PHILIP ANGELL

POTENTIAL RELATIONSHIPS AMONG LANGUAGE-COMPLEXITY VARIABLES, HOME-LANGUAGE VARIABLES AND RANGE OF READING ABILITY: EVIDENCE FROM PIRLS 2016 AND PISA 2018

PHILIP ANGELL

A THESIS SUBMITTED

FOR

THE DEGREE OF MASTER OF ARTS

IN

CURRICULUM AND INSTRUCTION

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

SEPTEMBER 2022

2022

To Elham,

my inspiration,

whose unwavering love and support

touches me every day

Potential Relationships Among Language-Complexity Variables, Home-Language Variables and Range of Reading Ability: Evidence from PIRLS 2016 and PISA 2018

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Philip Angell

In Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in

Curriculum and Instruction

Ankara

September 2022

İHSAN DOĞRAMACI BILKENT UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

Potential Relationships Among Language-Complexity Variables, Home-Language Variables and Range of Reading Ability: Evidence from PIRLS 2016 and PISA 2018

Philip Angell

September 2022

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

---------- ----------

Asst. Prof. Dr Ilker Kalender (Advisor) Asst Prof Dr Necmi Akşit (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

Asst. Prof. Dr. Tijen Akşit (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Curriculum & Instruction.

Prof. Dr. Ismail Hakkı Mirici, Hacettepe University (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Orhan Arıkan (Director)

ABSTRACT

POTENTIAL RELATIONSHIPS AMONG LANGUAGE-COMPLEXITY VARIABLES, HOME-LANGUAGE VARIABLES AND RANGE OF READING ABILITY: EVIDENCE FROM PIRLS 2016 AND PISA 2018

Philip Angell

MA in Curriculum and Instruction

Advisor: Asst. Prof. Dr. İlker Kalender, 2nd Advisor: Asst. Prof. Dr. Necmi Aksit

September 2022

Reading is one of the most important skills for children to master during their time in school. It is strongly connected to life outcomes, and as such, education ministries place it at the centres of their education policies. English is one of the most challenging alphabetic languages to learn to read, and governments of anglophone countries have spent many years working to improve the effectiveness of their literacy education. However, when examining International Large-Scale Assessments, it is notable that although students in anglophone countries are able to achieve among the highest reading levels, their poorest readers lag much further behind than the poorest readers in similarly successful non-anglophone countries. This study made use of data from PIRLS (2016) and PISA (2018) to investigate possible relationships between range of reading ability and language complexity variables related to orthography and morphology, as well as between range of reading ability and home-language disparity in anglophone countries. Pearson correlational analyses showed that orthographic complexity and morphological complexity were moderately correlated with range of reading ability in both datasets. Orthographic transparency was found to be strongly correlated with range of reading ability in the PISA dataset and very strongly correlated in the PIRLS dataset. Morphological unpredictability was not found to be correlated with either dataset. Home-language disparity was not shown to be connected with range of reading ability in the PISA dataset, but in the PIRLS dataset, students who never spoke English at home were shown to have a wider range of reading ability than other students.

Keywords: Orthographic depth, morphological complexity, monolingualism, reading ability, reading range, PISA, PIRLS

ÖZET

DİL KARMAŞIKLIĞI İLE ANA DİL DEĞİŞKENLERİ VE OKUMA YETENEK ARALIĞI ARASINDAKİ OLASI İLİŞKİLER: PIRLS 2016 VE PISA 2018'DEN BULGULAR

Philip Angell

Eğitim Programları ve Öğretim Yüksek Lisans Programı

Danışman: Dr. Öğr. Üyesi Prof. Dr. İlker Kalender

2. Danışman Dr. Öğr. Üyesi Necmi Aksit

Eylül 2022

Okuma, çocukların okulda hayatları boyunca ustalaşacakları en önemli becerilerden biridir. Yaşam sonuçlarıyla güçlü bir şekilde bağlantılıdır ve bu nedenle eğitim bakanlıkları okuma becerisini eğitim politikalarının merkezine yerleştirmektedir. İngilizce, okumasını öğrenmesi en zorlu alfabetik dillerden biridir ve İngilizce konuşulan ülkelerin hükümetleri, okuryazarlık eğitimlerinin etkinliğini artırmak için uzun yıllardır çalışmaktadırlar. Bununla birlikte, uluslararası büyük ölçekli değerlendirme çalışmaları incelendiğinde, İngilizce konuşulan ülkelerdeki öğrencilerin, en yüksek okuma seviyelerine ulaşabilseler de; okuma becerileri en zayıf olan öğrencilerin, benzer başarı düzeyinde olup da İngilizce konuşulmayan ülkelerdeki aynı grup öğrencilerden çok daha geride kalmaları dikkat çekicidir. Bu çalışmada, İngilizce konuşulan ülkelerde okuma yeteneği aralığı ile yazım ve morfoloji ile ilgili dil karmaşıklığı değişkenler arasındaki ilişkiler yanında okuma yeteneği aralığı ve çok dillilik arasındaki olası ilişkileri araştırmak için PIRLS (2016) ve PISA (2018) verileri kullanılmıştır. İlişkisel analizler, ortografik karmaşıklık ve morfolojik karmaşıklığın, her iki veri setinde de okuma yeteneği aralığı ile orta derecede ilişkili olduğunu göstermiştir. Ortografik şeffaflığın, PISA veri setindeki okuma yeteneği aralığı ile güçlü bir şekilde ilişkili olduğu ve PIRLS veri setinde ise çok güçlü bir şekilde ilişkili olduğu bulunmuştur. Morfolojik öngörülemezliğin, her iki veri seti ile de güvenilir bir şekilde ilişkili olduğu bulunmamıştır. Çok dilliliğin PISA veri setinde okuma becerisi ile bağlantılı olduğu görülmemiş, ancak PIRLS veri setinde evde hiç İngilizce konuşmayan öğrencilerin diğer öğrencilere göre daha geniş bir okuma becerisine sahip oldukları ortaya konmuştur.

Anahtar kelimeler: Ortografik derinlik, morfolojik karmaşıklık, tek dillilik, okuma yeteneği, okuma aralığı, PISA, PIRLS

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincerest appreciation to Prof. Dr. Ali DOĞRAMACI for establishing and supporting this program.

I am thankful to my thesis advisors, Asst. Prof. Dr. İlker Kalender and Asst. Prof. Dr. Necmi Aksit, for their guidance and comments.

I would also like to express my thanks, with gratitude, to Margaret Griffiths, for her time, insight and support with certain ideas central to this work.

Finally, a very special thanks to my wife, Elham, for her unwavering support, love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

Introduction

Given standard cognitive and linguistic development, English is one of the most demanding alphabetic languages to learn to read; It takes English-speaking (anglophone) children longer to read than for children of other languages – in some cases requiring several years longer, depending on the language. International comparisons show that a variety of factors influence population patterns of reading ability, spanning both linguistic and social factors (De Witt & Lessing, 2017; Frith et al., 1998; Lukie et al., 2014). This thesis is concerned with how linguistic factors affect patterns of reading ability, especially in the context of comparing Anglophone countries with non-Anglophone countries.

Over the latter half of the twentieth century, a wide variety of stakeholders in education became increasingly interested in making international comparisons regarding national educational metrics. One of the metrics of greatest interest has consistently been reading ability as literacy has long been understood to strongly influence both people's educational outcomes and their professional lives (Gibson et al., 2019; Sum, 1999). Various International Large-Scale Assessments (ILSAs) were developed for its measurement, and the results of these international comparisons and benchmarking tests have been used extensively to guide both educational policy and also development in various sectors of the teaching profession. In the main, the majority of changes have focused on greater standardisation and improvement of the educational experience of children, either through providing better subject knowledge, support materials and training for the teachers, or by changing the

expectations and timeframes laid out in various national curricula. This has led to significant improvements in the overall educational achievement in many countries, but many anglophone countries have noticed that despite increasing their overall reading ability levels, they still have a stubbornly long tail of students who remain poor readers (Knight et al., 2019; Thomson et al., 2016).

Background

How Do Languages Vary in Complexity?

Written languages vary in complexity and ease-of-reading across many measurable axes. Some of these are linked to the underlying morphological structures of the language, such as the degree to which a language's morphology is agglutinative and allows individual units of meaning (morphemes) to be combined or not and which particular grammatical structures are allowed (one type of morphological complexity). For example, English allows various units of meaning to be combined into single words:

Re|sand|ing

Re- (prefix) – amends the meaning of the root to imply a repetition

sand- (root) – verb meaning to make something smooth by rubbing

-ing (verb suffix) – identifies the verb as being continuous

On the other hand, some languages do not combine multiple units of meaning (morphemes) into single words, and instead keep each unit of meaning separate. Examples of language that do not allow morphemes to be combined or altered include Vietnamese and Chinese languages. Agglutinative morphologies give rise to much longer words than is the case for non-agglutinative morphologies. This has a detrimental impact on the language's readability – especially during early, less sophisticated phases of reading acquisition. Early readers find short words easier to

read than longer words, although this effect is mediated by word frequency (Gerth & Festman, 2021).

There are other variables that influence a language's readability, but without being intrinsically connected to the complexity of the language itself. An example of a variable that is independent from the language's underlying complexity is the spelling system, or orthography, employed by that language. Some languages, such as English have developed over a long period of time, incorporating and adapting spelling conventions from a variety of sources (Upward & Davidson, 2011). Others, such as French, have a history of centrally-imposed spelling and vocabulary conventions that have minimised the impact of foreign "loan" words (Fitzsimmons, 2017). Still others, such as Turkish, have fundamentally reformed their spelling systems (and even the choice of alphabet) with the majority of "loan" words being re-spelled according to new spelling conventions, thereby allowing new words to be assimilated into the national lexicon whilst maintaining standardized spelling conventions (Brendemoen, 2021). As such, a language's orthography can be seen to be, to some extent, arbitrary, without any overt link to any other underlying measure of linguistic complexity. The level of difficulty of a spelling system is known variously as its orthographic depth, transparency or complexity (Frith et al., 1998; Knight et al., 2019; Schmalz et al., 2015); a measure of the regularity or complexity of a written language's spelling. The more rules a language uses to turn spellings into sounds, and the higher the proportion of words whose spellings do not conform to those rules, the more orthographically deep or complex it is said to be. The orthographic depth of any given language can be placed on a continuum, from very shallow or transparent, such as Turkish, to very deep or complex, such as English (Marjou, 2021; van den Bosch et al., 1994). Some languages, such as Turkish, have

relatively few rules to learn and also have few exception words whose spellings do not conform; some languages, such as French, have many more spelling rules while having relatively few exception words; some languages, such as English, have both a large number of spelling rules and also a large number of exception words (Schmalz et al., 2015).

The level of orthographic complexity has been shown to affect the ease with which beginner readers can become fluent and independent, and therefore how much teaching time and additional support is required to ensure all children become literate (Knight et al., 2019). At one extreme, educators in Taiwan spend only the first ten weeks of school teaching children to read a phonetic alphabet called Zhu-Yin-Fu-Hao, after which nearly all children achieve a good level of independence (Huang & Hanley, 1997). This orthography is then used to support the learning of Chinese characters during the rest of elementary education. At the other extreme, it takes anglophone children over two years to achieve a basic level of fluency, and many more to become truly independent (Seymour, 2005).

How do Orthographic Depth and Other Language Complexity Factors Affect Population-Level Patterns of Reading Ability?

Although orthographic depth is known to affect how long it takes students to learn to read, all written languages do allow normally-developing readers to eventually become fluent, sophisticated readers (Schmalz et al., 2015). That is, all readers who do not suffer from any form of developmental reading impairment are able to reach a level of reading ability that allows them to independently read and efficiently gain understanding from text, although the time it takes to attain this level of independence differs from language to language. On the other hand, it is also known that languages with deep orthographies pose significant extra problems for

students with reading disabilities (Landerl et al., 1997, 2013). In other words, while anglophone readers without reading disabilities are able to reach the highest possible levels of reading ability, people with various forms of reading difficulty can take much longer to achieve fluency, with some even finding themselves at a lifelong disadvantage. Not only that, but as severity of reading disability has been shown to exist on a continuum (Shaywitz & Shaywitz, 2005; Spanoudis et al., 2019), deep orthographies, such as English (Frith et al., 1998), might be expected to give rise to a more visible range of extra reading difficulties at a population level. This does not seem to be the case for readers of languages with orthographies generally considered to be shallow, such as Italian, as pointed out by Sprenger-Charolles et al. (2011), with adult dyslexic readers of shallow orthographies showing almost no reading impairment. Patterns of reading acquisition for individuals with reading difficulties in languages with shallow orthographies are instead characterised by slightly slower initial rates of progress, but with nearly all readers being able to become fully sophisticated readers in time (Silinskas et al., 2020; Torppa et al., 2012, 2013). As incidences of developmental reading difficulties remain fairly stable across different countries, this suggests that languages with different orthographic depths might give rise to different patterns of reading ability at population levels, and that these patterns may become more noticeable during the time given over to formal education.

While orthographic complexity has been shown to affect the speed of early reading acquisition, it is becoming apparent that morphological factors (that is, factors related to the rules governing how words are created) can also affect patterns of later reading acquisition (see Borleffs et al., 2019 for review). Children growing up speaking Turkish, Finnish or Basque as their mother tongues (all highly morphologically complex languages) have been shown to develop an ability to read

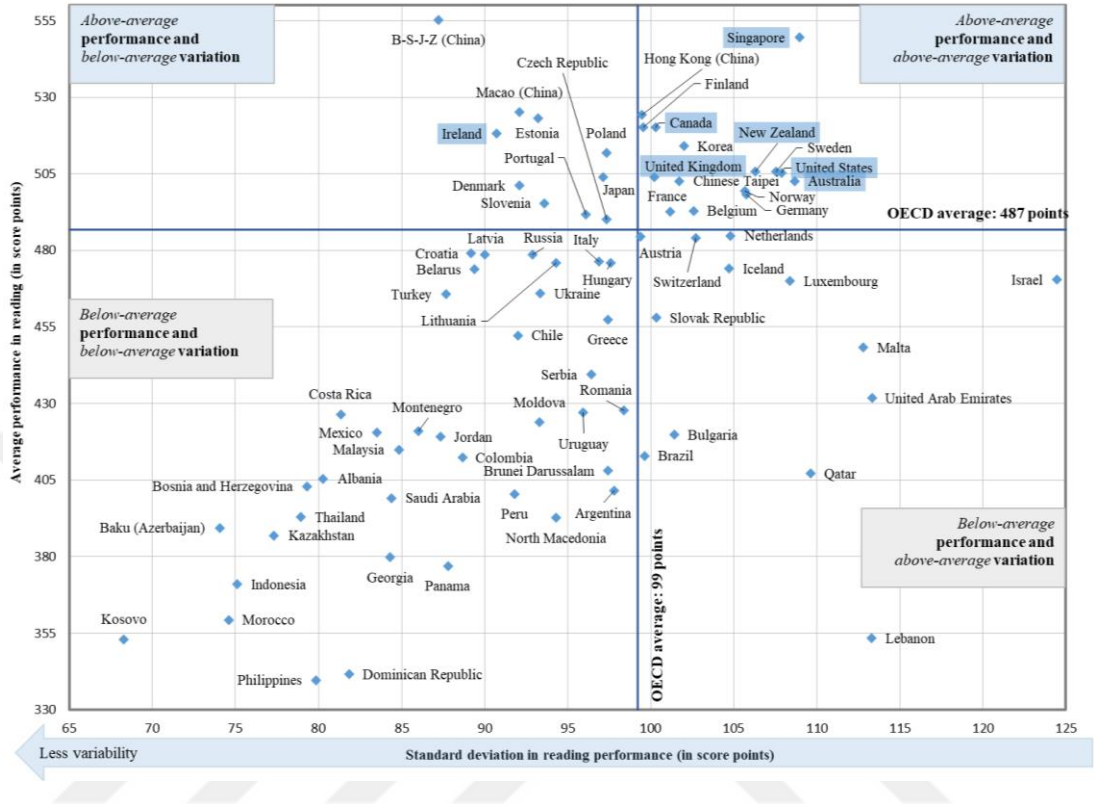
morphologically complex words more easily than morphologically simple words as they become more proficient readers.

Problem

As described above, language-complexity factors, such as orthographic depth or morphological complexity, can affect reading acquisition in ways that will likely lead to patterns of reading ability that would be visible at population levels. At the same time, data from ILSAs and benchmarking studies have been used by education ministries around the world for many years to plan improvements to the teaching of reading and to track their success (Addey & Sellar, 2018; Arikan et al., 2020). Two of the most widely-used and respected of these ILSAs are the Progress in International Reading Literacy Study (PIRLS) from the International Association for the Evaluation of Educational Achievement (IEA) which is sat by children in the fourth grade, and the Programme for International Student Assessment (PISA) from the Organisation for Economic Co-operation and Development (OECD), which is sat by 15-year-old children. Anglophone countries in particular have been interested in identifying possible areas of improvement as they have consistently shown a surprisingly large number of students performing below their peers in other countries. As a result, these anglophone countries have invested heavily over many years in attempts to close the gaps with other countries, but with seemingly paradoxical outcomes. As can be seen in Figure 1, a scatter graph showing the average reading performances and variations in PISA (2018), nearly all the countries that use English as their main language of instruction find themselves in the top right quadrant, being above average for both performance and variation.

Figure 1

Average Performance in Reading and Variation in Performance in PISA (2018)



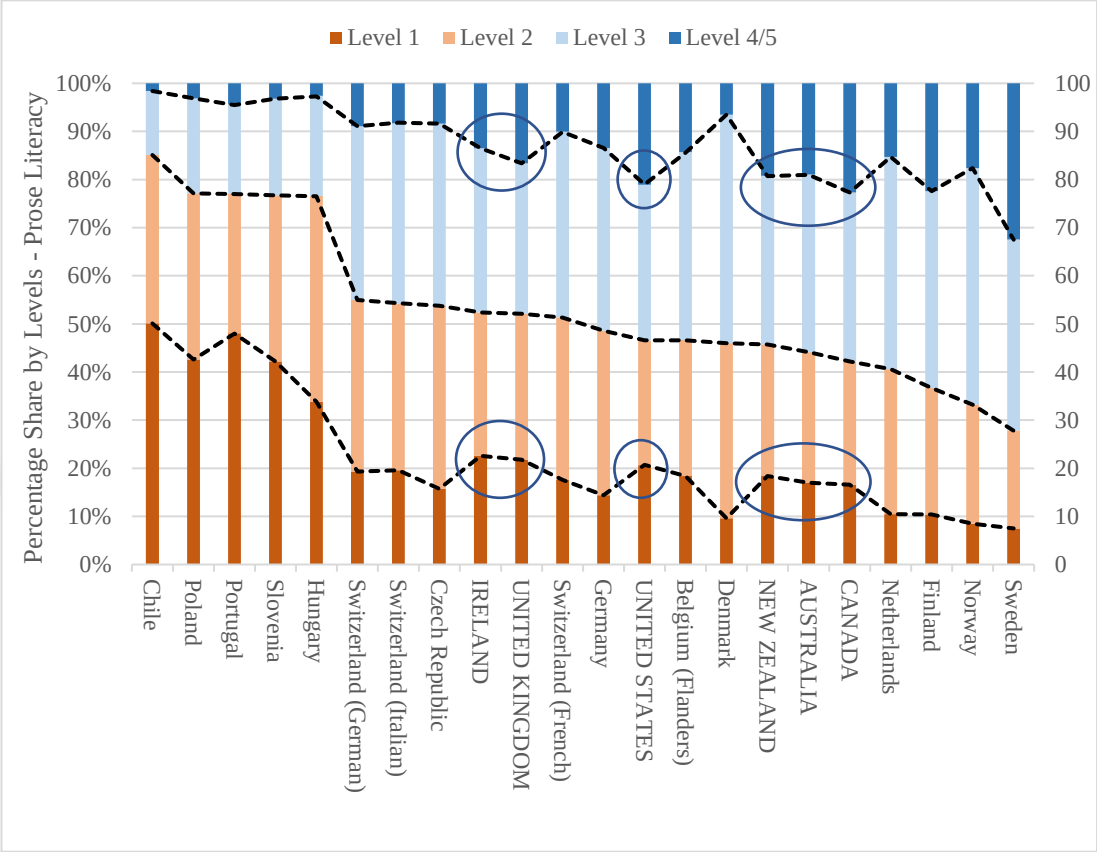
Note. Figure adapted from OECD (2019b) with anglophone nations highlighted.

https://doi.org/10.1787/888934028349

Not only that, but these differences persist beyond the end of compulsory education, leading to elevated levels of adults who lack what The United Nations Educational, Scientific and Cultural Organization (UNESCO) refers to as “functional literacy” (Burnett, 2005). As can be seen in Figure 2, anglophone nations might have impressive numbers of highly literate adults, but they also seem to have higher proportions of adults with very low levels of literacy compared to their non-anglophone ranked peers.

Figure 2

Distribution Of Adult Prose Literacy Proficiency Levels, 1994 – 1998



Note. This figure shows adults' prose proficiency, organised into levels of increasing ability (5 is the highest). Anglophone countries' names are capitalized and their lowest and highest ability groups are circled. Adapted from Burnett (2005) and OECD / Statistics Canada (2000).

Despite high levels of funding, teacher training and support available for individuals, these examples show that reading skills amongst both students and adults from anglophone countries stubbornly continue to show significantly higher numbers of poor readers than other countries, notwithstanding their success in improving their overall outcomes. This begs the question as to whether there is perhaps something intrinsic to the language itself which is holding their young readers back.

Given the investment in time and money poured into the teaching of reading in anglophone countries, it is surprising how little attention education leaders pay to language-complexity variables when comparing their results to those from other countries, as opposed to differences in the mechanics of reading education, although this perhaps may be attributed to the lack of relevant data reported by either the OECD or the IEA.

It seems that the most influential factor when it comes to ease of initial reading acquisition (at least for alphabetic languages) is orthographic depth (Schmalz et al., 2015). Other variables such as the language's morphological complexity might also have an effect on the distribution of reading acquisition and comprehension ability as early readers grow into more advanced readers (Borleffs et al., 2017, 2019). It therefore seems strange that, as Knight et al. (2019) point out, comparative reading assessments such as PISA and PIRLS do not include "orthographic complexity as a variable that can differentiate nations' reading and academic achievement" (p. 7). Variations in morphological complexity remain, likewise, unrepresented in the data collected. This may not be an altogether fair complaint to level at the testing organisations, given the absence of any universally agreed-upon measures of either variable. Nevertheless, implementing changes in education while ignoring or being unaware of the influences of these language-complexity factors runs a variety of risks for the various stakeholders.

Purpose

The purpose of this study is to examine firstly whether language complexity variables are associated with different patterns of reading ability distribution. Specifically, variables such as morphological complexity or orthographic depth may give rise to groupings of languages with similar patterns of reading ability

distribution. This study will also investigate whether there are any differences in relation to language clusters in monolingual vs. non-monolingual anglophone educational contexts. Reading ability data from the PISA 2018 and the PIRLS 2016 cycles will be analysed using correlational and analysis of variance analyses to provide the basis for these investigations.

Research Questions

In this study, the following questions will be addressed:

1. Are the distribution patterns of reading ability range in the data from the PISA 2018 and PIRLS 2016 cycles associated with orthographic depth?
2. Are the distribution patterns of reading ability range in the data from the PISA 2018 and PIRLS 2016 cycles associated with morphological complexity?
3. Are there any differences in the distribution patterns of reading ability in the PISA 2018 and PIRLS cycles associated with a disparity between home-language and the language of the test in anglophone contexts?

As mentioned above, various international governmental departments and education bodies afford PISA and PIRLS results and rankings a significant degree of importance. Changes in a country's ranking are often reported widely and can have significant impacts for the various stakeholders in education systems. Many anglophone countries are placed well above the mean for reading ability scores, but they also display surprisingly large numbers of children who perform below the mean. There is frequent discussion, both in the media and from politicians as to the cause of this, with the blame often laid at the doors of the educators, or the prevailing

educational philosophy of the day. If it can be shown that an intrinsic aspect of the language itself, such as its orthography, is connected to this anomalous distribution, this will be of value to stake holders all across the education sector. It may also provide evidence for the value in renewed research into alternative spelling systems for the deepest orthographies.

