, and Finland (PISA):

$$Y_{ij} = \beta_o + \beta_1 shalloworthography_{ij} + \beta_2 poor_{ij} + \beta_3 verypoor_{ij} + \beta_4 female_{ij} + \boldsymbol{\beta_5 SES_{ij}} + \beta_6 homelang_{ij} + \beta_7 poor_{ij} \times orthography_{ij} + \beta_8 verypoor_{ij} \times language_{ij} + \zeta_j + \varepsilon_{ij}$$

Switzerland (PISA):

$Y_{ij} = \beta_o + \beta_1 italian_{ij} + \beta_2 german_{ij} + \beta_3 poor_{ij} + \beta_4 verypoor_{ij} + \beta_5 female_{ij} + \boldsymbol{\beta_6 SES_{ij}} + \beta_7 homelang_{ij} + \beta_8 poor_{ij} \times italian_{ij} + \beta_9 poor_{ij} \times german_{ij} + \beta_{10} verypoor_{ij} \times italian_{ij} + \beta_{11} verypoor_{ij} \times german_{ij} + \zeta_j + \varepsilon_{ij}$

Where Y_{ij} (for all the above models) is the reading comprehension score for the i^{th} student in the j^{th} school and $\zeta_j | x_{ij} \sim N(0, \psi)$ and $\varepsilon_{ij} | x_{ij}, \zeta_j \sim N(0,\theta)$ and $\mathbf{SES_{ij}}$ is a vector of dummy variables for SES (reference group: some ISCED level 1 or 2 or did not go to school). In the models other than Switzerland, orthography is specific to each country: Belgium (1 = Dutch, 0 = French), Canada (1 = French, 0 = English), and Finland (1 = Finnish, 0 = Swedish).

If Hypothesis 1 is confirmed there will be a pattern of positive significant effects of $shalloworthography_{ij}$ (and $italian_{ij}$ and $german_{ij}$ in Switzerland) across countries. If Hypothesis 2 is confirmed, there will be a pattern of positive significant interaction effects across countries for $poor_{ij}$ by $orthography_{ij}$ and $verypoor_{ij}$ by $orthography_{ij}$ (PIRLS), $poor_{ij}$ by $orthography_{ij}$ and $verypoor_{ij}$ by $orthography_{ij}$ (PISA), and $poor_{ij}$ by $italian_{ij}$, $poor_{ij}$ by $german_{ij}$, $verypoor_{ij}$ by $italian_{ij}$, and $verypoor_{ij}$ by $german_{ij}$ in Switzerland. If Hypothesis 3 is confirmed, the $verypoor_{ij}$ by $orthography_{ij}$ interactions will be of a greater magnitude than the $poor_{ij}$ by $orthography_{ij}$ interactions across countries. If Hypothesis 4 is confirmed, effect sizes for the reading ability by orthographic depth interactions will be larger at the 15-year-old level compared to the fourth grade level. Additionally, these findings will hold regardless of whether the shallow or deep orthography reflects the dominant language spoken in each country and when controlling for the covariates.