Definition of Key Terms

Agglutinative Morphology: A morphology that allows morphemes to be stacked together to combine meanings into a single word.

Grapheme: "A unit of a writing system consisting of all the written symbols or sequences of written symbols that are used to represent a single phoneme" (Collins English Dictionary, n.d.)

Morpheme: "The smallest grammatical unit of speech" (Britannica, n.d.). This could be a whole word, such as "*drink",* or it could be an element of a word, such as *"un-"* and *"-able"* in *"undrinkable"*.

Morphology: "The system of word-forming elements and processes in a language" (Merriam-Webster, n.d.). Analytical morphologies (such as Mandarin Chinese) only allow for a single morpheme per word, whereas synthetic morphologies (such as Turkish) allow many morphemes to be added together in a single word.

Orthography: "The representation of the sounds of a language by written or printed symbols." (Merriam-Webster, n.d.)

Orthographic Complexity: The total number of GPCRs needed to describe the language's orthography. The greater the number of GPCRs required, the higher the complexity.

Orthographic Transparency: A measure of how many words in a language cannot be pronounced using the GPCRs of that language. The lower the proportion of words that can be correctly pronounced, the lower the language's transparency.

Phoneme: "The smallest unit of sound which is significant in a language." (Collins English Dictionary, n.d.)

CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

In this study, the aim is to explore the relationships between language-complexity measures of readability across different languages and the distribution patterns of reading ability as found in PIRLS and PISA data from the 2016 and 2018 cycles, respectively. Reading is a skill that is regarded as fundamental in all developed societies, and a central part of their education systems. Although the reading ability scores of anglophone countries in both PISA and PIRLS assessments show that many of their children score well above the median, they also show that anglophone countries have many more children who score poorly than would be expected given their places in the rankings and the patterns of reading ability displayed by similarly ranked countries. There is a good degree of variety between the teaching approaches taken by the various Anglophone countries, but so far, no Anglophone country has found a way to avoid large numbers of poorly-performing readers. Is it possible that language-specific differences might be playing a role in these seemingly unavoidable "failures" of their education systems?

Research into the differences in complexities between languages falls under the field of typological linguistics, which Croft describes broadly as having to do with "cross-linguistic comparison of some sort" (2002, p. 1). To facilitate these comparisons, languages are classified according to their structural differences, such as variations in their "phonological, morphological, grammatical, syntactic, lexical, pragmatic, semantic, etc. systems" (Velupillai, 2012, p.15). As this study is interested in differences in the *current* state of reading abilities across different

countries, it will focus on *synchronic* linguistic typological measures of morphological, orthographic and phonological structures, as opposed to how those structures have changed over time (or *diachronic* measures). With reference to the orthographies of the different languages, this study will focus purely on the *spelling* conventions within that language, and ignore other aspects such as punctuation, capitalisation, etc. This study will also briefly look at lexical structures, but only in the context of describing one of the models involved, rather than identifying any measures associated with them.

Potential Causes of Differences in Patterns of Reading Across Languages and Cross-Cultural Differences in Reading Acquisition

The patterns of development and speed of achievement that children display when they are learning to read are very heavily dependent on which language they are learning to read in (Frost et al., 1987; Galletly & Knight, 2017; Huang & Hanley, 1997; Katz & Frost, 1992). Perhaps the most unavoidable underlying cause of the differences between patterns of reading acquisition across different languages is the fact that written languages must necessarily represent particular spoken languages, which themselves vary in important ways. For example, languages can vary on how frequently consonant clusters occur, whether vowel clusters (diphthongs) are present, or whether the morphology of the grammar is analytical or synthetic. Analytical morphologies (such as Mandarin Chinese) only allow for a single morpheme per word, whereas synthetic morphologies (such as Turkish) allow many morphemes to be added together in a single word.

Orthographic Depth

One area of difference which has been found to be particularly important when investigating differences between reading acquisition in different languages is

the regularity of the spelling system – its orthography – utilised by a language (Schmalz et al., 2015). Towards the end of the 1980's, the term "orthographic depth" started to be used to capture some of this variation in the complexity of the writing systems of different languages, with Katz and Frost proposing the Orthographic Depth Hypothesis. Frost et al. (1987) describe a shallow orthography as one in which "lexical word recognition […] is mediated primarily by phonemic cues generated pre-lexically by grapheme – to – phoneme translation" (p.113). They go on to further contrast this with a deep orthography, which "relies strongly on orthographic cues, whereas phonology is derived from the internal lexicon". To illustrate this concept, it is perhaps useful to compare the Turkish orthography that is generally considered to be *shallow* with English orthography that is considered *deep* (Landerl et al., 1997; Öney & Durgunoğlu, 1997). If we take the Turkish grapheme <ç>, it will be pronounced as /ch/ in nearly every word it appears in. Likewise, if a spoken Turkish word contains the phoneme /ch/, it is nearly certain it will contain a <ç> at that point. This pattern is repeated for nearly all the graphemes and phonemes used in Turkish with the result that the Turkish orthography has one of the lowest levels of depth or complexity of all languages. Consequently, it is possible to reliably and accurately read nearly any Turkish word using simple *grapheme-phoneme correspondence rules* (GPCRs), even if the reader has never encountered that word before. English, on the other hand, shows an extreme level of orthographic complexity in both directions when compared to other languages (van den Bosch et al., 1994). For example, the phoneme /or/ can be represented by the graphemes:

- <or> as in (fork)
- <ore> as in (fore)
- <au> as in (aught)

- <aw> as in (thaw)

Equally, the letter ‘g’ could appear in the graphemes:

- <g> pronounced /g/ (go)
- <gn> pronounced /n/ (gnome)
- <ng> pronounced /ŋ/ (sing)
- <ge> pronounced /ʒ/ (mirage) OR /dʒ/ (cage)
- <gi> pronounced /g/ (gift) OR /dʒ/ (gilet)
- <gh> pronounced /g/ (ghost) OR /əʊ/ (though) OR /f/ (tough) OR /ə/ (thorough)

English is clearly a very complex orthography, and there have been many attempts to calculate its total number of GPCRs (Berndt et al., 1987; Brooks, 2015; Gontijo et al., 2003). While there does not seem to be a definitive answer (partly due to different outcomes depending on British vs American English, differences in accents, counting methods and other sources of disparity), nearly one word in six does not conform to the most common GPCRs. As a result, the only way to accurately generate these differing pronunciations is to use a lexical, top-down process; using GPCRs alone will likely result in a mistake. It has been argued that a language with both such high *complexity* and *irregularity* would impose a higher cognitive load on the reader (Knight et al., 2017, 2019), and would therefore require a higher degree of cognitive development in order to become an accomplished reader, as has been illustrated by the following studies.

These arguments all relate to writing-systems based on alphabetic orthographies. However, it should be remembered that not all languages use alphabetic writing systems, with syllabic and logographic writing systems being used

in a variety of other languages. As such, the concept of orthographic depth cannot be taken as being exhaustively relevant for all orthographies.

Early Studies into the Impact of Orthographic Depth on Reading Ability. Frith et al. (1998) conducted two studies to examine how the differences in orthographic consistency between English and German affected phonological recoding skills in German and English speaking 7- to 12-year-olds (German orthography is significantly more consistent that English). In the first study, the authors asked children to read from sets of words and non-words which had been created to control for readability across the two languages. Analysis of their speed and accuracy showed that English speaking children up to age 9 made more mistakes and read more slowly when reading non-words than their German-speaking counterparts. By age 12, the English-speaking children had caught up with respect to accuracy, but were still lagging behind on speed. In addition, vowel errors and word substitutions were found to be significantly more common in English than in German. The authors concluded that the greater level of orthographic inconsistencies present in English "imposes a heavy burden on the beginning reader" (p. 40).

In their second study (Frith et al., 1998) 8- and 12-year-old English-speaking and German-speaking children read a larger collection of one-, two- and three-syllable words and non-words. Lexical frequency was varied from low- to high-frequency for all word lengths. For one- and two-syllable words and non-words, low-frequency words were read less accurately than high-frequency words or non-words in both languages (although the German children were much more accurate than their English counterparts). At age twelve, English- and German-speaking children read all words with similar latencies and levels of accuracy, although English-speaking children still made more errors when reading complex three-syllable non-words. The

authors concluded that “learning to read in English implies a delay in the acquisition of phonological recoding” (p. 49).

English is a very orthographically deep language, but is the effect still noticeable when examining the reading acquisition characteristics of speakers of less deep orthographies? Defior et al. (2002) studied the differences in reading acquisition in Portuguese and Spanish children in grades 1-4. The authors chose those two languages specifically as they considered the Spanish and Portuguese orthographies to be at the shallow end of the spectrum, although the GPCRs (*grapheme-phoneme correspondence rules)* are slightly more consistent and fewer in number in Spanish than in Portuguese. The results of the study showed that even in two shallow orthographies, slight differences in orthographic depth led to measurable differences in patterns of reading acquisition. The Portuguese children made more mistakes and read more slowly when reading pseudo-words than their Spanish counterparts, although the differences in accuracy and speed reduced as the children grew older. The authors pointed out that while Spanish achieved their maximum reading speed by grade 2, the Portuguese children did not achieve it until grade 3. Interestingly, when the pseudo-word reading errors were analysed, Spanish children of all ages made very few *lexical* errors (that is, they did not substitute many real words for the pseudo-words). Portuguese children, however, made many more of this type of error in grades 1 and 2, suggesting that the higher number and more complex GPCRs they had to learn lead them to use alternative *top-down* techniques to read unfamiliar words. The Spanish children were perhaps able to learn the complete suite of GPCRs necessary to accurately decode their language faster than the Portuguese children. These results should be compared with the English-speaking children

studied by Frith et al. (1998) who only reached speed parity with their German-speaking counterparts by age twelve.

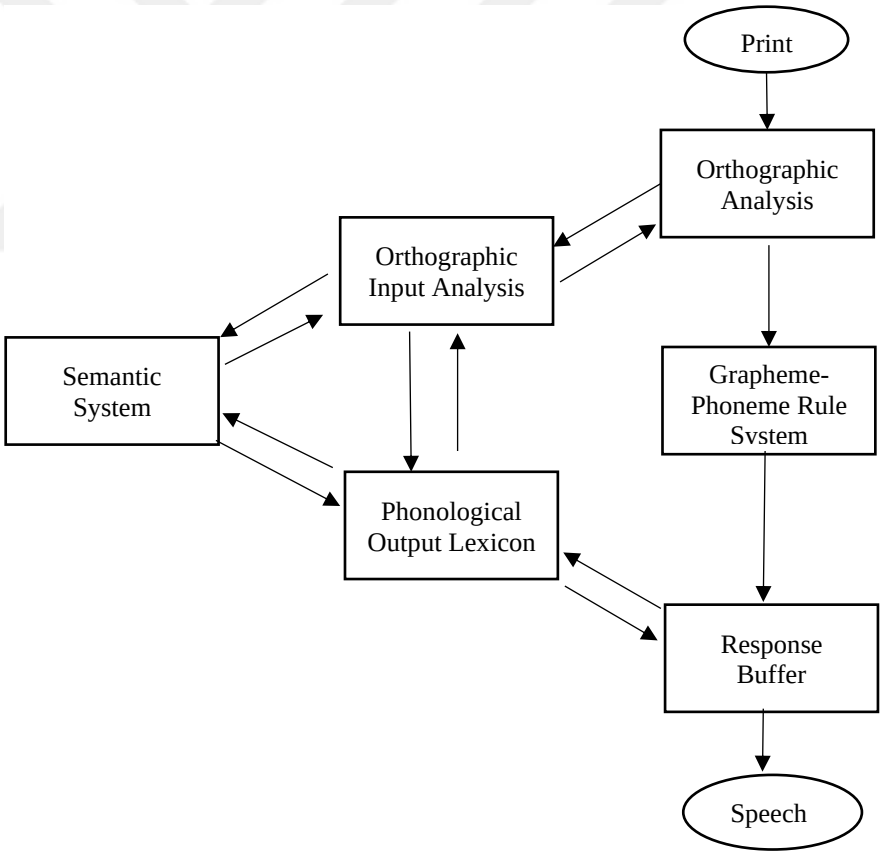
Towards a Unified Understanding of Orthographic Depth. The two studies discussed above used the term orthographic depth in slightly different ways, illustrating the incomplete state of academic understanding or agreement on the concept at the time. As discussed by Frith et al. (1998), English contains many *irregular* words whose pronunciations do not conform to the GPCRs and which therefore cannot be read accurately except by using a *lexical* route. Defior et al. (2002), by contrast, highlighted the increased *number* of GPCRs in Portuguese compared to Spanish, as well as the greater number of *context-dependent* GPCRs.

Schmalz et al. (2015) suggested a more nuanced way of conceiving of orthographic depth by bringing these subtly different components together. They proposed that "orthographic depth is a conglomerate of two separate constructs: the complexity of print-to-speech correspondences and the unpredictability of the derivation of the pronunciations of words on the basis of their orthography" (p. 1614). Complexity refers to how many GPCRs an orthography employs, whilst unpredictability refers to the likelihood that using the most common GPCRs will not result in the correct pronunciation. The authors show that because it is important to be able to consider the concepts of complexity and unpredictability separately, approaches such as measurements of *Onset Entropy* (Borgwaldt et al., 2005) are less than ideal as they tend to confound the two in certain languages where irregularity tends to show up after the first few letters in a word. Therefore, the authors chose to focus on the computationally derived *Dual Route Cascaded* (DRC) model (Ziegler et al., 2000). This makes use of a closely defined corpus of monosyllabic words to determine the numbers of single-letter rules, multi-letter rules and context-sensitive

rules, as well as the number of irregular words in a language. Irregular words are defined as those which cannot be correctly pronounced by making use of the GPCRs – in other words, those which have to be read by a lexical route rather than a sublexical route. In the DRC, a reader will initially attempt to read a word via the phonological (*sublexical)* route. If that method does not yield a meaningful pronunciation, a secondary *lexical* route will provide the pronunciation. See Figure 3 for details.

Figure 3

Basic Architecture of the Dual-Route Cascaded Model of Visual Word Recognition and Reading Aloud



Note. (redrawn from Coltheart et al., 2001, p. 213)

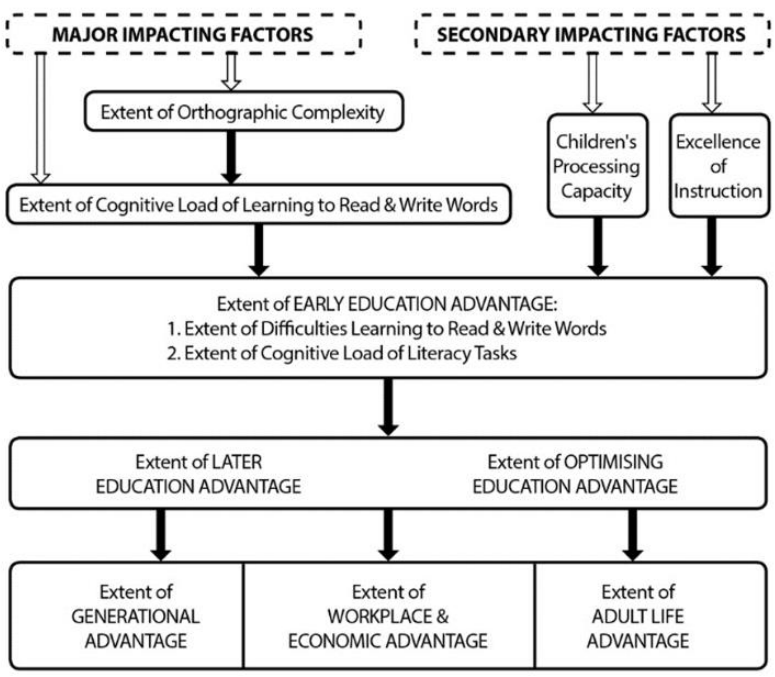
Schmalz et al. (2016) use the DRC model from Coltheart et al. (2001) to illustrate their proposal. In the DRC model, a reader will initially attempt to use

regular GPCRs to decode a word presented to them. If that route does not produce a meaningful result, a second *lexical* route will intervene to provide the most plausible result. This provides two axes on which to compare orthographies: How many GPCRs exist (complexity) and; how many exception-words there are (unpredictability). French, for example, has more than three times as many GPCRs as Dutch, but has a smaller percentage of irregular words (high complexity, low unpredictability). English meanwhile, has a high number of both GPCRs *and* irregular words (high complexity *and* unpredictability). The authors point out that while their approach does better at capturing the differences in orthographic depth across different languages, it is not complete. For instance, it does not capture the difficulties presented by words whose lexical information is *incomplete* – that is, words which need *semantic* information to be correctly read. For example, “wind” (/wɪnd/) and “wind” (/waɪnd/) are heteronyms whose correct pronunciation can only be derived by taking into account the sentence context. This form of depth is most clearly illustrated by un-pointed Hebrew and Persian, where texts intended for adult readers omit the diacritics that give information about vowels, leading to many words having identical letter strings. Another omission which the authors identify in their proposal is how a language deals with lexical stress assignment, that is, whether a language has strict rules about which syllable is stressed or not. These variables may yet prove to independently affect the readability of orthographies, although there is not currently enough experimental data to arrive at conclusions. The authors suggest future studies which could be conducted to tease apart the impacts of *complexity* vs *predictability* on reading behaviour, both for acquisition and skilled reading, and highlight the importance of controlling for these factors when designing future experiments.

Effect of Increased Orthographic Depth on Reading. Making use of arguments predicated on the different levels of orthographic complexities that exist between languages, Knight et al. (2019) propose an Orthographic Advantage Theory. Namely, that according to the level of their primary language of instruction's orthographic complexity, "nations experience disadvantage and potential advantage" across six "areas of education and national functioning" (p.5). As they explain, regular orthographies create a "low cognitive load […] for beginning readers" (p. 6) which allows for a more effective teaching and learning environment that has significant, life-long implications. Figure 4 illustrates these six areas of advantage:

Figure 4

Orthographic Advantage Theory



Note. This figure illustrates the various areas of advantage conferred on readers of languages with shallow orthographies, or conversely, those areas of extra difficulty experienced by readers of deep orthographies. From Knight et al., 2019, p.17

The authors emphasise the importance of the high cognitive load imposed by several aspects of English orthography when trying to make sense of the persistent failure to improve the reading outcomes of the poorest readers. While Knight et al. (2019) point out that "it seems likely that it is achieving success in the earliest years of Anglophone schooling that is crux if success in later school years is to be achieved" (p.20), they also point out the limits inherent in focusing purely on early instruction:

> Certainly, there seems value in nations carefully considering the type and number of GPCRs which are to be introduced for beginning readers, the order and timing of their introduction, and the need for children to experience ongoing strong success, including being highly successful when reading early books and texts. (pp. 24-25)

Morphological Complexity

A second source of complexity when learning to read rests on the underlying complexity of the language itself, and especially its morphological typology. Inflectional morphology is the means by which morphemes can be added to root words to impart extra information within a single, longer word. There are many different classes of information that languages allow for, such as number (singular or plural), tense, agreement, comparative or case, to name but a few. Even within those examples, different languages apply them in different ways. For example, English uses a verb suffix to denote the past perfect (I walk/walked), but makes use of an auxiliary verb for the future tense (I will walk) and both an auxiliary verb and a verb suffix for the past imperfect (I was walking.) French, by comparison, uses verb suffixes for the future (je marche/marcherai) and the past imperfect (je marchais), but uses an auxiliary verb and a verb suffix for the past perfect (j'ai marché). There is

also a “near future” tense in French that *does* use an auxiliary plus the infinitive of the verb, also indicated by a suffix (je vais marcher). Inflectional morphology is in fact hugely complicated, with confusing and seemingly arbitrary differences between languages, even ones that otherwise share significant similarities. As Baerman et al. (2015) put it: “At the micro-level, inflectional morphology is idiosyncratic because each language tells its own self-contained story, much more so than with other linguistic components” (p. 3). They also point to this seemingly vast array of potential differences between languages as having been a barrier to developing a systematic approach to describing, categorising and measuring them. Having said that, attempts have started to be made more recently to tackle the question of how to quantify the different levels of morphological complexity between different languages. These attempts can be broadly categorised into three approaches: Counting-based, entropy-based and description-based (Baerman et al., 2015, p. 5). Baerman et al. explain these approaches to morphological typology as follows:

- *Counting*-based approaches require the researcher to enumerate some comparable aspect of each language, such as the “number of morphemes found in words, however these are defined”.
- *Entropy*-based approaches calculate the degree of predictability in a language. Baerman et al. explain that “there is a greater degree of entropy […] if new instances are harder to predict”.
- *Description*-based approaches (formally known as *Kolmogorov Complexity)* relate the “minimum size of the rule required to generate the data” with complexity.

Although *counting*-based approaches are more frequently attempted, the authors argue that *entropy*-based and *description*-based approaches are perhaps

"better able to capture" differences between languages (p.6). This is especially true as it is difficult to determine or justify which morphological features of different languages should be compared when employing a *counting*-based approach.

Effect of Morphological Complexity on Reading. The vast majority of research investigating language-complexity factors affecting differences in reading acquisition has focused on the impact of orthographic factors. By contrast, the impact of factors associated with the morphological structures of different languages has seemingly attracted much less interest. Borleffs et al. (2019) reviewed studies investigating the impact of morphology in several highly complex agglutinative languages (Finnish, Basque, Turkish). The authors showed that children whose mother tongue is highly morphologically complex initially make use of grapheme to phoneme correspondences when they first start to learn to read (as do early readers of all alphabetic orthographies), but start to make more use of morphological information as they become more proficient readers. For example, Finnish children in their third year of schooling become better at reading morphologically complex words than morphologically simple ones.

Monolingualism vs Bilingualism or Multilingualism in Anglophone Countries

People who are only able to speak a single language (monolinguals) make up a minority of the world's population (Grosjean, 2021, p. 27). There are therefore many children who speak one language at home, but a different language at school. However, bilinguals or multilinguals are not equally proficient in the various languages that they speak. The degree of proficiency in secondary (or subsequent) languages exists on a continuum from low ability to near parity in ability (Incera & McLennan, 2018), and both the number of years and age at which speakers start to learn additional languages affect their proficiency (Bialystok, 2001).

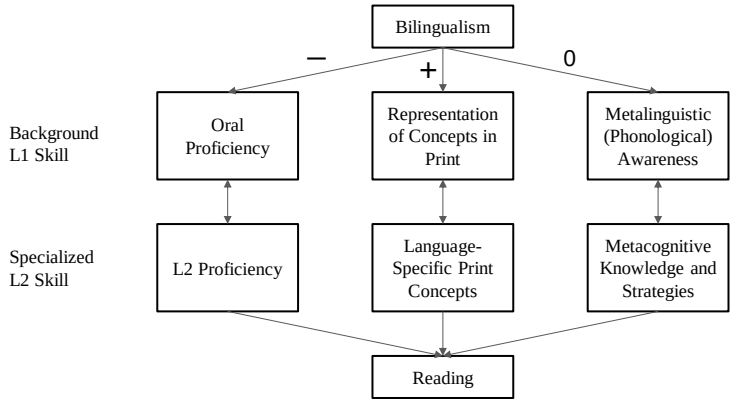
At the same time, it is known that infants who are raised in monolingual environments display different patterns of sensitivity to various linguistic cues in heard speech than infants raised in non-monolingual environments (Burns et al., 2007; Polka et al., 2017; Sundara & Scutellaro, 2011). Equally, it is also known that certain factors connected to growing up in multilingual environments have an impact both on early reading acquisition and on skilled adult reading ability (Bialystok, 2007; Verhoeven, 2007). For example, it has been shown that bilingual children have a smaller vocabulary in each of the languages they speak than do monolingual children in either of those languages (Bialystok, 2007). In England, it is also known that while a significant proportion of primary school pupils (15% in 2010) speak English as an Additional Language (EAL), these students do not perform as well as their monolingual peers (Burgoyne et al., 2011). It is therefore legitimate to ask whether population patterns of monolingualism vs. multilingualism might influence patterns of reading development, both at the individual and population levels. Indeed, among the Anglophone countries that participated in the 2012 PISA assessment round, Ireland and Canada had the highest achievement levels. As Knight et. al point out, this may have been because they experienced an "advantage built from multilingualism with [for example, Irish] children taught to speak and read Irish as well as English from the start of school, perhaps in combination with explicit, systematic word-reading instruction" (2017, p.8).

Effect of Bilingualism on Reading. Whilst it is known that there are factors that influence the patterns of reading development of children who grow up in multilingual environments, the picture is not uniformly positive. Bialystok (2007) identified three different categories of development and influence that affect reading acquisition in multilingual children: "Concepts of print, […] oral language

competence [… and] metalinguistic competence" (p.45). The author reviews a variety of research studies to demonstrate that differences in these competencies can provide either a positive, neutral or negative influence on reading acquisition, respectively (see Figure 5):

Figure 5

Relation Between First- and Second-Language Literacy Acquisition and Bilingualism



Note. This figure illustrates how different background skills, influenced by bilingualism, can affect specific skills related to reading. From Bialystok, 2007, p52.

"Concept of print is the understanding of the constituent notations that comprise the writing system and how those notations represent spoken language" (p. 59). Two sub-concepts of this that were investigated in bilingual children were "the understanding that the notational forms are invariant representations of meaning" and "the set of correspondences between forms and referents that are used in individual writing systems" (p. 61). The second of these concepts corresponds with knowledge of the specific orthographic depth of each language. The authors showed that 4-year-old bilingual children perform better on tests of invariance of print meaning, but when testing their understanding of print-meaning correspondence, the outcomes depended on the specific language-pair comparisons made. In some language-pair

comparisons, bilingual children performed no better than their monolingual counterparts whilst in others, the bilingual children performed significantly better, although bilingual children never performed any worse than their monolingual peers in this regard.

Oral language skills are another important difference between monolingual children and their multilingual peers. It has been known for some time that there exists a strong connection between oral language skills and reading acquisition (see review in Adams, 1990, as cited by Stahl et. al., 1990). The second language (L2) oral language skills in children who learn their second language after they start school necessarily lag behind those of their monolingual peers, meaning that this is a negative source of influence on L2 reading acquisition.

Metalinguistic competence includes ideas such as phonological and syllabic awareness, and it has been shown to be of fundamental importance to early reading acquisition in both alphabetic and non-alphabetic languages. If bilingualism confers a developmental advantage on metalinguistic capabilities, this would be a strong argument for bilingualism providing a net overall benefit for reading acquisition. A review of relevant studies by Bialystok (2007) showed that the answer is not completely clear. They showed that "[First,] the specific task rather than a global assessment of phonological awareness determines the outcome. […] Second, the bilingual advantages that were found occurred in kindergarten and usually disappeared by first grade" (p. 67). In other words, the influence of growing up in a multilingual setting does not seem to confer a specific metalinguistic advantage, especially by the age that children engage in either PIRLS or PISA assessments.

Conclusion

It is clear that there are many different variables that could potentially impact population level patterns of reading acquisition. Some of these variables are connected to the languages spoken and read in different countries. Two such variables are orthographic and morphological complexity. Research into these has shown that across all alphabetic orthographies so far studied, readers make use of GPCRs during their earliest stages of reading acquisition. For readers of simple, regular orthographies, the use of GPCRs remain a useful, reliable and rapid mechanism for reading and these children become proficient, rapid readers in a short space of time. While readers of orthographically complex orthographies also make early use of GPCRs, these are less reliable than in shallow orthographies. Readers of these orthographies therefore have to develop additional, more advanced and cognitively demanding skills, and therefore take longer to become fully proficient readers. For readers of morphologically complex languages, readers also start to make more use of language-specific morphological information embedded within the words (and non-words) that they read as they become more proficient. Languages vary independently in their degrees of orthographic and morphological complexities, so it might be expected that these factors have varying and independent levels of influence on population-level patterns of reading acquisition and reading ability. It is also possible that the patterns of this influence these factors exert might change at different stages of reading development. A third linguistic factor that could be influencing population-level patterns of reading acquisition is the proportion of students within any given country who are living and learning in multilingual environments. Whilst some differences between children growing up in multilingual environments and the monolingual peers have been shown to confer advantages

when it comes to reading acquisition, others have been shown to be either neutral or negative in their influence. It has also been shown that the specific language pairs involved can determine whether multilingual children perform better than their monolingual peers or not. However, as this thesis is mostly concerned with reading development in English, at least one of the comparison languages will always be the same.

Anglophone countries have made extensive use of international comparison and benchmarking reading assessments such as PISA and PIRLS with the aim of improving the educational outcomes of their young people. Whilst they have been able to improve the overall outcomes of their students, they continue to find that their lower ability students lag further behind their peers than is the case in many other countries. Given the potential influence of language-complexity variables on patterns of reading acquisition, it might therefore be valuable to investigate potential correlations between population level reading ability outcomes and measures of orthographic and morphological complexity, as well as patterns of monolingual vs multilingual English-reading students.

CHAPTER 3: METHOD

Introduction

The purpose of this study in the first instance was to investigate possible associations between language-complexity factors related to orthography and morphology and patterns of reading ability distribution across different countries, based on data from the PISA 2018 and PIRLS 2016 cycles. In the second instance, this study investigated whether monolingual and bi- or multi-lingual students in Anglophone countries display differences in their patterns of reading ability distributions. This chapter presents the methodology employed in this study.

Research Design

This is primarily a correlational study that used datasets from the 2016 cycle of the Progress in International Reading Literacy Study (PIRLS) conducted by the IEA as well as the OECD's Programme for International Assessment (PISA) 2018 cycle to search for potential relationships between language-complexity variables and patterns of reading ability distributions. The fields of developmental psychology and education research have long used correlational analyses to investigate relationships between variables of interest, often as a first step towards developing experimental studies to research causal connections. For example, the correlation between critical reading ability and mathematical critical thinking (Wikanengsih et al., 2020), the correlation between phonological awareness and different sub-skills of reading fluency (Elhassan et al., 2017), or the correlation between certain aspects of syntactic awareness and reading acquisition (Tunmer et al., 1987).

This study employed an explanatory correlational approach to answer the first two research questions, aimed at finding relationships between the absolute reading ability ranges in various countries as measured by the PIRLS (2016) and PISA (2018) datasets and the language-complexity variables Orthographic Complexity, Orthographic Transparency, Morphological Complexity and Morphological Unpredictability of the languages in those countries. For the third research question, a causal comparative approach was employed. For both datasets, the dependent variable was the range of reading ability and the independent variable was Home Language. For the PISA 2018 dataset, there were two groups for Home Language (children who speak the language of the test at home vs. children who speak a different language at home) and so an independent samples t test was employed. For the PIRLS 2016 dataset there were four groups in the Home Language variable (children who always/almost always/sometimes/never speak the test language at home) so one-way ANOVA was conducted.

It is important to bear in mind that correlational studies cannot be interpreted as providing evidence of causal connections. Any correlations that might be found would need to be the subject of future experimental studies to investigate the existence or direction of any causality.

Context

Around the world's educational communities, the late 20th Century saw a marked increase in the interest in conducting benchmarking and international comparisons (Addey & Sellar, 2018; Steiner-Khamsi, 2020). A plethora of different International Large-Scale Assessments and benchmarking studies were developed, designed to assess a variety of aspects of educational outcomes at different ages and in different countries.

The PISA assessment is one of the most widely recognised and respected such tests. It is sat by fifteen-year-old students around the world every three years and covers various academic disciplines, including reading. Its purpose is to allow for cross-cultural and international comparisons of educational success, as well as tracking progress over time. In 2018, some "600,000 students representing about 32 million 15-year-olds in the schools of the 79 participating countries and economies sat the 2-hour PISA test" (OECD, 2019, p.19). The students ideally comprise a representative cross-section of the participating countries' populations, although it can sometimes be difficult to ensure this – especially in countries with diverse geographical and linguistic settings.

Whilst reading is only a sub-section of the PISA test, PIRLS is developed with the specific purpose of monitoring trends in reading achievement in Grade Four. PIRLS focuses on the ability of children to engage with text in a variety of meaningful ways, providing this definition of reading literacy:

> Reading literacy is the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment. (Mullis et al., 2015, p.12)

Whilst PISA and PIRLS both collect data on the reading ability of school pupils, the two assessments are sat by children at very different ages and, therefore, at different stages of reading development. For this thesis, the two studies are used in a complementary manner, allowing the present study to investigate and illustrate changes in the patterns of influence linked to the various factors of interest as the children develop into advanced readers.

Anglophone countries represent a significant proportion of the world's population. As well as the large number of schools which exist in monolingual, English speaking contexts, there also exist a significant number of bilingual or multilingual linguistic contexts across different countries. Results from PIRLS and PISA cycles are regularly used by national governments and ministries of education to inform changes in educational methods in attempts to improve national achievement levels (Addey & Sellar, 2018; Arikan et al., 2020). Despite several decades of significant investment in reading education across many anglophone countries, they seem to be unable to avoid much larger numbers of low-performing students than might be expected (Knight et al., 2019). Other countries, by contrast, seem to have been able to improve the reading results of both their weaker and their stronger readers alike. It is therefore of interest to ask whether there may be factors particular to English that play a role in the differences in reading abilities of poor and strong readers.

Sampling

Sampling by OECD

The OECD state that when developing the PISA assessment, they use a "stratified two-stage sample design, where schools are sampled using probability proportional to size […] and students are sampled with equal probability within schools" (OECD, 2018).

PISA's target population is "students between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing period, attending educational institutions located within the country, and in grade 7 or higher" (OECD, 2016, p. 2). Whole schools may be excluded if they provide "instruction only to students in the excluded categories […] such as schools for the

blind" (OECD, 2016, p. 3). Countries are able to specify within-school exclusion categories, such as those students with "severe dyslexia, dysgraphia, or dyscalculia" (OECD, 2018, p. 6). Once the schools have been selected for inclusion, an equal number of students who fall within the necessary age range are selected at random, usually 35 students in countries that are taking part in Paper Based Assessment (PBA) and 42 students in countries that are taking part in Computer Based Assessment (CBA). Of the total level of exclusions from the 2018 round, the OECD state:

> In 31 of the 79 countries and economies that participated in PISA 2018, the percentage of school-level exclusions amounted to less than 1%; it was 4% or less in all except five countries. When the exclusion of students who met the internationally established exclusion criteria is also taken into account, the exclusion rates increase slightly. However, in 2018, the overall exclusion rate remained below 2 % in 28 participating countries and economies, below 5 % in 63 participating countries and economies, and below 7 % in all countries except Sweden (11.1 %), Israel (10.2 %), Luxembourg and Norway (both 7.9 %). (OECD, 2019c, p. 37)

Notwithstanding any exclusions, the goal is to obtain data representing at least 95% of the target populations for each country.

Sampling by IEA

When developing the PIRLS assessment, the IEA selects subjects via a "two-stage random sample design", the first step identifying individual schools and the second step selecting whole classes within those schools (Laroche et al., 2017, p. 3.1). The sampling definitions include all schools with grade 4 classes, whether or not they are run within the national education systems of their respective countries.

However, as the age at which children start school is not the same for all countries, the target population is defined as follows: “All students enrolled in the grade that represents four years of schooling […] providing the mean age at the time of testing is at least 9.5 years” (Laroche et al., 2017, p. 3.3). As with PISA, there are rules governing exclusions at both school and individual levels. For example, a school could be excluded if it had an especially small number of students in the target grade level. An individual student might be excluded, for example, if they have specific functional or intellectual disabilities, or if they are non-native speakers of the test language. Of the total level of exclusions from the 2016 round, the IEA state:

> For most PIRLS participants, the overall percentage of excluded students (combining school and within-school levels) was 5 percent or less after rounding. However, Austria, Belgium (French), Canada, Denmark, Hong Kong SAR, Latvia, Malta, and Portugal, as well as the benchmarking participants Denmark (3) and Madrid (Spain), had exclusions accounting for between 5 and 10 percent of the desired population. Israel and Singapore had exclusions exceeding 10 percent. (LaRoche & Foy, 2017, p. 5.8)

Notwithstanding any exclusions, the sample should represent at least 95% of the target population in each country. In addition, only 2% of the target population should be excluded due to attending very small schools.

Sampling Used for the Current Study

For the current study, the countries chosen for inclusion depended on the availability of data on the language-complexity variables for the language(s) in which students took the test. That is, whether there was a reliable measure available for the Orthographic Complexity, Orthographic Depth, Morphological Complexity or Morphological Unpredictability of the language. Developing such scales was without

the scope of this study, so it was necessary to rely on the work of others, although this meant that much of the data from the PIRLS (2016) and PISA (2018) assessments could not be included in all analyses due to missing data. Data was considered at a country-level for those countries whose languages were represented in the Orthographic and/or Morphological measures, mediated by the test-languages used: For countries that administered the test in more than one language, reading range data was considered for each language separately. There were seventy-nine participant countries in PISA 2018 and 61 participant countries in PIRLS 2016 (including 11 benchmarking entities). Of those countries, the countries which could be included in analyses involving Orthographic Depth (Complexity and/or Transparency) are listed in Table 1, organised according the language (or languages) they administered the test in. For the same information organised by country rather than language, see Appendix A. The countries whose data were included in analyses involving Morphological Complexity and Unpredictability are listed in Table 2.

Table 1

Languages Included in Analyses Involving Orthographic Depth, Listing the Countries That Administered PISA (2018) and/or PIRLS (2016) in Those Languages

Language	PISA (only)	PIRLS (only)	PISA and PIRLS
Arabic	Jordan	Bahrain, Oman	Israel, Morocco, Qatar, Saudi Arabia, United Arab Emirates
Danish	Denmark	-	-
Dutch	-	-	Belgium, Netherlands
English	Brunei Darussalam, Hong Kong, Lebanon, Luxembourg, Malaysia, Malta, Panama, Philippines,	Bahrain, Oman, South Africa, Trinidad & Tobago	Australia, Canada, Ireland, Macao, New Zealand, Qatar, Saudi Arabia, Singapore, United Arab Emirates, United Kingdom, United States
French	Lebanon, Luxembourg, Switzerland	-	Belgium, Canada, France
German	Luxembourg, Switzerland	-	Austria, Belgium, Germany, Italy
Italian	Switzerland	-	Italy
Korean	Korea	-	-
Portuguese	Brazil	-	Macao, Portugal

Language	PISA (only)	PIRLS (only)	PISA and PIRLS
Russian	Belarus, Estonia, Georgia, Moldova, Moscow Region (RUS), Tatarstan (RUS), Ukraine	-	Baku (Azerbaijan), Kazakhstan, Latvia, Lithuania, Russian Federation
Serbo-Croatian	Bosnia & Herzegovina, Croatia, Serbia	-	-
Spanish	Colombia, Costa Rica, Dominican Republic, Mexico, Panama, Peru, Uruguay	Spain	Argentina, Chile
Turkish	Turkey	-	-

Note. Some countries appear more than once, depending on how many languages they administer the tests in.

Table 2

Languages Included in Analyses Involving Morphological Complexity and Unpredictability, Listing the Countries That Administered PISA (2018) and/or PIRLS (2016) in Those Languages

Language	PISA	PIRLS	PISA & PIRLS
Afrikaans	-	South Africa	-
Albanian	Albania, Kosovo, Montenegro, North Macedonia	-	
Arabic	Jordan	Bahrain, Oman	Israel, Morocco, Qatar, Saudi Arabia, UAE
Azerbaijani	-	-	Georgia, Baku (Azerbaijan)
Bulgarian	-	-	Bulgaria
Czech	-	-	Czech Republic
Danish	-	-	Denmark
Dutch	-	-	Belgium, Netherlands
English	Brunei Darussalam, Hong Kong, Lebanon, Luxembourg, Malaysia, Malta, Panama, Philippines	Bahrain, Oman, South Africa, Trinidad & Tobago	Australia, Canada, Ireland, Macao, New Zealand, Qatar, Saudi Arabia, Singapore, UAE, UK, USA
Estonian	Estonia	-	-
Finnish	-	-	Finland
French	Lebanon, Luxembourg, Switzerland	-	Belgium, Canada, France
Georgian	-	-	Georgia
German	Luxembourg, Switzerland	-	Austria, Belgium, Germany, Italy
Hebrew	-	-	Israel
Hungarian	Romania, Serbia	-	Hungary, Slovak Republic
Icelandic	-	Iceland	-
Indonesian	-	Indonesia	-
Italian	Switzerland	-	Italy
Japanese	Japan	-	-
Korean	Korea	-	-
Latvian	-	-	Latvia
Lithuanian	-	-	Lithuania
Macedonian	North Macedonia	-	-
Malay	Malaysia	-	-
Maltese	-	Malta	-

Language	PISA	PIRLS	PISA & PIRLS
Polish	-	-	Lithuania, Poland
Portuguese	Brazil	-	Macao, Portugal
Romanian	Moldova, Romania	-	-
Russian	Belarus, Estonia, Georgia, Moldova, Moscow Region (RUS), Tatarstan (RUS), Ukraine	-	Baku (Azerbaijan), Kazakhstan, Latvia, Lithuania, Russian Federation
Slovak	-	-	Slovak Republic
Slovenian	-	-	Slovenia
Serbo-Croat	Bosnia & Herzegovina, Croatia, Serbia	-	-
Spanish	Colombia, Costa Rica, Dominican Republic, Mexico, Panama, Peru, Uruguay	Spain	Argentina, Chile
Swedish	-	-	Finland, Sweden
Thai	Thailand	-	-
Turkish	Turkey	-	-
Ukrainian	Ukraine	-	-
Zulu	-	South Africa	-

Note. Some countries appear more than once, depending on how many languages they administer the tests in.

It can be noted that Spain does not appear in the PISA data. This is surprising, given that Spanish is one of the languages included in both Orthographic Depth (OD) and Morphological Complexity (MC) analyses. However, as the OECD explain, some of Spain's reading data "show implausible response behaviour amongst students" (2019b, p.208). This makes it impossible to calculate reliable results, and so the data from Spanish students was not included.

When searching for patterns of reading ability affected by monolingual vs. bilingual or monolingual settings, only countries for which English was one of the main languages of education were included. See Table 3 for a list of which countries were included in the analyses.

Table 3

Countries Included in Analyses Involving Monolingual and Non-Monolingual Students, Listing the Countries That Administered PISA (2018) and/or PIRLS (2016)

PISA (2018)	PIRLS (2016)
Australia	Australia

PISA (2018)	PIRLS (2016)
Canada	Canada
Ireland	England
New Zealand	Ireland
Singapore	Northern Ireland
United Kingdom	Singapore
United States	United States

Details of which measures were chosen for the variables measuring different aspects of Orthographic and Morphological differences are addressed in the instrumentation section below.

Instrumentation

Selection Criteria for Measures of Language-Complexity Variables

In order to facilitate correlational analyses, the language-complexity variables were selected according to the following criteria:

- They should contain data points in common with the languages tested by PISA (2018) and/or PIRLS (2016).
- Their method of calculation should be well documented and peer-reviewed.

Orthographic Depth

Various attempts to develop a measure of orthographic depth have been developed over the latter half of the twentieth century. Schmalz et al. (2015, 2016) reviewed several of these and showed that it was important to dissociate the concepts of *complexity* and *unpredictability.* Using an approach taken from the Dual Route Cascaded (DRC) theory (Gibson et al., 2001), the authors collated results for Dutch, English, French, German and Italian from several different sources (Schmalz et al., 2015, p. 1620), although they did not report an *irregularity* score for Italian (see Table 4). Schmalz explained in an email (personal communication, 17 August, 2021) that this was because the lexical route of the model had not been implemented as

there are so few monosyllabic words in Italian. In addition, the Italian orthography is so regular that the only irregular words in the whole Italian lexicon would most likely be 'loan' words from other languages. As the data collated by the authors only covered five languages (and one of these was incomplete), further datapoints were identified from additional studies. Ulicheva et al. (2016) calculated values for the Russian language while Greek values were calculated by Kapnoula et al. (2017) and Protopapas et al. (2009). The Greek values were not calculated in exactly the same way because the DRC model was adapted to include multi-syllabic words. See Table 4 for the full set of DRC values collected.

Table 4

Measures of Complexity and Irregularity (Unpredictability) For Dutch, English, French, German, Italian, Russian and German Based on the DRC Model.

Measure	Dutch[i]	English[i]	French[ii]	German[i]	Italian[iii]	Russian[iv]	Greek[v]
Total rules (DRC)	104	226	340	130	59	101	118
Single-letter rules (DRC)	51 (49.0%)	38 (16.9%)	46 (13.5%)	44 (38%)	19 (32.2%)	33 (32.7%)	N/A
Multi-letter rules (DRC)	42 (40.4%)	161 (71.2%)	218 (64.1%)	55 (42.3%)	8 (13.6%)	5 (5.0%)	N/A
Context-sensitive rules (DRC)	11 (10.6%)	27 (11.9%)	76 (22.4%)	31 (23.8%)	32 (54.2%)	63 (62.4%)	N/A
Irregular words (%)	6.3	16.9	5.6	10.5	NA	0.26	4.9

Note. The values listed in this table were collated from the following sources: i (Ziegler et al., 2000), ii (J. C. Ziegler et al., 2003), iii (Schmalz et al., 2015), iv (Ulicheva et al., 2016), v (Kapnoula et al., 2017; Protopapas & Vlahou, 2009)

Although the DRC values collected here are a useful start, there are several potential problems, especially for the irregularity (unpredictability) measure. Firstly,

the measures were calculated by different research groups, making it uncertain that they were all calculated in exactly the same way. Indeed, the Greek values are different by definition, given that they make use of a corpus of multi-syllabic words, rather than the original mono-syllabic models for the other languages. Secondly, the restriction of the corpora of words used to monosyllables is potentially problematic in itself as it excludes such large and non-uniform proportions of the words in each language's lexicon, each of which will impose a variety of different phonotactic constraints on syllabic structure, stress patterns and so on. Thirdly, values for only seven languages were found, two of which were incomplete. Therefore, alternative potential measures of orthographic complexity were identified that might mitigate these potential problems. It was not possible to identify any alternative measures for orthographic *complexity*, but a more appropriate measure of orthographic unpredictability was identified: Marjou (2021) approached the problem from the field of computational linguistics and natural language processing. Using language corpora culled from https://wiktionary.org, Marjou trained an Artificial Neural Network (ANN) to calculated transparency scores for both reading and writing across sixteen languages, all of which were incorporated into a single analysis. This approach gave rise to the name the author coined for their measure: Orthographic Transparency Estimation with an Artificial Neural Network (OTEANN).

The work of Marjou (2021) has two main advantages in the context of this study: Firstly, Marjou calculated a transparency score (analogous to the *irregularity* score for the DRC model) for a greater number of languages; Secondly, the scores were all calculated by the same researcher with the same tool and using closely matched linguistic corpora, meaning that the scores are more reliably comparable, and; Thirdly, the corpora were not limited to any particular number of syllables. The

OTEANN was trained on words of a full range of lengths, and so might reasonably be supposed to be more reliably representative of the orthographies in their entirety. OTEANN scores were calculated for a total of sixteen languages, thirteen of which were relevant for this study (Arabic, German, English, Spanish, Finnish, French, Italian, Korean, Dutch, Portuguese, Russian, Serbo-Croatian and Turkish).

No further metrics were identified that could specifically measure the *complexity* of different languages without confounding their *irregularity* or *unpredictability*, as discussed by Schmalz et al. (2015).

For the remainder of this study, *Orthographic Complexity* and *Orthographic Transparency* will be understood as two independent aspects of *Orthographic Depth*, each with their own measure.

Morphological Complexity

As with a language's orthographic depth, a language's morphology can be conceived of as varying in difficulty along two different axes: *Complexity* (how many rules or possibilities exist for how words can be constructed from their constituent morphemes) and *(un)Predictability* (how easy or difficult it is to correctly guess how a word will finish, given how it starts). Gutierrez-Vasques et al. (2019) approached this problem by using Natural Language Processing (NLP) methodology to quantify the "morphological complexity by combining two different measures over parallel corpora: (a) the type-token relationship (TTR); and (b) the entropy rate of a sub-word language model as a measure of predictability" (p. 2). The two corpora the authors identified were (i) a portion of the Bible Parallel Corpus that overlapped over 47 languages, and (ii) the JW300 parallel corpus, composed of magazine articles in 133 languages from the Jehovah's Witnesses website. Analysing two independent corpora allowed the authors to validate their results, and the two corpora

were chosen as they covered a large number of languages, but importantly, they both fell within the same broad category of writing (religious). For the purposes of this study, the values derived from the JW300 corpus were used for two reasons. Firstly, the JW300 corpus covered a larger range of languages; And secondly, the JW300 corpus consisted of magazine articles, which were therefore closer in written style to the reading matter used in the PISA and PIRLS reading tests than bible verses would be. Gutierrez-Vasques et al. (2019) calculated the TTR value for each language by dividing the total number of word types (vocabulary size) by the number of word tokens (total number of words in each corpus). To calculate the morphological predictability of a language, the authors used a feed-forward "neural probabilistic language model" to "estimate a stochastic matrix P, where each cell contains the transition probability between two sub-word units in that language" (p. 4). The authors calculated a value for each language using sub-word units of size one character (H1) and three characters (H3). They found that the predictability value as calculated by H1 correlated very strongly with the TTR complexity value, but the H3 value did not and was therefore seemingly capturing a different aspect of the morphologies. As such, this study made use of the TTR complexity values and the H3 uncertainty values. For TTR, higher values represent greater complexity. For H3, higher values represent higher uncertainty (lower predictability). TTR and H3 values for thirty-eight languages were used in this study, covering test-languages from either PISA (2018), PIRLS (2016) or both.

It is possible that other, more appropriate, accurate or reliable metrics may exist (or could be developed) to measure the orthographic and morphological complexities of the languages involved in this study. However, it is beyond the scope of this study to enter into extended analysis of this question.

Method of Data Collection

This descriptive correlational study involved the use of pre-existing data from:

- the PISA 2018 Student Questionnaire dataset, as available directly from the OECD's website (https://www.oecd.org/pisa/)
- the PIRLS (2016) Student dataset, as available directly from the IEA's website (https://timssandpirls.bc.edu/pirls2016/international-database/index.html)

These datasets were manipulated to incorporate additional variables as described above, but no additional or original data was collected.

Variables

Reading Ability Range

Instead of examining either mean or maximum reading abilities, this study examined the magnitude of the range of reading ability in each country. Analyses of ability range for ILSAs such as PISA and PIRLS usually ignore the very top and bottom in order to avoid outliers unduly affecting results – choosing to examine either 10^{th} to 90^{th} percentile (NCES, 2018; OECD, 2019b) or 5^{th} to 95^{th} percentile (Mullis et al., 2017; Tunmer et al., 2013). Using the IDB Analyzer from the IEA to create the necessary SPSS code, the 1^{st} to 10^{th} Plausible Value 1 in Reading (PVREAD01-10) were used to calculate the 5^{th} and 95^{th} percentiles. In SPSS, the difference between these two percentile scores was used to calculate the magnitude of the range of reading ability for each language in each country, which was recorded in a separate variable. Skewness and Kurtosis calculations show both to be between ±2, so this variable can be taken as being normally distributed.

Orthographic Complexity

Orthographic Complexity scores for seven languages (see Table 5) as measured by the total number of grapheme-phoneme correspondence rules in the Dual Route Cascaded model (Gibson et al., 2001) were entered into SPSS by recoding the LANGTEST_COG values for each country into a new variable. Skewness and Kurtosis calculations show both to be between ±2, so this variable can be taken as being normally distributed.

Table 5

DRC Total GPC Rules

	Dutch[i]	English[i]	French[ii]	German[i]	Italian[iii]	Russian[iv]	Greek[v]
Total rules (DRC)	104	226	340	130	59	101	118

Note. The total number of GPCR rules, as listed in this table, were collated from the following sources: i (Ziegler et al., 2000), ii (J. C. Ziegler et al., 2003), iii (Schmalz et al., 2015), iv (Ulicheva et al., 2016), v (Kapnoula et al., 2017; Protopapas & Vlahou, 2009)

Orthographic Transparency

Orthographic Transparency scores for thirteen languages (see Table 6), as measured by OTEANN (Marjou, 2021) were entered into SPSS by recoding the LANGTEST_COG values for each country into a new variable. Skewness and Kurtosis calculations show both to be between ±2, so this variable can be taken as being normally distributed.

Table 6

Orthographic Reading Transparency Scores for Thirteen Languages Used in This Study

Language	OTEANN (Read)	Language	OTEANN (Read)
Arabic	99.4 ± 0.3	Korean	97.5 ± 0.5
German	78.8 ± 1.5	Dutch	55.7 ± 2.2
English	31.1 ± 1.3	Portuguese	82.4 ± 0.9
Spanish	85.3 ± 1.3	Russian	97.2 ± 0.5
Finnish	92.3 ± 0.8	Serbo-Croatian	99.3 ± 0.3
French	79.6 ± 1.7	Turkish	95.9 ± 0.6
Italian	71.6 ± 0.9		

Note. Values taken from Marjou (2021)

Morphological Complexity

Morphological Complexity and Morphological Uncertainty scores for thirty-eight languages (see Table 7) were entered into SPSS by recoding the LANGTEST_COG values for each country into a new variable. Skewness and Kurtosis calculations show both to be between ±2, so these variables can be taken as being normally distributed.

Table 7

Measures of Morphological Complexity (TTR) and Morphological Uncertainty (H3) For Languages Analysed in This Study

Language	TTR	H3	Language	TTR	H3
Afrikaans†	0.05	0.67	Italian$^{\Delta}$	0.08	0.61
Albanian•	0.07	0.72	Japanese•	0.02	0.91
Arabic$^{\Delta}$	0.17	0.83	Korean•	0.06	0.91
Azerbaijani•	0.15	0.73	Latvian$^{\Delta}$	0.12	0.75

Language	TTR	H3	Language	TTR	H3
Bulgarian$^{\Delta}$	0.09	0.68	Lithuanian$^{\Delta}$	0.17	0.71
Croatian$^{\bullet}$	0.11	0.74	Macedonian$^{\bullet}$	0.08	0.70
Czech$^{\Delta}$	0.13	0.78	Malayalam$^{\bullet}$	0.27	0.70
Danish$^{\Delta}$	0.06	0.70	Maltese†	0.08	0.68
Dutch$^{\Delta}$	0.06	0.68	Polish$^{\Delta}$	0.15	0.75
English$^{\Delta}$	0.05	0.71	Portuguese$^{\Delta}$	0.08	0.70
Estonian$^{\bullet}$	0.16	0.66	Romanian$^{\bullet}$	0.07	0.70
Finnish$^{\Delta}$	0.18	0.63	Russian$^{\Delta}$	0.14	0.72
French$^{\Delta}$	0.07	0.67	Slovak$^{\Delta}$	0.12	0.77
Georgian$^{\Delta}$	0.18	0.73	Slovenian$^{\Delta}$	0.11	0.69
German$^{\Delta}$	0.08	0.69	Spanish$^{\Delta}$	0.08	0.65
Hebrew$^{\Delta}$	0.17	0.76	Swedish$^{\Delta}$	0.07	0.71
Hungarian$^{\Delta}$	0.17	0.75	Thai$^{\bullet}$	0.01	0.74
Icelandic$^{\bullet}$	0.09	0.70	Turkish$^{\bullet}$	0.18	0.65
Indonesian$^{\bullet}$	0.05	0.62	Zulu†	0.24	0.61

Note. (•) PISA only, (†) PIRLS only**,** (Δ) PIRLS & PISA. Values taken from Gutierrez-Vasques et al. (2019)

Bilingualism

While neither the IEA, nor the OECD, collect detailed information on bilingualism or multilingualism amongst the students sitting PISA or PIRLS assessments, they both collect information that identifies whether their home language matches the language of the test:

- PISA gives only two possible answers (I speak Language of the test / Other language at home) – Item ST022Q01TA
- PIRLS gives four possible answers (I always / almost always / sometimes / never speak <language of test> at home) – Item ASBG03

IDB Analyzer was used to prepare SPSS files for both PIRLS and PISA data. For each file, cases were selected according to country and language of test, such that

only the English language test results for the countries listed in Table 3 were analysed (see Sampling in Chapter 3 for further details).

Method of Data Analysis

Preliminaries - Sample Weights and Plausible Values

When conducting analyses of International Large-Scale Assessments (ILSAs) datasets such as PISA or PIRLS, two important features must be kept in mind. The first is that the stratified two-stage sample design of PISA gives rise to the need to employ sample weights to ensure that the "representation of the population is guaranteed and the estimations are precise. […] Analysis without considering weights would mislead the estimations related to the population" (Arikan et al., 2020, pp 46-47). The second feature to keep in mind is that the aim of ILSA datasets is to produce *population* performance estimates, rather than focusing on the performances of *individual* students. Results are therefore given in the form of Plausible Values, but these cannot be analysed directly, as if they were individual test scores. As the OECD make clear:

> It cannot be emphasised too strongly that the plausible values are not a substitute for test scores for individuals. Plausible values incorporate responses to test items and information about the background of responses; therefore, they cannot be used to compare individuals.[…] When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics. (OECD, 2017, p.147)

In order to ensure the appropriate analyses were conducted, reading ability ranges were therefore calculated within the IDB Anlayzer software from the International Association for the Evaluation of Educational Achievement (IEA), which is designed to work with sample weights and plausible values. All analyses

were therefore conducted using either the IDB Analyzer or the Statistical Package for the Social Sciences (SPSS).

Approaching the Research Questions

To address the first two questions, Pearson Correlations were used to explore possible connections between distributions of range of reading ability, as found in the PISA (2018) and PIRLS (2016) datasets, and language-complexity variables related to the orthographic and morphological complexities of the test languages.

In this study, the specific measure of reading ability identified as of most interest was the range of reading ability for each country's students, as defined by the difference between the fifth and ninety-fifth percentiles (OECD, 2019b). For those countries where students sit the test in more than one language, a range value was calculated for each language. The language-complexity variables were Orthographic Complexity, as measured by the total number of GPC rules in the Dual Route Cascaded model (Gibson et al., 2001), Orthographic Transparency, as measured by OTEANN (Marjou, 2021) and Morphological Complexity and Unpredictability, as measured by Gutierrez-Vasques & Mijangos (2019). As stated earlier, skewness and kurtosis analysis of all variables showed them to fall between ±2 so they can be treated as being normally distributed (see Tables 8 and 9 for descriptive statistics of all variables).

Table 8:

Descriptive Statistics for Variables (Weighted, PISA)

Variable	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	167	189.65	407.08	304.76	37.18	-0.50	1.09
Orth. Complexity	56	59	340	179.89	84.34	0.52	-0.76
Orth. Transparency	87	31.1	99.4	74.17	26.35	-0.82	-0.96

Variable	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Morph. Complexity	142	0.01	0.27	0.10	0.05	0.79	0.46
Morph. Uncertainty	142	0.61	0.91	0.72	0.06	1.30	3.190

Table 9

Descriptive Statistics for Variables (Weighted, PIRLS)

Variable	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	125	169.63	351.24	249.43	45.04	0.75	-0.56
Orth. Complexity	44	59	340	204.39	86.38	0.08	-0.92
Orth. Transparency	65	31.1	99.4	68.53	27.93	-0.42	-1.55
Morph. Complexity	98	0.05	0.24	0.10	0.05	0.64	-0.86
Morph. Uncertainty	98	0.61	0.83	0.71	0.05	0.60	0.45

To ensure the validity of Pearson Correlations, it was also necessary to ensure that the range variable remained normally distributed once it was organized by language-complexity variable, and when split by number of languages spoken. See Tables 10, 11 and 12 for further descriptive statistics.

Table 10

Descriptive Statistics for Range of Reading Ability when Split by Language for Orthographic Complexity, As Measured by DRC

Language	DRC Complexity	*M*	Median	*N*	*SD*	Min.	Max.	Kurtosis	Skewness
PISA (2018)									
Italian	59	305.68	305.68	3	16.15	289.53	321.83	.	0.00
Russian	101	288.48	289.50	13	18.06	235.11	308.80	6.94	-2.29
Dutch	104	341.81	341.81	3	0.60	341.21	342.41	.	0.00
Greek	118	322.06	322.06	2	0.00	322.06	322.06	.	.
German	130	330.19	330.19	7	17.11	304.01	347.34	-1.40	-0.42
English	226	337.19	337.19	21	37.44	261.22	407.08	-0.21	-0.17
French	340	336.43	331.20	7	20.40	316.34	371.32	-0.17	0.92
Total		322.93	322.95	56	330.38	235.11	407.08	0.31	0.10
PIRLS 2016									
Italian	59	215.53	215.53	2	0.00	215.53	215.53	.	.
Russian	101	211.30	209.17	7	17.07	194.04	244.96	2.37	1.43

Language	DRC Complexity	*M*	Median	*N*	*SD*	Min.	Max.	Kurtosis	Skewness
Dutch	104	198.99	198.99	3	0.84	198.14	199.83	.	0.00
German	130	235.13	235.13	3	21.54	213.58	256.67	.	0.00
English	226	285.29	283.43	21	34.15	242.47	351.24	-0.92	0.54
French	340	226.73	226.89	8	13.54	209.78	256.35	3.86	1.55
Total		250.40	245.08	44	42.89	194.04	351.24	-0.31	0.74

Table 11

Descriptive Statistics for Range of Reading Ability When Split by Language for Orthographic Transparency, As Measured by OTEANN

Language	OTEANN Transp.	*M*	Median	*N*	*SD*	Min.	Max.	Kurtosis	Skewness
PISA (2018)									
English	31.1	337.19	337.19	21	37.44	261.22	407.08	-0.21	-0.17
Dutch	55.7	341.81	341.81	3	0.60	341.21	342.41	.	0.00
Italian	71.6	305.68	305.68	3	16.15	289.53	321.83	.	0.00
German	78.8	330.19	330.19	7	17.11	304.01	347.34	-1.40	-0.42
French	79.6	336.43	331.20	7	20.40	316.34	371.32	-0.17	0.92
Portuguese	82.4	314.49	313.63	4	8.58	304.95	325.76	1.50	0.59
Spanish	85.3	293.40	293.86	10	18.71	267.17	321.20	-0.96	-0.03
Finnish	92.3	328.38	328.38	2	0.00	328.38	328.38	.	.
Turkish	95.9	288.96	288.96	2	0.00	288.96	288.96	.	.
Russian	97.2	288.48	289.50	13	18.06	235.11	308.80	6.94	-2.29
Korean	97.5	340.69	340.69	2	0.00	340.69	340.69	.	.
Serbo-Croat	99.3	281.49	281.49	6	23.03	256.36	317.03	-0.54	0.47
Arabic	99.4	292.12	292.12	7	26.68	243.10	324.06	1.10	-0.95
Total		313.58	312.37	87	32.65	235.11	407.08	0.26	0.27
PIRLS (2016)									
English	31.1	285.29	283.43	21	34.15	242.47	351.24	-0.92	0.54
Dutch	55.7	198.99	198.99	3	0.84	198.14	199.83	.	0.00
Italian	71.6	215.53	215.53	2	0.00	215.53	215.53	.	.
German	78.8	235.13	235.13	3	21.54	213.58	256.67	.	0.00
French	79.6	226.73	226.89	8	13.54	209.78	256.35	3.86	1.55
Portuguese	82.4	192.67	192.67	3	23.05	169.63	215.72	.	0.00
Spanish	85.3	228.23	219.62	6	29.34	196.61	269.40	-1.63	0.57
Finnish	92.3	218.84	218.84	2	0.00	218.84	218.84	.	.
Russian	97.2	211.30	209.17	7	17.07	194.04	244.96	2.37	1.43
Arabic	99.4	323.26	321.63	10	18.38	286.73	347.55	0.52	-0.46
Total		255.92	245.87	65	48.98	169.63	351.24	-1.01	0.44

Table 12

Descriptive Statistics for Range of Reading Ability When Split by Language for Morphological Complexity, As Measured by the H3

	Morph. Ent. H3	*M*	Median	*N*	*SD*	Min.	Max.	Kurtosis	Skewness
PISA (2018)									
Albanian	0.72	247.83	247.83	5	27.66	217.40	283.80	-1.79	0.23
Arabic	0.83	292.12	292.12	7	26.68	243.10	324.06	1.10	-0.95
Azerbaijani	0.73	189.65	189.65	2	0.00	189.65	189.65		
Bulgarian	0.68	331.26	331.26	2	0.00	331.26	331.26		
Croatian	0.74	281.49	281.49	6	23.03	256.36	317.03	-0.54	0.47
Czech	0.78	318.82	318.82	2	0.00	318.82	318.82		
English	0.71	337.19	337.19	21	37.44	261.22	407.08	-0.21	-0.17
Estonian	0.66	305.29	305.29	2	0.00	305.29	305.29		
Finnish	0.63	328.38	328.38	2	0.00	328.38	328.38		
French	0.67	336.43	331.20	7	20.40	316.34	371.32	-0.17	0.92
Georgian	0.73	272.65	272.65	2	0.00	272.65	272.65		
German	0.69	330.19	330.19	7	17.11	304.01	347.34	-1.40	-0.42
Hebrew	0.76	369.82	369.82	2	0.00	369.82	369.82		
Hungarian	0.75	309.64	309.64	5	6.47	301.10	319.28	1.98	0.42
Icelandic	0.70	347.27	347.27	2	0.00	347.27	347.27		
Indonesian	0.62	248.52	248.52	2	0.00	248.52	248.52		
Italian	0.61	305.68	305.68	3	16.15	289.53	321.83		0.00
Japanese	0.91	320.29	320.29	2	0.00	320.29	320.29		
Korean	0.91	340.69	340.69	2	0.00	340.69	340.69		
Macedonian	0.70	300.27	300.27	2	0.00	300.27	300.27		
Malayalam	0.70	265.87	265.87	2	0.00	265.87	265.87		
Maltese	0.68	341.81	341.81	3	0.60	341.21	342.41		0.00
Polish	0.75	311.41	320.13	5	14.32	295.78	323.54	-3.29	-0.58
Portuguese	0.70	314.49	313.63	4	8.58	304.95	325.76	1.50	0.59
Romanian	0.70	308.41	303.01	5	8.87	301.58	322.44	0.64	1.30
Russian	0.72	288.48	289.50	13	18.06	235.11	308.80	6.94	-2.29
Slovak	0.77	332.55	332.55	2	0.00	332.55	332.55		
Slovenian	0.69	309.11	309.11	2	0.00	309.11	309.11		
Swedish	0.71	322.65	311.65	5	20.70	307.13	354.34	-0.17	1.13
Thai	0.74	261.52	261.52	2	0.00	261.52	261.52		
Turkish	0.65	292.66	291.58	12	17.02	267.17	321.20	-0.45	0.13
Ukrainian	0.78	309.78	309.78	2	0.00	309.78	309.78		
Total		308.08	309.40	142	35.62	189.65	407.08	1.31	-0.43
PIRLS (2016)									
Arabic	0.83	323.26	321.63	10	18.38	286.73	347.55	0.52	-0.46
Bulgarian	0.68	279.64	279.64	2	0.00	279.64	279.64		
Czech	0.78	220.84	220.84	2	0.00	220.84	220.84		
English	0.71	285.29	283.43	21	34.15	242.47	351.24	-0.92	0.54
Finnish	0.63	218.84	218.84	2	0.00	218.84	218.84		
French	0.67	248.40	227.44	10	47.22	209.78	335.07	0.81	1.53
Georgian	0.73	251.10	251.10	2	0.00	251.10	251.10		

	Morph. Ent. H3	*M*	Median	*N*	*SD*	Min.	Max.	Kurtosis	Skewness
German	0.69	235.13	235.13	3	21.54	213.58	256.67		0.00
Hebrew	0.76	245.47	245.47	2	0.00	245.47	245.47		
Hungarian	0.75	289.16	289.16	3	42.23	246.93	331.39		0.00
Italian	0.61	215.53	215.53	2	0.00	215.53	215.53		
Maltese	0.68	236.86	199.83	5	51.87	198.14	293.67	-3.33	0.61
Polish	0.75	217.32	210.04	5	14.00	206.74	238.70	-0.21	1.12
Portuguese	0.70	192.67	192.67	3	23.05	169.63	215.72		0.00
Romanian	0.70	224.70	224.70	2	0.00	224.70	224.70		
Russian	0.72	211.30	209.17	7	17.07	194.04	244.96	2.37	1.43
Slovak	0.77	255.97	255.97	2	0.00	255.97	255.97		
Slovenian	0.69	237.94	237.94	2	0.00	237.94	237.94		
Swedish	0.71	218.02	222.05	5	14.43	197.25	230.57	-1.05	-0.77
Turkish	0.65	228.23	219.62	6	29.34	196.61	269.40	-1.63	0.57
Zulu	0.61	251.62	251.62	2	0.00	251.62	251.62		
Total		253.79	245.47	98	45.19	169.63	351.24	-0.69	0.59

To address the third question, the different approaches taken by PISA 2018 and PIRLS 2016 to the question of what language(s) the participants spoke at home necessitated different analytical approaches. Item ST022Q01TA of PISA 2018 gave participants two possible answers (I speak language of the test / another language at home), meaning that an Independent Samples T-Test could be used to identify any difference between the groups. Item ASBG03 of PIRLS 2016 gave participants four possible answers (I always / almost always / sometimes / never speak <language of test> at home). Therefore, a one-way ANOVA was conducted to determine whether there were any significant differences in the ranges of reading ability between the groups.

CHAPTER 4: RESULTS

Introduction

This study used correlational analyses to investigated potential relationships between language-complexity factors (orthographic complexity, orthographic transparency, morphological complexity and morphological unpredictability) and patterns of reading ability, as measured by the range of reading ability for countries represented in the PISA 2018 and PIRLS 2016 datasets. Two specific questions were addressed:

1. Are the distribution patterns of reading ability range in the data from the PISA 2018 and PIRLS 2016 cycles associated with orthographic depth?
2. Are the distribution patterns of reading ability range in the data from the PISA 2018 and PIRLS 2016 cycles associated with morphological complexity?
3. Are there any differences in the distribution patterns of reading ability in the PISA 2018 and PIRLS cycles based on / caused by anglophone monolingual and bilingual contexts?

Orthographic Depth and Range of Reading Ability

Orthographic Complexity

This section reports the relationships between range of reading ability and Orthographic Complexity, as measured by the total number of Grapheme-Phoneme Correspondence Rules (GPCRs) from the Dual Route Cascaded model (Ziegler et al., 2000). As described in Chapter 3, the data from the PISA (2018) and PIRLS (2016) datasets were assessed using Pearson correlational analyses.

PISA (2018) Dataset. There was a positive Pearson correlation between Range (5% - 95%) and DRC Complexity ($r(54) = .416$, $p < .001$), indicating a moderate positive relationship between two variables. See Table 13 for the descriptive statistics. Figure 6 shows the scattergram of this relationship. Explained variance shows that scores share approximately 17% common variance.

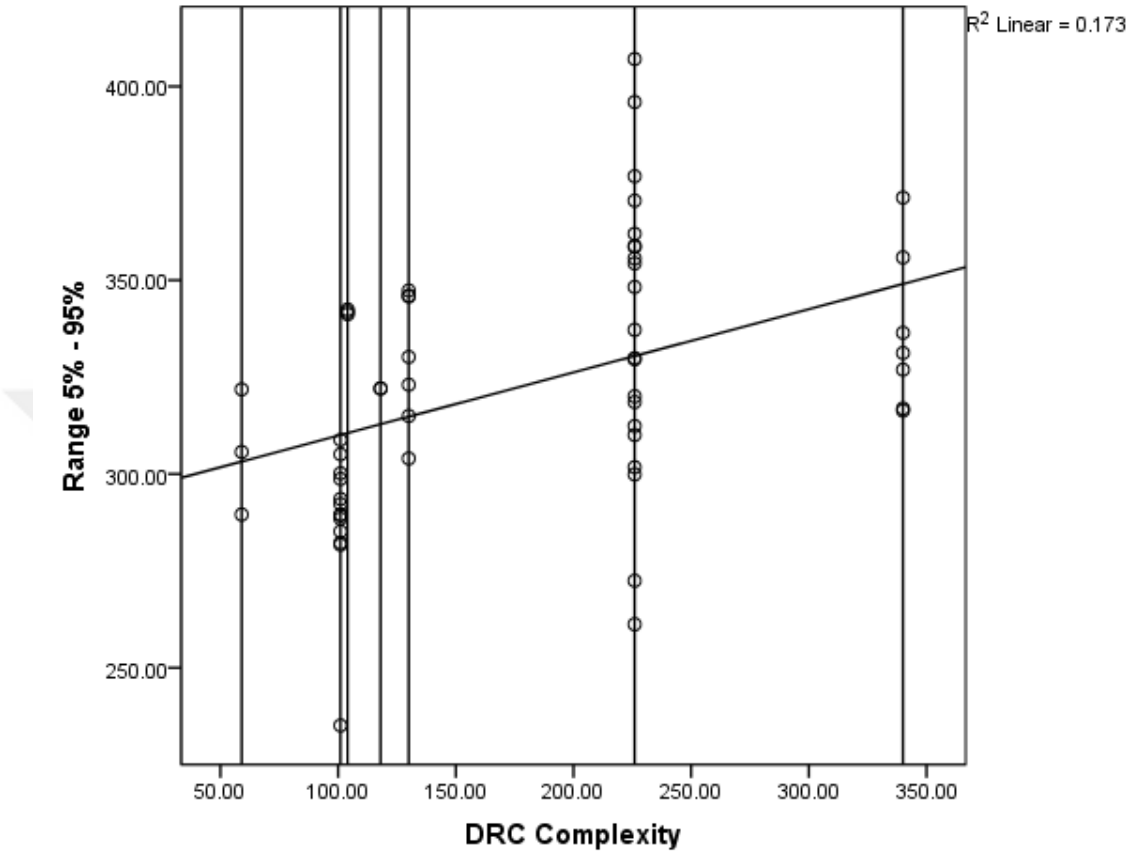
Table 13

Descriptive Statistics for Range of Reading Ability and Orthographic Complexity in PISA (2018)

Variable	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	56	235.11	407.08	322.93	33.04	0.10	0.31
DRC Complexity	56	59	340	179.89	84.34	0.52	-0.76

Figure 6

Scattergram Showing the Shared Variance Between Range of Reading Ability and Orthographic Complexity (DRC) in PISA (2018)



PIRLS (2016) Dataset. The same relationship calculated based on the PIRLS 2016 dataset was statistically significant but weaker (r(42) = .298, p = .049). Descriptive statistics are shown in Table 14. Common variance between the variables is almost 9%, as shown on the scatterplot of the variables in Figure 7.

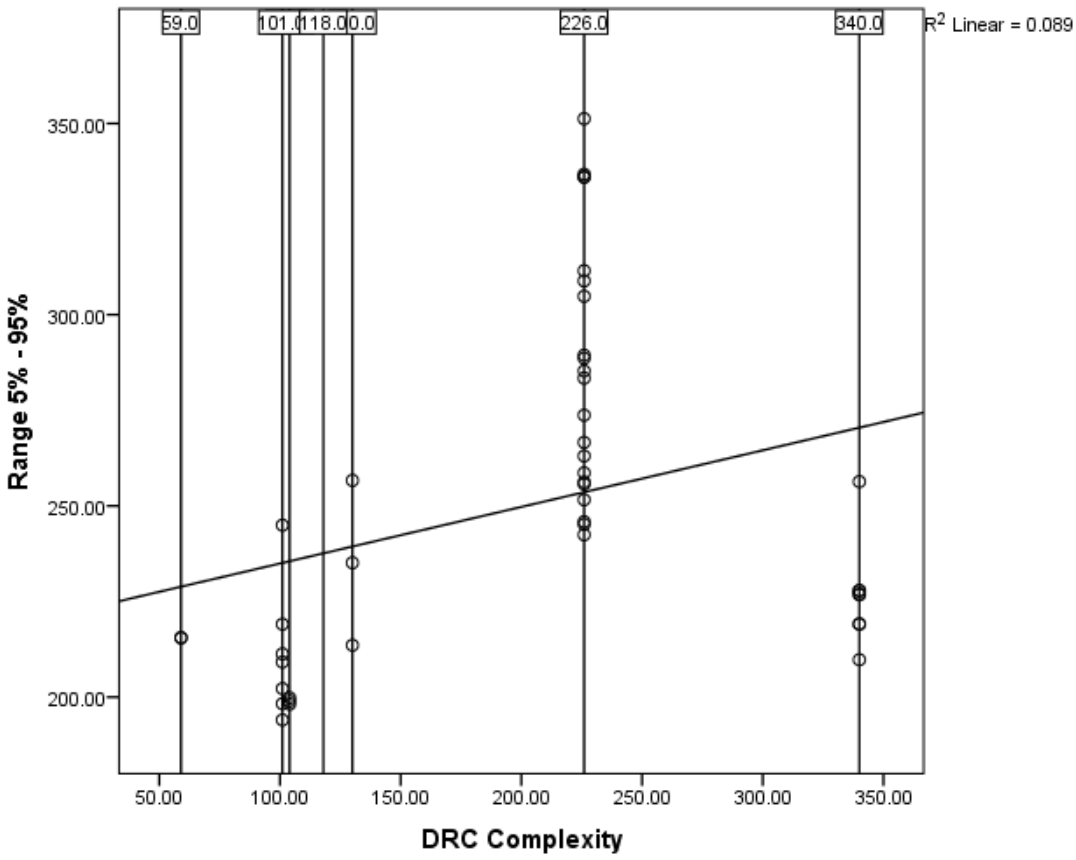
Table 14

Descriptive Statistics for Range and Orthographic Complexity PIRLS (2016)

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	44	194.04	351.24	250.40	42.89	0.74	-0.31
DRC Complexity	44	59	340	204.39	86.38	0.08	-0.92

Figure 7

Scattergram Showing the Shared Variance Between Range of Reading Ability and Orthographic Complexity (DRC) in PIRLS (2016)



Orthographic Transparency

This section reports the relationships between range of reading ability and Orthographic Transparency, as measured by Orthographic Transparency Estimation with an Artificial Neural Network (OTEANN) (Marjou, 2021).

PISA (2018) Dataset. There was a negative correlation between Range (5% - 95%) and Orthographic Transparency ($r(85) = -.528$, $p < .001$). That is, as the orthography becomes more transparent, the magnitude of the range of reading ability reduces. See Table 15 for descriptive statistics). OTEANN and reading share a common variance of 28%, as shown in the scattergram in Figure 8.

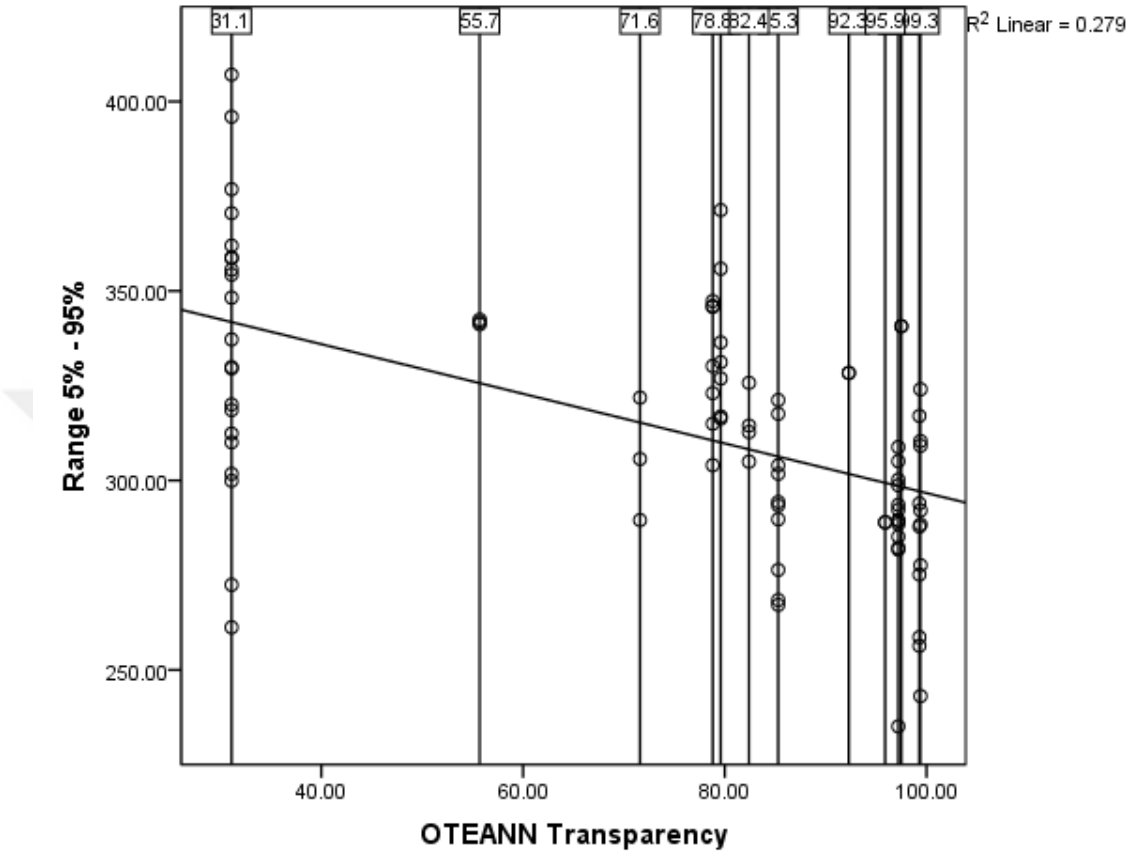
Table 15

Descriptive Statistics for Range and Orthographic Transparency PISA (2018)

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	87	235.11	407.08	313.58	32.65	0.27	0.26
OTEANN Transparency	87	31.1	99.4	74.17	26.35	-0.82	-0.96

Figure 8

Scattergram Showing the Shared Variance Between Range of Reading Ability and Orthographic Transparency (OTEANN) in PISA (2018)



PIRLS (2016) Dataset. PIRLS data present a much weaker, insignificant relationship (r(63) = -.225, p = .071) between range of reading ability and OTEANN (see Table 16 for descriptive statistics), indicating a smaller share of common variances (5%), as shown in the scattergram in Figure 9.

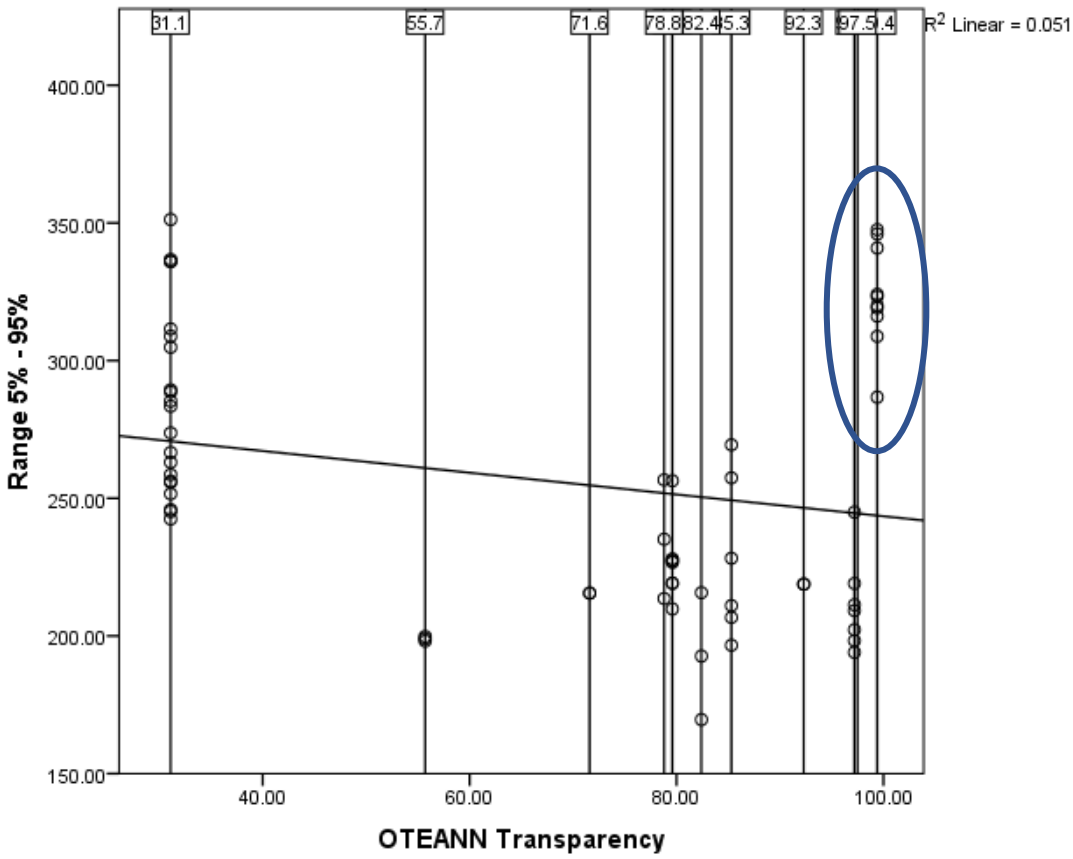
Table 16

Descriptive Statistics for Range and Orthographic Transparency PIRLS (2016)

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	65	169.63	351.24	255.92	48.98	0.44	-1.01
OTEANN Transparency	65	31.1	99.4	68.53	27.93	-0.42	-1.55

Figure 9

Scattergram Showing the Shared Variance Between Range of Reading Ability and Orthographic Transparency (OTEANN) in PIRLS (2016)



However, upon examination of the scatterplot, it appeared that the results for Arabic-speaking countries seemed to fall outside of the pattern for the rest of the countries (circled in Figure 9). In private correspondence, the IEA confirmed that all countries (including Arabic speaking countries) were responsible for producing their own copies of the test materials, and not all Arabic materials had been printed with diacritics. As the OTEANN scores were calculated using pointed Arabic, it would therefore seem to be reasonable to exclude results from Arabic speaking countries in the analysis. Following their removal (see Table 17 for descriptive statistics), the relationship between range of reading ability and OTEANN gets significantly stronger ($r(53) = -.757$, $p < .001$), indicating a common variance of 52% between the two variables, as shown in the scattergram in Figure 10.

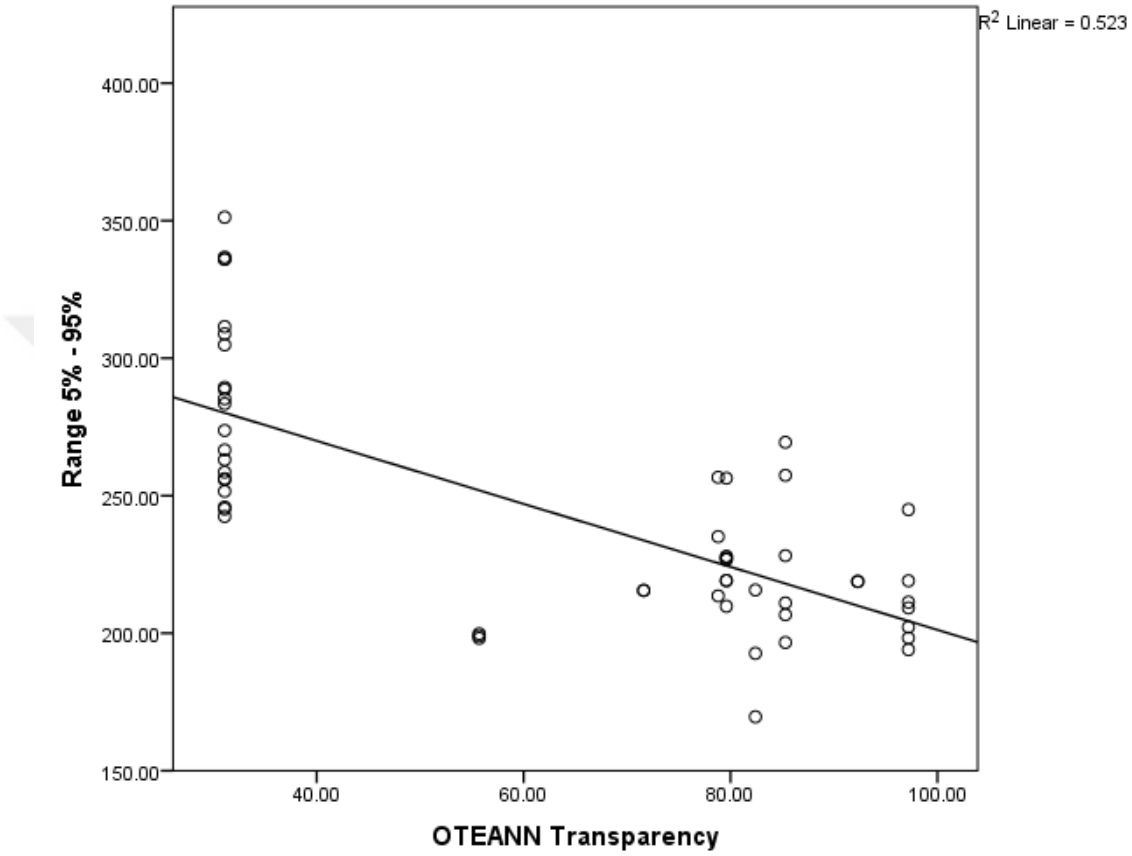
Table 17

Descriptive Statistics for Range and Orthographic Transparency PIRLS (2016) with Arabic-Speaking Countries Removed

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	55	169.63	351.24	243.68	42.36	0.82	0.05
OTEANN Transparency	55	31.1	97.2	62.92	26.75	-0.21	-1.74

Figure 10

Scattergram Showing the Shared Variance Between Range of Reading Ability and Orthographic Transparency in PIRLS (2016) with Arabic-Speaking Countries Removed



Morphology and Range of Reading Ability

Morphological Complexity (TTR)

This section reports the relationships between range of reading ability and Morphological Complexity, as calculated by Gutierrez-Vasques et al. (2019*)* using a Type/Token Ratio (TTR).

PISA (2018) Dataset. Range of reading ability and Morphological Complexity were found to be lightly negatively correlated ($r(140) = -.25$, $p = .002$) sharing a common variance of 6.5%. See Table 18 for descriptive statistics and Figure 11 for a scattergram showing shared variance.

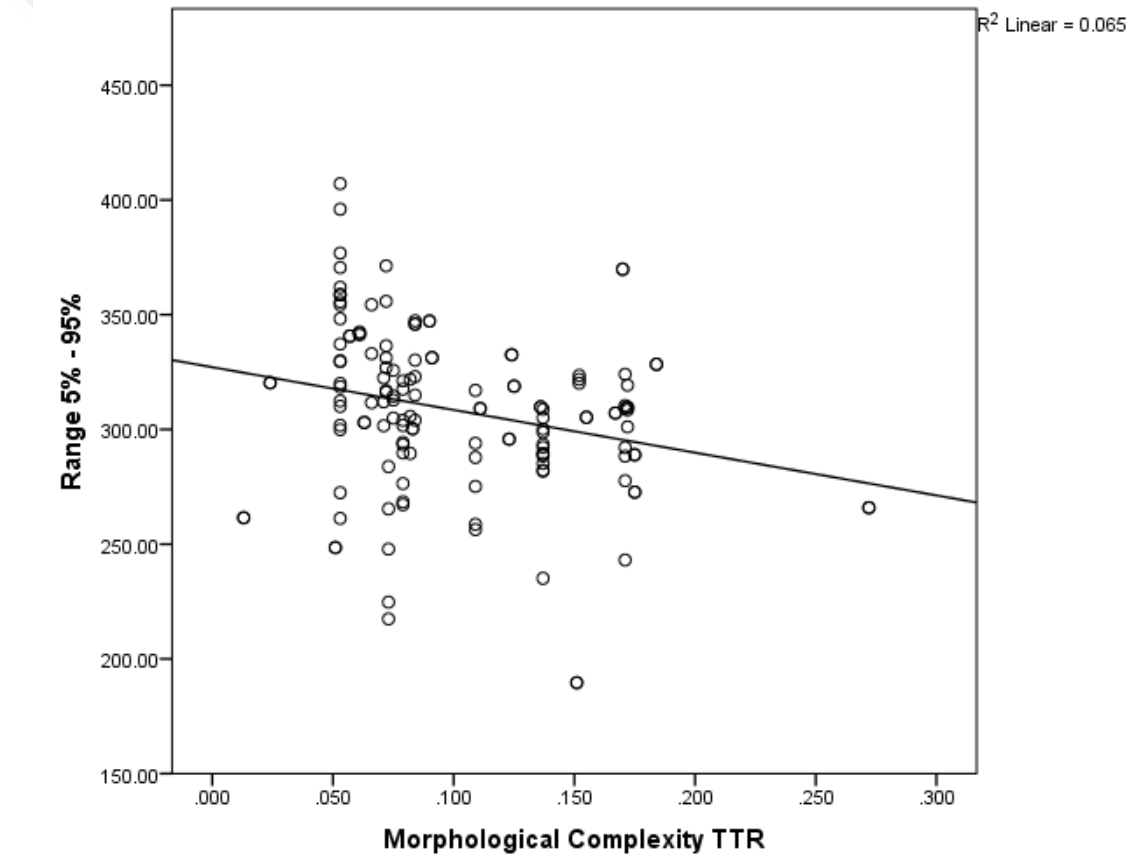
Table 18

Descriptive Statistics for Range and Morphological Complexity in PISA (2018)

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	142	189.65	407.08	308.08	35.62	-0.43	1.31
Morphological Complexity TTR	142	0.013	0.27	0.10	0.05	0.79	0.46

Figure 11

Scattergram Showing the Shared Variance Between Range of Reading Ability and Morphological Complexity in PISA (2018)



PIRLS (2016) Dataset. After excluding Arabic speaking countries (as for the previous PIRLS analysis), range of reading ability and Morphological Complexity were found to be moderately negatively correlated ($r(86) = -.23$, $p = .029$) sharing a common variance of 5.4%. See Table 19 for descriptive statistics and Figure 12 for a scattergram showing shared variance.

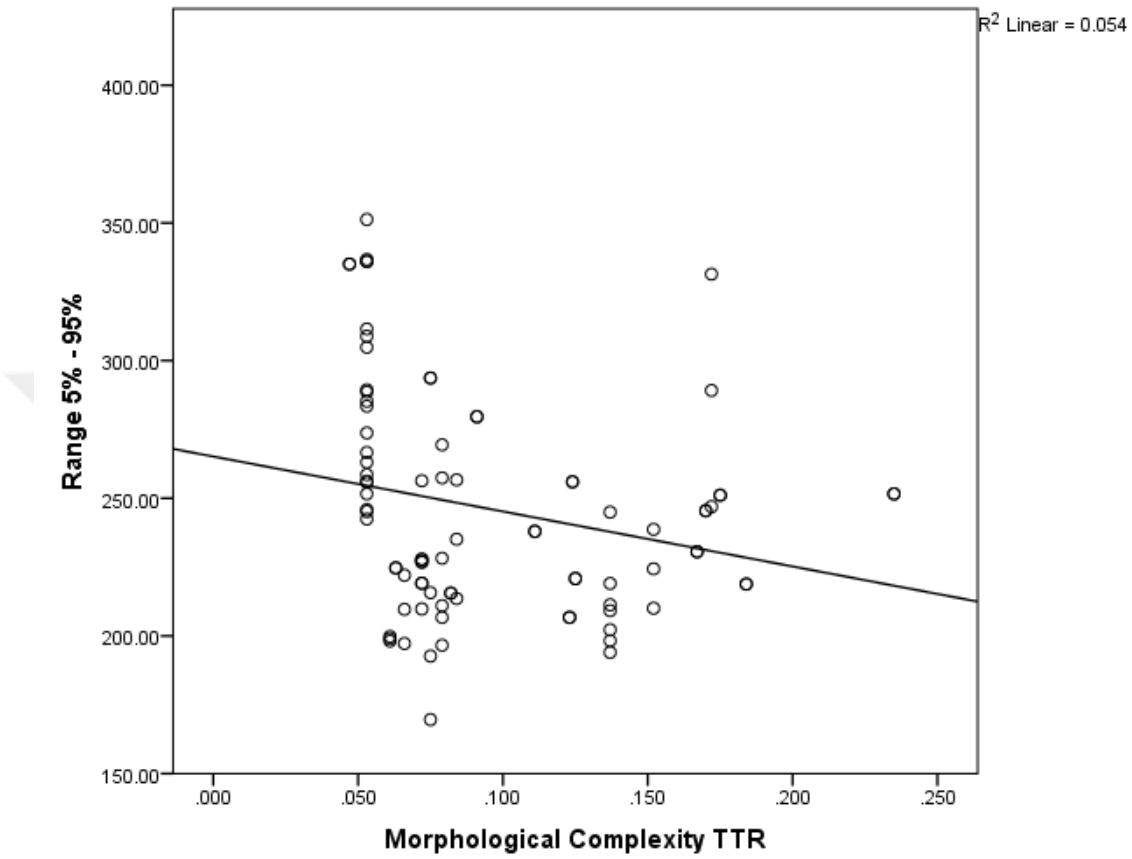
Table 19

Descriptive Statistics for Range and Morphological Complexity in PIRLS (2016) with Arabic-Speaking Countries Removed

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	88	169.63	351.24	245.89	40.30	0.83	0.11
Morphological Complexity TTR	88	0.05	0.24	0.10	0.05	1.02	0.11

Figure 12

Scattergram Showing the Shared Variance Between Range of Reading Ability and Morphological Complexity in PIRLS (2016) with Arabic-Speaking Countries Removed



Morphological Unpredictability (H3)

This section reports the relationships between range of reading ability and Morphological Unpredictability. Gutierrez-Vasques et al. (2019*)* calculated transition probabilities between consecutive sub-word units, three characters long (H3).

PISA (2018) Dataset. Range of reading ability and Morphological Unpredictability were uncorrelated (r(140) = .01, p = .903). See Table 20 for descriptive statistics and Figure 13 for a scattergram showing shared variance.

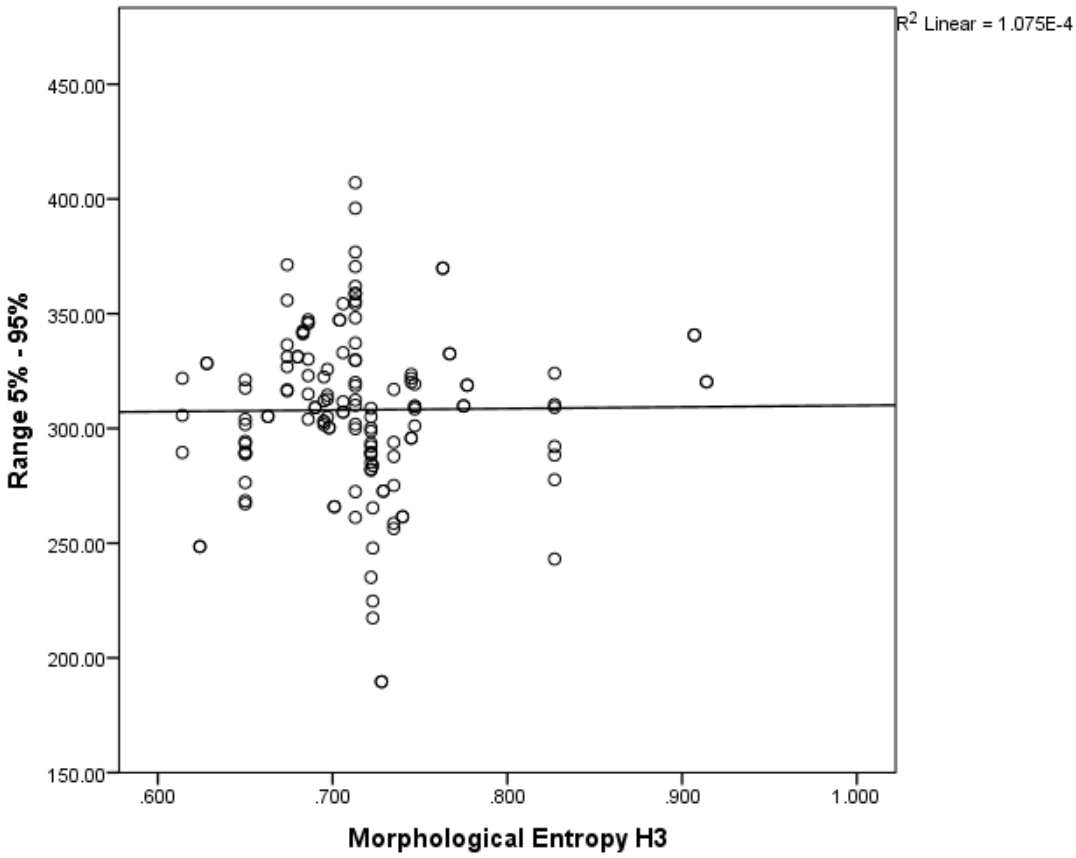
Table 20

Descriptive Statistics for Range and Morphological Unpredictability in PISA (2018)

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	142	189.65	407.08	308.08	35.62	-0.43	1.31
Morphological Unpredictability H3	142	0.61	0.91	0.72	0.06	1.30	3.19

Figure 13

Scattergram Showing the Absence of Shared Variance Between Range of Reading Ability and Morphological Unpredictability in PISA (2018)



PIRLS (2016) Dataset. When disregarding results from Arabic test scores, range of reading ability and Morphological Unpredictability were also not significantly correlated in the PIRLS (2016) dataset ($r(86) = .118$, $p = .273$). See table 21 for descriptive statistics and figure 14 for a scattergram showing lack of shared variance.

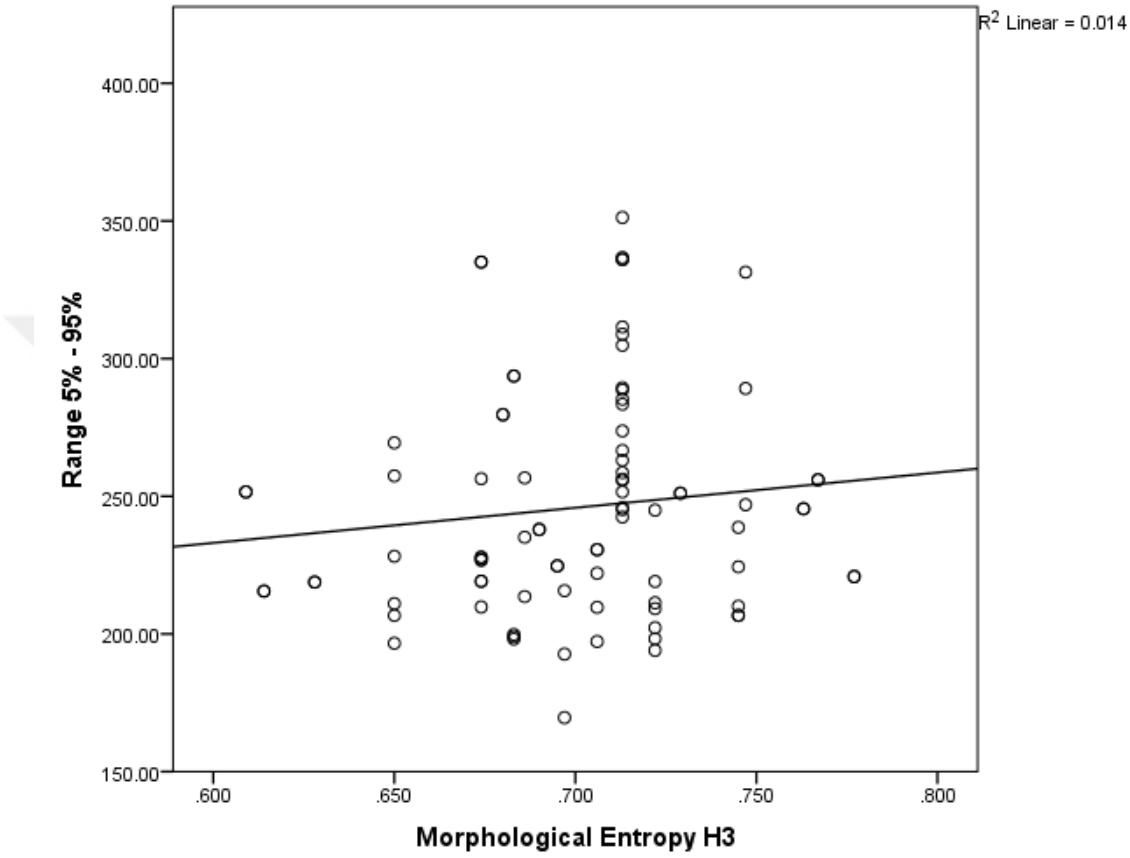
Table 21

Descriptive Statistics for Range and Morphological Unpredictability in PIRLS (2016) ***Excluding*** *Results from Arabic-Speaking Countries*

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	88	169.63	351.24	245.89	40.30	0.83	0.11
Morphological Unpredictability H3	88	0.61	0.78	0.70	0.04	-0.38	0.25

Figure 14

Scattergram Showing the Absence of Shared Variance Between Range of Reading Ability and Morphological Unpredictability in PIRLS (2016) Excluding Results from Arabic-Speaking Countries



Surprisingly, if the Arabic language tests were included in the analysis, the results changed dramatically, leading to a significant, moderate correlation ($r(96) = .451$, $p < .001$), with a shared common variance of 20.4%. See Table 22 for descriptive statistics and Figure 15 for a scattergram showing lack of shared variance.

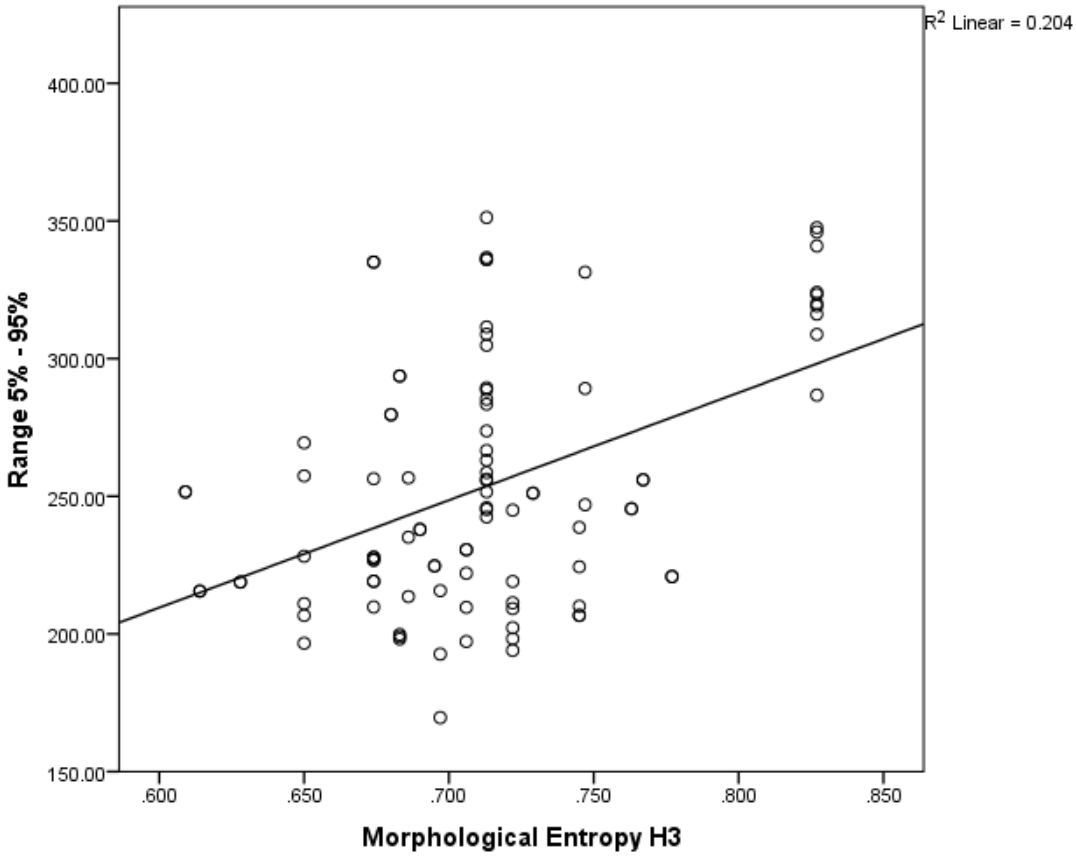
Table 22

Descriptive Statistics for Range and Morphological Unpredictability in PIRLS (2016) ***Including*** *Results from Arabic-Speaking Countries*

Statistics	*N*	Minimum	Maximum	*M*	*SD*	Skewness	Kurtosis
Range 5% - 95%	98	169.63	351.24	253.79	45.19	0.59	-0.69
Morphological Entropy H3	98	0.61	0.83	0.71	0.05	0.60	0.45

Figure 15

Scattergram Showing the Absence of Shared Variance Between Range of Reading Ability and Morphological Unpredictability in PIRLS (2016) Including Results from Arabic-Speaking Countries



Bilingualism and Range of Reading Ability

This section reports the relationships between range of reading ability and the home language of students in countries whose main language of instruction is English.

PISA (2018) Dataset. When looking at the differences in range of reading ability between the two groups, students who spoke a different language at home showed a higher ranger of reading ability in all countries (see Figure 16). As a result, the mean range of reading ability was lower for students who spoke English at home than for students who spoke a different language (see Table 23) but overall effect of home language on range of reading ability did not reach significance, $t(12) = -1.24$, $p = .24$).

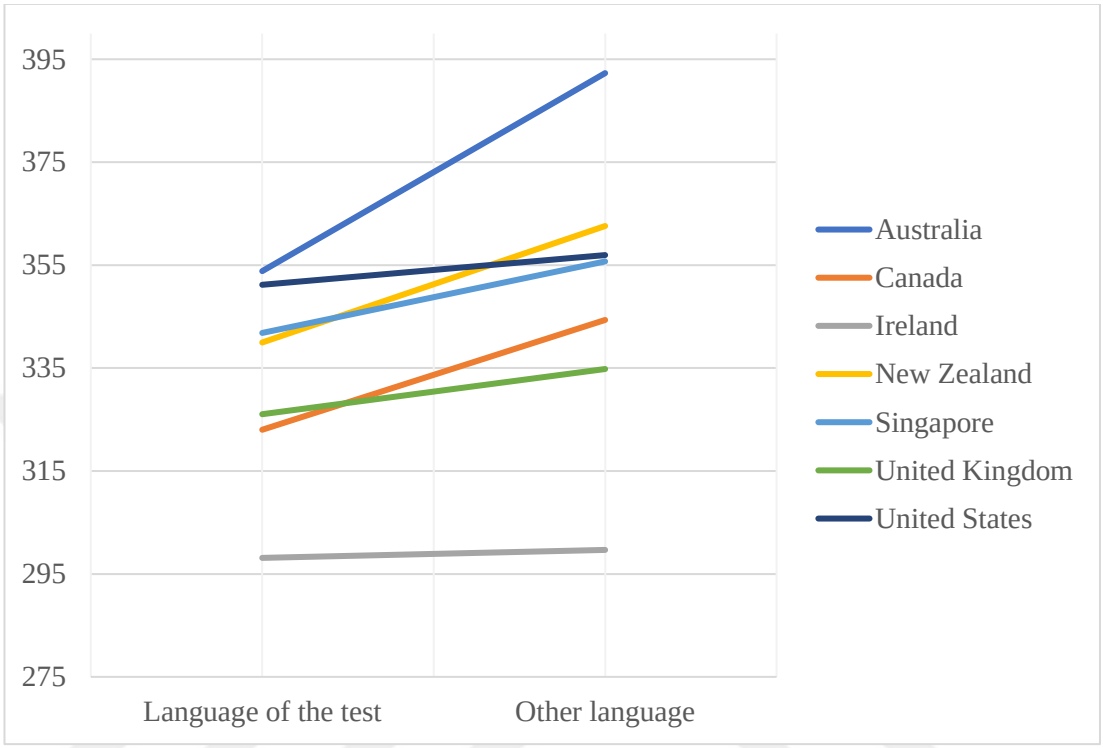
Table 23

Descriptive Statistics Showing Mean Ranges Depending on What Language the Students Speak at Home in PISA (2018)

What language do you speak at home most of the time?		*N*	*M*	*SD*	Std. Error Mean
Range	Language of the test	7	333.43	19.39	7.33
	Other language	7	349.49	28.35	10.72

Figure 16

Line Graph Showing Range of Reading Ability by Country and Home Language, PISA (2018)

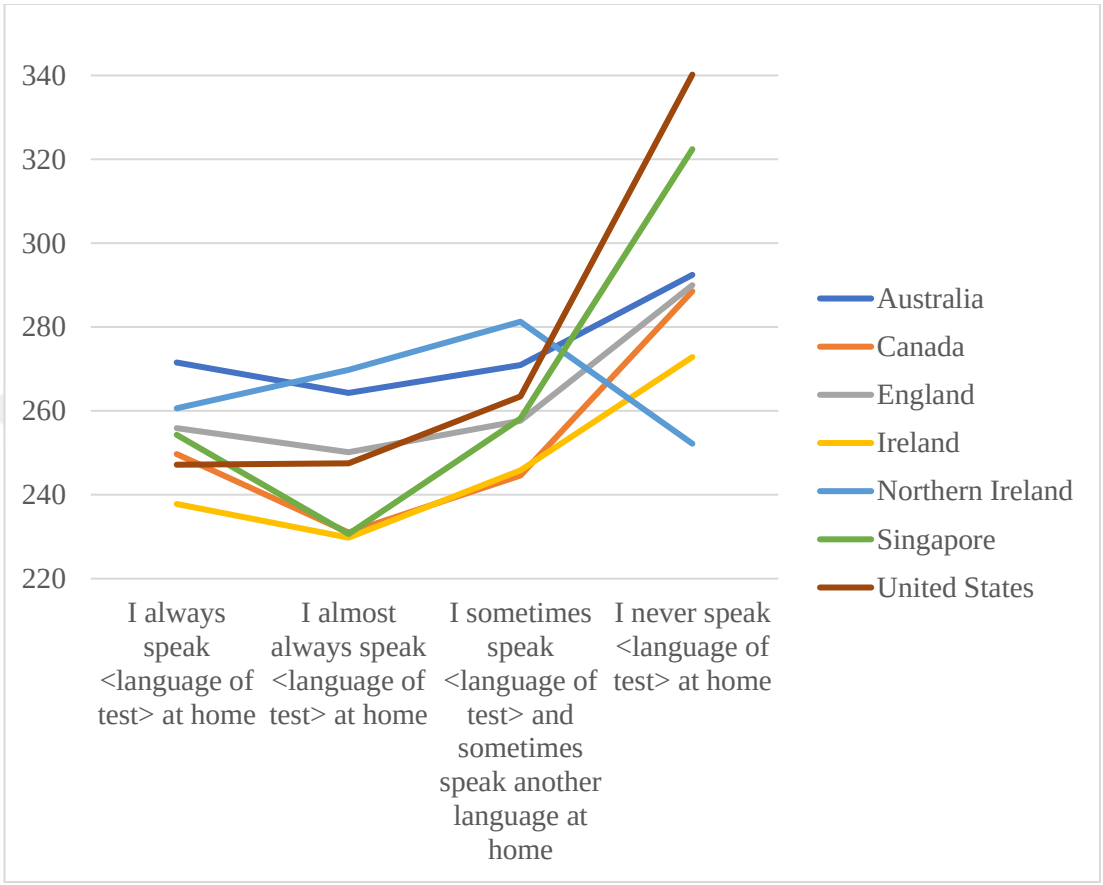


PIRLS (2016) Dataset. As previously described, PIRLS (2016) gave students four possible answers to the question of what language they spoke at home. The mean range of reading ability of students who *always* spoke English at home was 253.85 (SD = 9.87); for students who *almost always* spoke English at home, it was 246.17 (SD = 15.32); for students who *sometimes* spoke English at home, it was 260.28 (SD = 12.12); and for students who *never* spoke English at home, it was 294.09 (SD = 27.22). Using Welch's robust test of Equality of Means showed that a significant effect of home language on range of reading ability, $F(3, 14.998) = 6.196$, $p = .006$. Tukey's HSD test for multiple comparisons demonstrated that the mean ranges of reading ability for the first three groups were not significantly different from each other, but the range of reading ability for students who *never* speak English at home was significantly higher than for all other conditions (*always* $p < .001$, 95% CI = [16.40, 64.08]; *almost always* $p < .001$, 95% CI = [24.08, 71.76]; *sometimes* $p = .003$, 95% CI = [9.97, 57.65]). Figure 17 illustrates how reading ability ranges differed between groups for each country.

Figure 17

Line Graph Showing Range of Reading Ability by Country and Home Language

PIRLS (2016)



CHAPTER 5: DISCUSSION

Introduction

The initial purpose of this study is to examine potential correlations between certain language-complexity variables associated with the orthographies and morphologies of different languages and the range of reading abilities in those languages, as found in the PISA (2018) and PIRLS (2016) datasets. This study also examined the potential impact of speaking more than one language on the range of reading ability of children in countries whose main language of instruction is English. This study found that two aspects of orthography (its complexity and its transparency) are associated with an increase in the range of reading ability. That is, in more complex or less transparent orthographies, the poorest readers lag further behind in reading ability from the strongest readers than for simpler orthographies. In contrast, this study found that only the complexity of a language's morphology is associated with an increase in the range of reading ability; its unpredictability did not initially show any correlation with range of reading ability. The final part of this study found that speaking a different language at home was not associated with a difference in the range of reading ability in the PISA (2018) dataset, but it was in the PIRLS (2016) dataset. That is, in the younger children, those children who never spoke English at home showed a higher range of reading ability.

Overview of the Study

The aim of this study is to make use of from the PISA (2018) and PIRLS (2016) assessments to investigate possible connections between certain language-complexity variables and population patterns of reading ability distribution. In

particular, this study is concerned with differences in the orthographies and morphologies of different languages, as well as whether students speak more than one language. With regard to the orthographic and morphological structures of languages, four separate characterizing variables were identified: Orthographic Complexity (how many Grapheme-Phoneme Correspondence Rules a language uses); Orthographic Transparency (what proportion of words in a language can be correctly pronounced using the GPCRs); Morphological Complexity (a measure of the ratio of Types to Tokens in the language), and; Morphological Uncertainty (a measure of how easy it is to predict upcoming letters based on previous letter-strings). With regard to the impact of monolingualism vs multilingualism, students from countries whose main language of instruction is English were split into groups according to the language they spoke at home. For PISA (2018), students were split into two groups (those who spoke English at home and those who did not.) For PIRLS (2016), students were split into four groups (those who always / almost always / sometimes / never speak English at home).

This study found that orthographic and morphological factors were associated with different distribution patterns of range of reading ability at different age levels. Orthographic Complexity was positively correlated with range of reading ability for both the PISA (2018) and PIRLS (2016) datasets, although the correlation was stronger for the older children. Orthographic Transparency was negatively correlated with range of reading ability for both datasets, and the effect size was strongest for the younger children. Morphological Complexity was found to be lightly negatively correlated with range of reading ability in both the PISA (2018) and PIRLS (2016) datasets, with both ages exhibiting a similar level of shared variance. Morphological Uncertainty did not initially show any correlation with range of reading ability in

either of the datasets. However, when the data from Arabic speaking countries was included in the analysis, the correlation became significant and strong in the younger children.

With regard to the second aim, this study found that speaking a different language at home was not associated with a significant difference in the range of reading ability in the PISA (2018) dataset, but it was in the PIRLS (2016) dataset.

Discussion of Major Findings

Reading has always been recognized as one of the most important skills that children need to master by the time that they finish school. National analyses of career progression and lifetime earning potential show not only that a minimum level of literacy is needed to access much of the job market, but higher levels of literacy are strongly associated with higher lifetime earning potential (Gibson et al., 2019; Sum, 1999). As such, politicians and education ministries are interested in being able to measure levels of reading ability in their populations, as well as monitoring how those levels change over time in response to any changes they may implement. In order to serve this demand, a variety of ILSAs have been developed. Two highly regarded such ILSAs are the OECD's Program for International Student Assessment (PISA) and the IEA's Progress in International Literacy Study (PIRLS). PISA covers a wide range of domains, including literacy, and is sat by students aged fifteen, while PIRLS is purely focused on reading ability and is sat by students in the fourth grade.

As it is important for educational organizations to be able to understand the factors affecting the outcomes of the students whose education they are responsible for, the results of PIRLS and PISA assessments are analysed in great detail at both national and international levels. To support those analyses, both PISA and PIRLS collect a wide variety of supplementary information about the students who sit the

assessments, including data covering home life factors, motivational factors and school-based factors. At the same time, there is a significant amount of research surrounding the impact of language-complexity factors, including (but not limited to) details of languages' orthographies and morphologies, showing that these factors do play a role in the ease with which students are able to learn to read different language (Frith et al., 1998; Katz & Frost, 1992; Knight et al., 2017; Landerl et al., 1997; Rastle, 2019). Some researchers are therefore starting to comment that it is surprising that the results of ILSAs such as PIRLS and PISA do not routinely factor in the impact of differences in orthographies, morphologies or any other language-complexity factors (Knight et al., 2019).

The current study took as its starting point the observation that while many anglophone countries have been able to raise the overall reading achievement level of their students, a surprisingly persistent long tail of lower-performing students can be seen across these countries. This is not necessarily seen in the results of other countries with different languages of instruction. That is, the disparity in reading ability between the lowest ability and highest ability readers seems to be anomalously high in anglophone countries. Could this indicate that there is something unique to English that makes it especially difficult for those readers at the bottom of the distribution to become good readers?

The current study approached this apparent anomaly by identifying a variety of metrics pertaining to the orthographies and morphologies of different languages and analysing how they correlate with population-level ranges of reading ability. Analysing the results of both PISA and PIRLS assessments allowed for a view of how the different variables might be related to range of reading ability at different ages, and therefore at different stages of reading development.

Findings for Research Question 1 (Orthography)

Orthographic Complexity

The results for ten-year-old (PIRLS) and fifteen-year-old (PISA) children both showed significant, positive correlations between the total number of GPCRs and the range of reading ability. The results for the older children showed a stronger correlation than the younger children, with shared variances of 17% and 9% respectively. That is, as the spelling complexity of a language increases, the poorest readers in a country tend to fall further behind the strongest readers, and this pattern is more visible amongst the older readers. The fact that the correlation is stronger for older readers than younger readers is surprising, as it might be expected that the extra years of reading tuition would have helped them to master the higher number of GPCRs and so the correlation would be less strong among older children. Potentially, this finding might be connected to the reading-level of the passages included in the two ILSAs. It is conceivable that as the vocabulary used in the PISA reading assessment is more advanced than in PIRLS, it would include a greater number or proportion of words that require more complex context-dependent GPCRs. This will be discussed in the section dealing with implications for further research.

Orthographic Transparency

Older children showed a significant negative correlation between how likely a word can be pronounced accurately by using standard GPCRs and the range of reading ability. That is, as a language's orthography becomes less transparent or predictable, the poorest readers fall further behind the strongest readers. Conversely, as a language's orthography becomes more predictable, the poorest readers get closer in reading ability to the strongest readers. For the older children (PISA), Pearson's coefficient of correlation was -.528 ($p < .05$), which is considered a large effect size

in social sciences research – Cohen (1977) asserts that an effect size of r = .50 "falls around the upper end of the range of (nonreliability) r's one encounters in […] educational psychology" (p. 80). For the younger children, the coefficient of correlation was initially low and did not reach significance. This may indicate that the sample size (65) was too low to reach significance when taking into account the effect size. However, upon examining the scatter diagram of the results, another possibility became apparent, as the results for one language (Arabic) did not seem to match the overall trend. After communicating with the IEA, it was found that due to the way that the test materials were produced in each country, the Orthographic Transparency score that had been assigned to Arabic was not necessarily reliable. Upon removing the results from the tests in Arabic, the correlation improved dramatically, reaching -.723, indicating a stronger negative correlation. It is highly unusual to find a correlation of this strength in the fields of educational research or psychology, so although correlational analysis does not allow one to infer causality from the results, it could be a strong indicator that further, experimental research would be fruitful. *If* it were possible to attribute causality, an effect size of r = -.723 would suggest that over 50% of the variance could be attributed to Orthographic Transparency.

The results seem to indicate that as a language's orthography becomes less and less transparent, poorer readers of that language will fall further and further behind their peers. This is true for both readers in only their fourth year of formal education, as well as more mature readers towards the end of their time in school. Not only that, but this effect appears to be more pronounced in the younger readers, which may reflect differences in the reading development of students at these different ages. As hypothesised by Knight et al. (2017), the higher cognitive load

imposed by reading orthographically opaque languages may affect younger readers more noticeably.

Findings for Research Question 2 (Morphology)

Morphological Complexity

Results from both PISA (2018) and PIRLS (2016) showed very similar, moderate but significant negative correlations between the ratio of Types to Tokens in a language's morphology and the range of reading ability ($r = -.254$ and $r = -.233$ respectively). In other words, the larger the number of permutations based on root words that are permitted in a language, the closer the poorest readers come to the reading ability of the highest readers. This could initially seem counter-intuitive, as it might be assumed that as a morphology becomes more complicated, poorer readers would find the task more challenging. However, as discussed earlier, it has been shown that children who learn to read a morphologically complex language become better at reading morphologically complex words than morphologically simple ones in their third year of schooling (Borleffs et al., 2019). Perhaps it is therefore possible that increasing the number of permutations results in additional help when reading new words.

Morphological Unpredictability

Results from both PISA (2018) and PIRLS (2016) initially showed no correlation between the unpredictability of a language's morphology and the range of reading ability. That is, a language whose morphology was more predictable was not associated with any advantage for the lowest ability readers when compared to readers of languages with less predictable morphologies. However, when results from Arabic speaking countries were included in the analysis of the PIRLS (2016) data, a highly significant, moderate-to-strong correlation ($r = .451$) appeared. In other

words, amongst the younger group of students, languages with more predictable morphologies were associated with smaller gaps between the poorest and strongest readers in those languages. There are two possible explanations for these varying results: Either, morphological unpredictability is indeed directly correlated with range of reading ability, or; the range of reading ability is correlated with quality of educational provision, and the ranking of Arabic's morphological unpredictability coincidentally matches the educational provision ranking of the countries that speak it. Further experimentation would be needed to tease apart these two possibilities.

Findings for Research Question 3 (Bilingualism)

Bialystok (2007) identified three routes by which speaking a second (or subsequent) language(s) can influence reading ability: Metalinguistic Competence, Concept of Print, and, Oral Language Skills. As discussed earlier, by the ages at which students sit the PIRLS and PISA tests, the influence of metalinguistic competence (phonological awareness) has fallen away. Bialystok identified some areas in which some aspects of concepts of print (essentially, knowledge and ability to accurately apply GPCRs) can benefit from speaking more than one language, but the influence was language-pair specific. Finally, oral language skills have been known for many years to be affected by the age at which students learn their second language. Students who only learn their second language once they start school are necessarily delayed in their learning, which impacts their reading ability. As Adams (1990, as cited by Stahl et al. 1990) showed, the younger a child is when they start to learn their second language, and the length of time that they have been speaking it for both affect their language ability.

When looking at the PISA (2016) data, the graph illustrating how each individual country's monolingual vs. multilingual students performed (Figure 16)

showed that monolingual students generally had a smaller range of reading ability than those students who spoke a different language at home. That is, for each country, the poorest-reading monolingual students were gene closer in reading ability to the strongest-reading monolingual students than was the case for those who spoke a different language at home. That being said, the overall result showed no significant difference in the ranges of reading ability between the two groups. However, in some countries (such as Ireland and America) the two groups showed very little difference in range of reading ability, whilst in others (such as Australia and Canada), those students who spoke a different language at home showed a much larger range of reading ability than their monolingual students. This is not altogether surprising, as students who speak a different language are more likely to be the children of relatively recent immigrants, and therefore might be less likely to have received their full education in English. However, this does not altogether tie in with comments by Galletly & Knight (2017) who suggest that the strong performances of Canada and Ireland in PISA results might be due to an "advantage built from multilingualism" (p. 8), although it is important to remember that these comments were based on overall reading achievements for each country rather than the size of the range of reading ability. It is true that the Irish students who speak a different language at home score very similarly to their monolingual peers, but the same does not appear to be the case for Canadian students.

When looking at the results of PIRLS (2016), the data allowed a finer-grained analysis of what language a child spoke at home. When comparing four groups (those who always / almost always / sometimes / never spoke English at home), the first three groups did not show any significant differences in their mean ranges of reading abilities. The fourth group (those who never spoke English at home),

however, did show a significantly larger range of reading ability than the other three groups. That is, the poorest readers amongst those children who never spoke English at home were significantly further behind the best readers in their group than was the case for any other group. This is not surprising, as those who never speak English at home are the most likely to be those who had immigrated relatively recently, and therefore the most likely to have had only a short amount of time learning to read in an English-language setting.

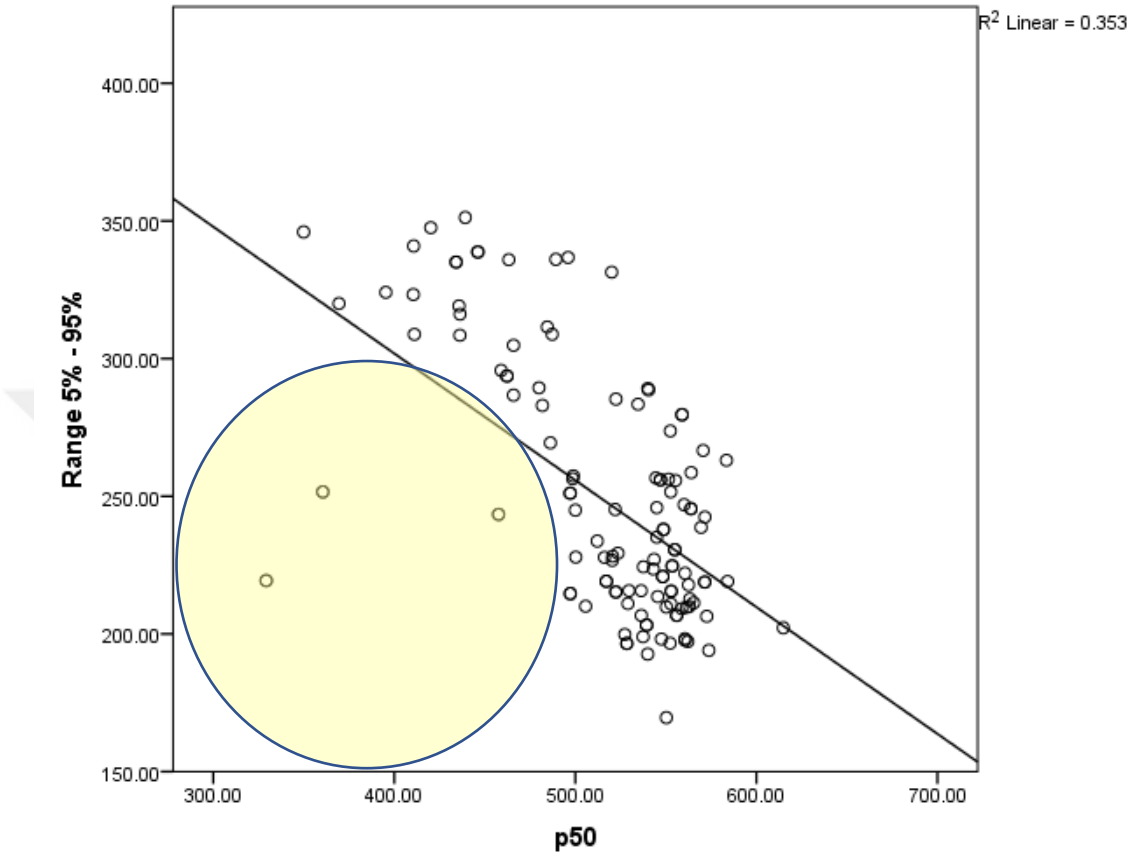
General Comments Concerning the Inclusion/Exclusion of National Results from Analyses

As discussed earlier, the decision whether or not to include the results from Arabic-speaking countries in the analyses was not altogether straightforward. While it is true that the IEA informed the researcher that the decision to include diacritics was not standardized (for PIRLS), the researcher was only able to get final confirmation from the National Research Coordinators of Qatar and Bahrain – who did in fact print the test materials with diacritics. Not only that, but the spread of ranges of the Arabic-speaking countries that took part in PIRLS (2016) was not especially large. If some countries had included diacritics and some had not, it might be reasonable to expect a wider spread of ranges. This suggests that all countries may, in fact, have printed their test materials with diacritics. An alternative explanation for why the results of the Arabic-speaking countries would not conform to the pattern of the other languages stems from the overall educational success of those countries. All the Arabic-speaking countries were found towards the bottom of the table of overall reading achievement (Mullis et al., 2017, p.20), which would seem to suggest that Arabic-speaking countries are not teaching reading as effectively as other countries in the first place. As such, the influence of language-

complexity factors in those countries might be insignificant in the face of poor overall educational outcomes. That possibility leads to a more general question regarding the choice of which languages or countries to include or exclude from this study. When selecting which data to analyse, results were included from all countries and languages for which variable values had been identified. This leaves open the possibility that alternative factors, such as the overall quality of teaching in any given country, might in fact be more important, but their impacts were effectively ignored. Overall quality of educational system is not a factor that was considered in this study, but its impact on the overall range of reading ability might be expected to be considerable. From a cursory view of the distribution of reading achievement (Mullis et al., 2017, p.20), it appears as if the range of reading ability does increase as you go down the table – could this be driving all the results found in this study? An initial correlational analysis seems to confirm this ($r(124) = -.594$, $p < .001$) and a scatter graph clearly illustrates this distribution (see Figure 18).

Figure 18

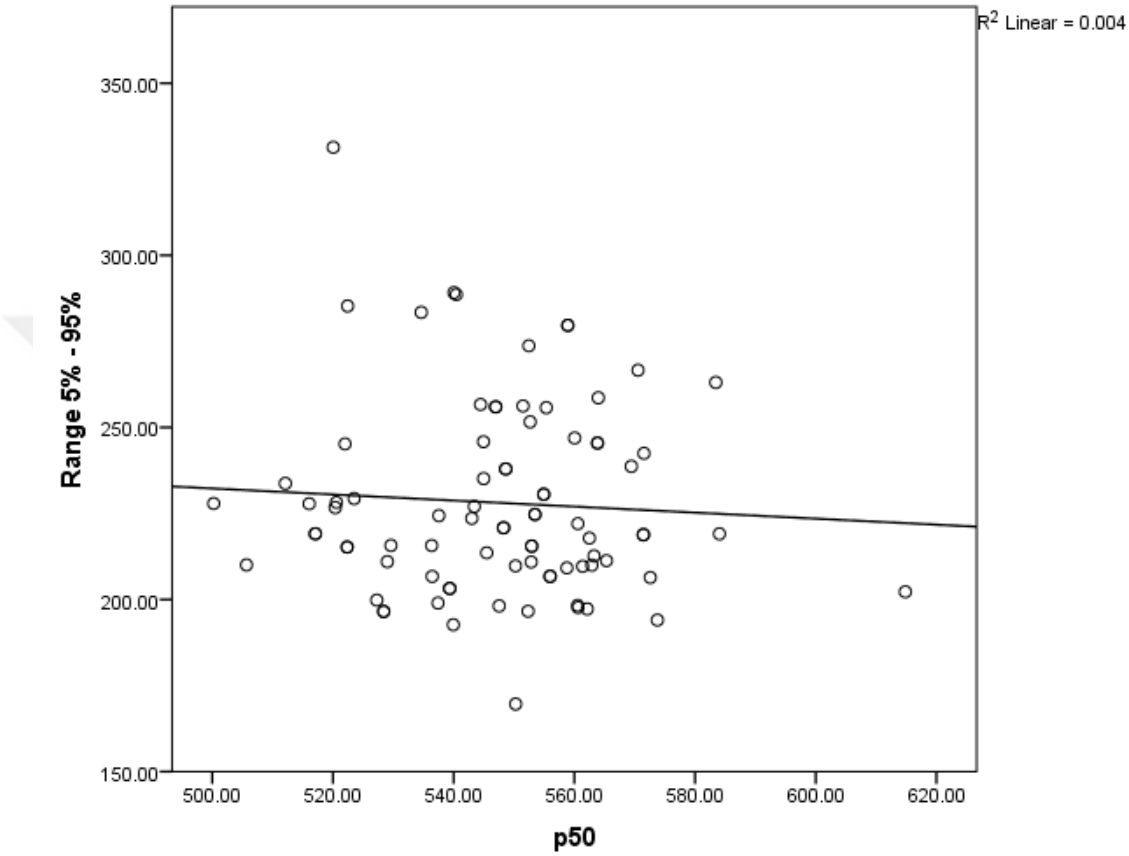
Scattergram Showing the Shared Variance Between Range of Reading Ability and Score of the 50th Percentile for All Countries, PIRLS (2016)



It seems clear that for countries whose 50th percentile scored under 500, the range of reading ability is increased – almost none of those countries fall under the line of best fit (area highlighted on graph). If we accept that countries which scored poorly overall are those countries with the least successful educational systems, it is almost to be expected that their poorest readers would lack the necessary support and would therefore fall even further behind their strongest readers. But what happens if we consider only those countries in the top half of the distribution, whose education systems we can assume are sufficiently well designed and implemented? See Figure 19.

Figure 19

Scattergram Showing the Shared Variance Between Range of Reading Ability and Score of the 50th Percentile for Those Countries Whose 50th Percentile Scored Above 500, PIRLS (2016)



As can be seen in this case, the correlation essentially disappears (r(83) = -.061, p = .581). However, re-conducting the initial analyses on this reduced dataset shows that the results still stand, and in fact increase in most cases:

- For orthographic complexity (DRC), the correlation increases slightly (r(33) = .355, p = .04)
- For orthographic transparency (OTEANN), the correlation increases (r(42) = -.761, p < .001)
- For morphological complexity (TTR), there remains no correlation (r(67) = -.019, p = .875)

- For morphological unpredictability (H3), the correlation reaches significance at the 5% level ($r(67) = .274$, $p = .024$)

In other words, it seems that the overall quality of education does correlate with a greater reduction in the reading ability of poorer readers than strong readers, subject to a floor effect, if the quality of education is below a certain level. For countries whose educational systems are sufficiently well implemented, that relationship falls away, whereas the influences of language-complexity factors seem to increase. This should be confirmed with further study.

With respect to the association between range of reading ability and morphological unpredictability, the picture is not altogether clear. When including all countries (including Arabic-speaking ones), the correlation effect is strong ($r(96) = .451$, $p < .001$). However, when analysing the results of countries whose 50^{th} percentiles scored 500 or above, the correlation reduces in strength ($r(67) = .274$, $p = .024$), and including all countries except Arabic-speaking ones, the effect disappears altogether ($r(86) = .118$, $p = .273$). These results do not seem to lend themselves to a simple explanation, and further study is called for. In particular, it would perhaps be helpful to examine the specific effect of morphological unpredictability on reading development, as well as its impact on reading ability at different ages. While there is growing research into the influence of morphological complexity (Borleffs et al., 2019), there seems to be a lack of similar interest in the influence of morphological entropy or unpredictability.

Implications for Further Research

The first part of this study is, at heart, an explanatory correlational study that has highlighted some interesting, possible connections between orthographic and morphological factors in different languages and the patterns of reading ability seen

in young readers of those languages. As such, although no causation can be inferred from the analyses conducted, they do suggest directions for further research.

As a first step, it is worth noting that the measures chosen in this study to quantify the orthographic and morphological variables may not have been the most appropriate for the particular analyses conducted. Efforts were taken to ensure that the variables were well documented and peer reviewed, but it is notable that there were far fewer datapoints for both of the orthographic variables, compared to the morphological variables. Not only that, but it is possible that the linguistic corpora with which the various measures were calculated might not be the most appropriate or representative when attempting to correlate with the specific results of PISA and PIRLS reading assessments. Running computer-learning algorithms on different corpora gives rise to different output values, even when the corpora are thematically linked, as illustrated by the differences in values calculated by Gutierrez-Vasques & Mijangos (2019) using bible verses and online magazine articles from the Jehovah's Witnesses. It might therefore be of interest to calculate values for the different variables using age-appropriate linguistic corpora, as the values used were all calculated with adult-level samples of reading matter. This would be of particular interest given the slightly paradoxical findings regarding orthographic complexity that showed an increased correlation with the range of reading ability amongst the older children. If analysis of the test materials could demonstrate a higher number of words using more complex GPCRs in the test materials for the older children, this could perhaps help to make sense of the findings.

As a second step, in order to move towards making claims of causality, it would be important to engage in experimental manipulation of the variables. However, when looking at the two broad categories of language-complexity factors

identified in this study, it seems clear that one (orthography) is more open to experimental manipulation than the other (morphology). That is, while it is relatively trivial to invent new spelling rules and even complete orthographies for a language, the same is not the case for a language's morphology – at least, perhaps not without inventing completely new languages, such as Esperanto. Indeed, there are many examples through history of countries that have changed the orthography of their language – ranging from fairly straightforward spelling reforms such as German in 1901 and again in 1996, to more fundamental changes of alphabet, such as the language and alphabet reforms in Turkey from Ottoman Turkish to modern Turkish, as undertaken by the Turkish Language Association and initiated by Mustafa Kemal Atatürk in 1932. That being the case, this study provides support for arguments underpinning spelling reform in English, as it seems to indicate that the orthographically opaque spelling of English could be fundamentally linked to the higher numbers of poor readers in anglophone nations. There has been a long history of (mostly failed) attempts to reform the spelling of English – indeed Milton (1996, p. 2) claims that "people have been proposing schemes of [English] spelling reform for at least four centuries." One attempt at spelling reform that had slightly more success than most was the development of the Initial Teaching Alphabet (ITA) during the 1960's and 1970's. As described by Knight et al. (2017), the ITA was developed in 1961 and "is an almost fully regular English orthography commonly using 44 GPCs […] designed to ease and expedite Anglophone children's early reading and writing development through minimising the complexity that Standard English spelling and letters create" (pp 26-27). There was a considerable amount of research on its benefits during the 1960's and 1970's, but it has almost completely disappeared from academic and educational discourse since then. The authors review

a considerable amount of cross-linguistic research findings covering reading acquisition in regular orthographies, the specific challenges presented by English orthography, and the benefits which were observed when children were taught using the ITA. Some of the most striking of these are perhaps the findings that:

- Use of ITA strongly expedited word reading, word writing, vocabulary and language skills for literacy, reading comprehension and written expression.
- It significantly reduced the number of children experiencing reading and literacy difficulties.
- It produced verbal-efficiency effects including enhanced vocabulary, and more mature ideation and thinking. (p. 33)

Not only that, but the authors discuss research by Stevenson on army personnel, reported in 1973 by Pitman. Personnel with low literacy rates were taught to read ITA, and some of the findings were as follows:

- The rapid success rate induced more controlled emotional responses and gain in self-respect.
- The students gained self-confidence and a new awareness and independence.
- The men concerned became more co-operative and less anti-social and, therefore, both better citizens and more proficient soldiers (p. 29)

In other words, using the ITA was highly effective when teaching both children and adults, and it allowed anglophones to achieve reading and writing results similar to those seen in learners of highly regular existing orthographies. It was also effective in countering some of the negative emotional and societal problems associated with low literacy levels. The teachers who implemented ITA in

their classrooms were apparently nearly unanimous in their praise of the positive outcomes it produced, and although there were influential sceptics, there was a significant amount of rigorous research carried out that backed up those positive anecdotal impressions. The authors were therefore surprised at how completely ITA had disappeared over the course of the 1970's, both from academic research and from practical teaching environments. They concluded that there were perhaps two main factors which influenced its disappearance: Firstly, for various reasons, ITA was only used in the initial stages of reading instruction, rather than being used in conjunction with regular English orthography over a period of several years (as is the case with Zhu-Yin-Fu-Hao in Taiwan), and might therefore have lost its effectiveness over time. Secondly, the authors point to the growing value placed on the Whole Language teaching philosophy during the 1970's, which specifically avoids specific word-reading instruction.

Here, it is perhaps worth also noting the work of Katzir et al. (2012), who looked at Fourth Grade readers of Hebrew and English. Hebrew orthography is of interest in this context as it has both a pointed (vowelized) version and an un-pointed (un-vowelized) version. The authors describe pointed Hebrew as a shallow orthography, and un-pointed Hebrew as a deep orthography, but they do not discuss whether this is due to complexity, unpredictability (opaqueness) or something else. Schmalz et al. (2015) discuss Hebrew as being a language whose orthographic depth is primarily due to its *incompleteness*, a variable that this study has not addressed. Students normally become proficient readers of pointed Hebrew by the end of the first year of school, and are expected to attain proficiency in un-pointed Hebrew by the end of Grade Four. Katzir et al. showed that Grade Four Hebrew-speaking students were better at word-identification than their English-speaking peers, but

were slower at reading in a timed task. This suggests that while the twin versions of Hebrew orthography supported the development of some aspects of reading development, it hindered others – even in comparison to English, a highly opaque orthography. However, it is worth remembering that as Israel falls towards the bottom of the results tables for both PISA (2018) and PIRLS (2016), the influence of educational system may outweigh the impact of any language-complexity factors, as discussed above.

It is likely that further study may identify other factors that are also correlated with how close the poor readers in a country are able to get to the strong readers. These might range from additional language-complexity variables, such as differences in syntax, to population-level variables, such as societal inequality. The Gini index (World Bank, n.d.) is one possible such measure that focuses on income inequality. However, an initial analysis of the PIRLS (2016) data did not show any significant correlation with range of reading ability, either for the full set of results ($r(41) = .164$, $p = .299$) or for those countries in the top half of the distribution ($r(36) = -.026$, $p = .884$).

With regard to the influence of additional home languages on reading ability, this study does not seem to wholly support the insight of Galletly & Knight (2017), who suggested that the bilingual environments may help to explain Ireland's and Canada's successes in PISA reading tests. That being said, the analysis conducted in this study only focused on the overall influence of bilingualism across a small number of English-speaking countries. It may be the case that further analysis, taking into account more countries and/or languages, would be able to find significant correlations. As the impact of immigration and time spent in education in English are likely a factor, as discussed above, it would be helpful to identify those bilingual

students who are either (i) native and whose home language was perhaps a different, recognized national language, (ii) children of immigrants who were either born in the country or who had arrived at a young enough age to have received their full education in that country, or (iii) immigrants who had arrived at some point after the start of school (and how long they had lived there).

Finally, it is worth noting that this study made use of ILSAs sat by children at ages ten and fifteen. It may also be of value to study the connection between range of reading ability and the various factors at both earlier and later stages in reading development, perhaps during the first year of learning to read and also in adults who have left full-time education. This would be of particular interest in the context of the impact of morphological complexity seeming to become useful to children after their third year of formal schooling (Borleffs et al., 2019).

Implications for Practice

In the absence of spelling reform, it seems clear that those who teach people how to read English need to do as much as possible to ameliorate the challenges imposed by its spelling. It would be especially important to keep in mind the order of development of the various cognitive processes needed for success. Various approaches to the teaching of reading can be identified in different anglophone countries, such as the Whole Language approach employed by New Zealand, compared to the synthetic phonics approaches that can be seen in English, American and Canadian classrooms. It is likely that these different approaches will have different strengths and weaknesses and as (Knight et al., 2017) suggested, these might lead to differences in the patterns of reading ability seen in those countries. Whilst synthetic phonics approaches may support the initial acquisition of the high number of context-dependent GPCRs in English that are the cornerstone of early

reading acquisition, it is conceivable that whole-language approaches might be better placed to make use of a possible benefit of English's relatively high morphological complexity.

If it were possible to reignite interest and research into spelling reform and alternative orthographies in English, earlier research into the ITA suggest that it may provide possibilities for improving literacy levels in readers of all ages. There are many reasons why English spelling has so far resisted reform, ranging from the lack of any regulatory body with a mandate to control its spelling of the type that exists for almost all other major languages, to the wide variety of local dialects and accents making it difficult to identify which pronunciation to determine should be classified as prototypical or "correct". It is also important to recognize the historical context of the evolution of modern English spelling. England has a long history of being invaded, variously by peoples of Germanic, Scandinavian, Celtic, French and Latin descent, each bringing their languages with them, not to mention the more universal phenomenon of loan words being adopted, unaltered, from other languages. As such, the spellings of English words hold not only a wealth of morphological information withing them, but also a sort of cultural history that many people feel a strong affinity for. That being said, it is possible to imagine that more up-to-date methods, including the use of computerized teaching paradigms that could take into account local accents, as well as the power of tools such as Google Translate to automatically transliterate the standard English orthography into a different one might be able to overcome some of these difficulties and roadblocks.

Research into self-teaching suggests that it is an important aspect of how individuals become successful, sophisticated readers (Nation et al., 2007; Nation & Angell, 2006; Share, 2004). At the same time, the Orthographic Advantage Theory

put forward by Knight et al. (2019) suggests that the particular challenges caused by the complexity of English spelling make it harder for children to gain a sufficiently high level of reading ability to be able to engage in self-teaching. Their model highlights six areas of educational and lifelong advantage and disadvantage associated with a language's orthography, including "simplified school instruction and learning across primary and secondary school" (p.5). Together with the results of this study, these ideas suggest that the development of an alternative, more transparent orthography would allow teachers to spend less time engaging in guided reading instruction, freeing up time for teaching of other skills, and would also give children (and adults) the chance to engage in greater amounts of self-teaching.

Limitations

The first limitation to this study stems from the fact that there is no formal or universally recognised definition that provides the basis for measuring a language's orthographic complexity or transparency. Not only that, but the measures that were found were calculated based on adult-level corpora, rather than on corpora of the same reading-levels as the test materials. As such, it is highly conceivable that more appropriate measures could be found or calculated that would allow for greater sensitivity in the analyses conducted.

The second limitation is similar to the first, but relates to morphological (rather than orthographic) complexity. Various different measures of morphological complexity and unpredictability have been suggested, and their scores have again been calculated for different groups of languages. As with the measures of orthographic complexity, these were calculated by making use of computer learning, meaning that the choice of training corpora could have significant impacts on the resulting values.

The third limitation is related to the means of collection of the student data that the analyses rely on: The student responses to the tests may not reflect their true abilities and opinions; More importantly, it is possible that the student samples may not be truly representative of each country's populations, despite the sampling designs of the IEA and the OECD. Research published by Jerrim (2021) suggests that high non-participation rates, especially amongst the lowest-performing students, might lead to "potential biases in the data" (p.3). If this were the case, it would suggest that the true range of reading ability in England might be higher than the results indicate.

Fourthly, if orthographic depth is shown to be connected to the distribution patterns of reading ability, this would beg the question of whether there is anything that can be done to help mitigate the impacts. The Initial Teacher Alphabet (ITA) was developed in the 1960's as an attempt to create a regular English orthography to support early reading acquisition, in much the same way as Zhu-Yin-Fu-Hao supports the early acquisition of Chinese in Taiwan. As described by Galletly et al. (2017), the ITA was developed as "an almost fully regular English orthography commonly using 44 GPCs […] designed to ease and expedite Anglophone children's early reading and writing development through minimising the complexity that Standard English spelling and letters create" (pp 26-27). Engaging in experimental study of any possible benefits of alternative orthographies is beyond the scope of this study, but it may add further weight to calls for such investigations.

Finally, the analysis of the impact of monolingualism on range of reading ability was limited to only looking at the overall impact of whether students spoke English at home or not in a relatively small number of anglophone countries. This

did not take into account several other factors that could be of interest and might merit further study, such as:

- The impact of specific home languages other than English (this information is not available in the datasets.)
- The impact of home language in individual anglophone countries (the results in this study were pooled as the numbers of students in individual groups for each country were very small)
- The impact of multilingualism in non-anglophone countries.

REFERENCES

Addey, C., & Sellar, S. (2018). Why do countries participate in PISA? Understanding the role of international large-scale assessments in global education policy. In A. Verger, M. Novelli, & H. K. Altinyelken (Eds.), *Global Education Policy and International Development* (2nd ed., pp. 97–118). Bloomsbury. https://doi.org/10.5040/9781474296052.ch-005

Arikan, S., Özer, F., Şeker, V., & Ertas, G. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(1), 43–60. https://doi.org/10.21031/EPOD.602765

Baerman, M., Brown, D., & Corbett, G. G. (2015). Understanding and measuring morphological complexity: An introduction. In M. Baerman, D. Brown, & G. G. Corbett (Eds.), *Understanding and measuring morphological complexity*. Oxford University Press. https://doi.org/10.1093/ACPROF:OSO/9780198723769.003.0001

Berndt, R. S., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, & Computers 1987 19:1*, *19*(1), 1–9. https://doi.org/10.3758/BF03207663

Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics*, *21*, 169–181. https://doi.org/10.1017/S0267190501000101

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, *57*(SUPPL. 1), 45–77. https://doi.org/10.1111/J.1467-9922.2007.00412.X

Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. M. B. (2005). Onset entropy matters: Letter-to-phoneme mappings in seven languages. *Reading and Writing*, *18*(3), 211–229. https://doi.org/10.1007/s11145-005-3001-9

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative review. *Reading and Writing*, *30*(8), 1617. https://doi.org/10.1007/S11145-017-9741-5

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2019). Cracking the code: The impact of orthographic transparency and morphological-syllabic complexity on reading and developmental dyslexia. *Frontiers in Psychology*, *9*, 1–19. https://doi.org/10.3389/fpsyg.2018.02534

Brendemoen, B. (2021). The Turkish language reform. In L. Johanson & É. Á. Csató (Eds.), *The Turkic Languages* (pp. 231–235). Routledge. https://doi.org/10.4324/9781003243809-15

Britannica. (n.d.). Morpheme. https://www.britannica.com/topic/morpheme

Brooks, G. (2015). The grapheme-phoneme correspondences of English, 2: Graphemes beginning with vowel letters. In *Dictionary of the British English spelling system*. Open Book Publishers. https://doi.org/10.11647/OBP.0053.10

Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2011). The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual, English-speaking peers. *British Journal of Educational Psychology*, *81*(2), 344–354. https://doi.org/10.1348/000709910X504122

Burnett, N. (2005). *Education for all, Literacy for life*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000141639

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(3), 455–474. https://doi.org/10.1017/S0142716407070257

Cohen, J. (1977). The Significance of a Product Moment rs. *Statistical Power Analysis for the Behavioral Sciences*, 75–107. https://doi.org/10.1016/B978-0-12-179060-8.50008-6

Collins English Dictionary. (n.d.-a). Grapheme. https://www.collinsdictionary.com/dictionary/english/grapheme

Collins English Dictionary. (n.d.-b). Phoneme. https://www.collinsdictionary.com/dictionary/english/phoneme

Croft, W. (2002). Introduction. In *Typology and universals* (2nd ed., pp. 1–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511840579.003

De Witt, M. W., & Lessing, A. C. (2017). The deconstruction and understanding of pre-literacy development and reading acquisition. *Early Child Development and Care*, *188*(12), 1841–1854. https://doi.org/10.1080/03004430.2017.1329727

Defior, S., Martos, F., & Cary, L. (2002). Differences in reading acquisition development in two shallow orthographies: Portuguese and Spanish. *Applied Psycholinguistics*, *23*, 135–148. https://doi.org/10.1017/S0142716402000073

Elhassan, Z., Crewther, S. G., & Bavin, E. L. (2017). The contribution of phonological awareness to reading fluency and its individual sub-skills in readers aged 9-to 12-years. *Frontiers in Psychology*, *8*(APR), 533. https://doi.org/10.3389/fpsyg.2017.00533

Fitzsimmons, M. P. (2017). Introduction. In *The place of words: The Académie Française and its dictionary during an age of revolution*. Oxford University Press. https://doi.org/10.1093/oso/9780190644536.001.0001

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German-and English-speaking children. *Scientific Studies of Reading*, *2*(1), 31–54. https://doi.org/10.1207/s1532799xssr0201_2

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*(1), 104–115. https://doi.org/10.1037/0096-1523.13.1.104

Gerth, S., & Festman, J. (2021). Reading development, word length and frequency effects: An eye-tracking study with slow and fast readers. *Frontiers in Communication*, *6*, 202. https://doi.org/10.3389/FCOMM.2021.743113

Gibson, L., Rodriguez, C., Ferguson, S. J., Zhao, J., & Hango, D. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(I), 204–256. https://doi.org/10.1037//0033-295X.108.1.204

Gibson, L., Rodriguez, C., Ferguson, S. J., Zhao, J., & Hango, D. (2019). *Insights on Canadian society: Does reading proficiency at age 15 affect employment earnings in young adulthood?* https://www.statcan.gc.ca

Gontijo, P. F. D., Gontijo, I., & Shillcock, R. (2003). Grapheme–phoneme probabilities in British English. *Behaviour Research Methods, Instruments & Computers*, *35*(1), 136–157. https://doi.org/10.3758/BF03195506

Grosjean, F. (2021). The extent of bilingualism. In *Life as a Bilingual* (pp. 27–39). Cambridge University Press. https://doi.org/10.1017/9781108975490.003

Gutierrez-Vasques, X., & Mijangos, V. (2019). Productivity and predictability for measuring morphological complexity. *Entropy*, *22*(1), 48. https://doi.org/10.3390/e22010048

Huang, H. S., & Hanley, J. R. (1997). A longitudinal study of phonological awareness, visual skills, and Chinese reading acquisition among first-graders in Taiwan. *International Journal of Behavioral Development*, *20*(2). https://doi.org/10.1080/016502597385324

Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, *25*(3), 443–463. https://doi.org/10.1080/13825585.2017.1319902

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *Review of Education*, *9*(3), e3270. https://doi.org/10.1002/REV3.3270

Kapnoula, E. C., Protopapas, A., Saunders, S. J., & Coltheart, M. (2017). Lexical and sublexical effects on visual word recognition in Greek: Comparing human behavior to the dual route cascaded model. *Language, Cognition and Neuroscience*, *32*(10), 1290–1304. https://doi.org/10.1080/23273798.2017.1355059

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology*, *94*(C), 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2

Katzir, T., Schiff, R., & Kim, Y. S. (2012). The effects of orthographic consistency on reading development: A within and between cross-linguistic study of fluency and accuracy among fourth grade English- and Hebrew-speaking children. *Learning and Individual Differences*, *22*(6), 673–679. https://doi.org/10.1016/J.LINDIF.2012.07.002

Knight, B. A., Galletly, S. A., & Gargett, P. S. (2019). Orthographic Advantage Theory: National advantage and disadvantage due to orthographic differences.

Asia Pacific Journal of Developmental Differences, *6*(1), 5–31. https://www.dyslexia.org.sg/images/publications/apjdd/apjddjan2019/APJDD-V6-1-A1.pdf

Knight, B. A., Galletly, S., & Gargett, P. (2017). Managing cognitive load as the key to literacy development: Research directions suggested by crosslinguistic research and research on Initial Teaching Alphabet (ITA). In *Progress in Education, Vol. 45* (Issue January, pp. 0–46).

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tõth, D., Honbolygõ, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J. F., … Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(6), 686–694. https://doi.org/10.1111/jcpp.12029

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*(3), 315–334. https://doi.org/10.1016/S0010-0277(97)00005-X

LaRoche, S., & Foy, P. (2017). Sample implementation in PIRLS 2016. In *Methods & Procedures in PIRLS 2016* (pp. 5.1-5.126). https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-5.html

Laroche, S., Joncas, M., & Foy, P. (2017). TIMSS & PIRLS sample design in PIRLS 2016. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016* (Issue 1). IEA.

Lukie, I. K., Skwarchuk, S. L., LeFevre, J. A., & Sowinski, C. (2014). The role of child interests and collaborative parent-child interactions in fostering numeracy

and literacy development in Canadian homes. *Early Childhood Education Journal*, *42*(4), 251–259. https://doi.org/10.1007/S10643-013-0604-7

Marjou, X. (2021). OTEANN: Estimating the transparency of orthographies with an artificial neural network. *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP*, 1–9. https://doi.org/10.18653/V1/2021.SIGTYP-1.1

Merriam-Webster Dictionary. (n.d.-a). Morphology. https://www.merriam-webster.com/dictionary/morphology

Merriam-Webster Dictionary. (n.d.-b). Orthography. https://www.merriam-webster.com/dictionary/orthography

Milton, R. (1996). *English spelling and the computer*. Longman. http://eprints.bbk.ac.uk/archive/00000469

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *International results in reading PIRLS 2016*. TIMSS & PIRLS International Study Center. http://pirls2016.org/download-center/

Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2015). PIRLS 2016 reading framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 assessment framework* (2nd ed., pp. 11–29). TIMSS & PIRLS International Study Center.

Nation, K., & Angell, P. (2006). Learning to read and learning to comprehend. *London Review of Education*, *4*(1), 77–87. https://doi.org/10.1080/13603110600574538

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*(1), 71–84. https://doi.org/10.1016/j.jecp.2006.06.004

NCES, N. C. for E. S. (2018). *Highlights of U.S. PISA 2018 results web report*. https://nces.ed.gov/surveys/pisa/pisa2018/index.asp#/

OECD. (2016). *Sampling in PISA*. https://www.oecd.org/pisa/pisaproducts/SAMPLING-IN-PISA.pdf

OECD. (2017). *PISA 2015 Technical Report*. OECD.

OECD. (2018). *Chapter 4 sample design: Target population and overview of the sampling design*. OECD.

OECD. (2019a). *PISA 2018 results, combined executive summaries, volumes i, ii & iii*. https://www.oecd.org/about/publishing/corrigenda.htm

OECD. (2019b). *PISA 2018 results (Volume I) What students know and can do*. https://doi.org/10.1787/5f07c754-en

OECD. (2019c). *PISA 2018 results (Volume II) Where all students can succeed*. https://doi.org/10.1787/b5fd1b8f-en

OECD, & Statistics Canada. (2000). *Literacy in the information age. Final report of the International Adult Literacy Survey*. https://www.oecd.org/education/skills-beyond-school/41529765.pdf

Öney, B., & Durgunoğlu, A. Y. (1997). Beginning to read in Turkish: A phonologically transparent orthography. *Applied Psycholinguistics, 18*(1), 1–15. https://doi.org/10.1017/s014271640000984x

Polka, L., Orena, A. J., Sundara, M., & Worrall, J. (2017). Segmenting words from fluent speech during infancy - challenges and opportunities in a bilingual context. *Developmental Science, 20*(1), 1–14. https://doi.org/10.1111/DESC.12419

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods 2009 41:4*,

41(4), 991–1008. https://doi.org/10.3758/BRM.41.4.991

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. https://doi.org/10.1016/J.CORTEX.2018.02.008

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: More evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658–672. https://doi.org/10.1080/20445911.2016.1182172

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, *22*(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

Seymour, P. H. K. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 296–315). Blackwell. https://doi.org/10.1002/9780470757642.ch16

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267–298. https://doi.org/10.1016/J.JECP.2004.01.001

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). In *Biological Psychiatry* (Vol. 57, Issue 11, pp. 1301–1309). https://doi.org/10.1016/j.biopsych.2005.01.043

Silinskas, G., Torppa, M., Lerkkanen, M.-K., & Nurmi, J.-E. (2020). The home literacy model in a highly transparent orthography. *School Effectiveness and School Improvement*, *31*(1), 80–101. https://doi.org/10.1080/09243453.2019.1642213

Spanoudis, G. C., Papadopoulos, T. C., & Spyrou, S. (2019). Specific language

impairment and reading disability: categorical distinction or continuum? *Journal of Learning Disabilities*, *52*(1), 3–14. https://doi.org/10.1177/0022219418775111

Sprenger-Charolles, L., Siegel, L. S., Jiménez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading*, *15*(6), 498–521. https://doi.org/10.1080/10888438.2010.524463

Stahl, S. A., Osborn, J., & Lehr, F. (1990). *Beginning to read: thinking and learning about print; A summary*. https://doi.org/10.2307/415121

Steiner-Khamsi, G. (2020). Conclusions: What policy-makers do with PISA. In F. Waldrow & G. Steiner-Khamsi (Eds.), *Understanding PISA's attractiveness: Critical analyses in comparative policy studies* (pp. 233–248). Bloomsbury Collections. https://doi.org/10.5040/9781350057319.CH-012

Sum, A. (1999). *Literacy in the labor force. Results from the National Adult Literacy Survey*. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999470

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505–513. https://doi.org/10.1016/J.WOCN.2010.08.006

Thomson, S., De Bortoli, L., & Underwood, C. (2016). *PISA 2015: A first look at Australia's results*. https://research.acer.edu.au/ozpisa/21/

Torppa, M., Georgiou, G., Salmi, P., Eklund, K., & Lyytinen, H. (2012). Examining the double-deficit hypothesis in an orthographically consistent language. *Scientific Studies of Reading*, *16*(4), 287–315. https://doi.org/10.1080/10888438.2011.554470

Torppa, M., Parrila, R., Niemi, P., Lerkkanen, M. K., Poikkeus, A. M., & Nurmi, J. E. (2013). The double deficit hypothesis in the transparent Finnish orthography: A longitudinal study from kindergarten to Grade 2. *Reading and Writing*, *26*(8), 1353–1380. https://doi.org/10.1007/s11145-012-9423-2

Tunmer, W. E., Chapman, J. W., Greaney, K. T., Prochnow, J. E., & Arrow, A. W. (2013). Why the New Zealand National Literacy Strategy has failed and what can be done about it: Evidence from the Progress in International Reading Literacy Study (PIRLS) 2011 and Reading Recovery monitoring reports. *Australian Journal of Learning Difficulties*, *18*(2), 139–180. https://doi.org/10.1080/19404158.2013.842134

Tunmer, W. E., Nesdale, A. R., & Wright, A. D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology*, *5*(1), 25–34. https://doi.org/10.1111/J.2044-835X.1987.TB01038.X

Ulicheva, A., Coltheart, M., Saunders, S., & Perry, C. (2016). Phonotactic constraints: Implications for models of oral reading in Russian. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(4), 636–656. https://doi.org/10.1037/XLM0000203

Upward, C., & Davidson, G. (2011). *The history of English spelling*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781444342994.CH1

van den Bosch, A., Content, A., Daelemans, W., & de Gelder, B. (1994). Analysing orthographic depth of different languages using data-oriented algorithms. *Proceedings of the 2nd International Conference on Quantitative Linguistics, QUALICO-94*, *December*, 26–30.

Velupillai, V. (2012). *An introduction to linguistic typology*. John Benjamins Publishing Company. https://doi.org/10.1075/z.176

Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics*, *28*(3), 425–439. https://doi.org/10.1017/S0142716407070233

Wikanengsih, San Fauziya, D., & San Rizqiya, R. (2020). The correlation between students' critical reading ability and their mathematical critical thinking. *Journal of Physics: Conference Series*, *1657*(1), 012041. https://doi.org/10.1088/1742-6596/1657/1/012041

World Bank. (n.d.). *Gini index | Data*. https://data.worldbank.org/indicator/SI.POV.GINI

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*(3), 413–430. https://doi.org/10.1080/09541440050114570

Ziegler, J. C., Perry, C., & Coltheart, M. (2003). Speed of lexical and nonlexical processing in French: The case of the regularity effect. *Psychonomic Bulletin & Review 2003 10:4*, *10*(4), 947–953. https://doi.org/10.3758/BF03196556

APPENDIX A

Countries Included in Analyses Involving Orthographic Depth, Listing the Pertinent Language or Languages They Used to Administer PISA (2018) and/or PIRLS (2016)

Country	PISA (2018)	PIRLS (2016)
Argentina	Spanish	Spanish
Australia	English	English
Austria	German	German
Bahrain	-	English, Arabic
Baku (Azerbaijan)	Russian	Russian
Belarus	Russian	-
Belgium	Dutch, French, German	Dutch, French, German
Bosnia & Herzegovina	Serbo-Croatian	-
Brazil	Portuguese	-
Brunei Darussalam	English	-
Canada	English, French	English, French
Chile	Spanish	Spanish
Colombia	Spanish	-
Costa Rica	Spanish	-
Croatia	Croatian	-
Denmark	Danish	-
Dominican Republic	Spanish	-
Estonia	Russian	-
Finland	-	Finnish
France	French	French
Georgia	Russian	-
Germany	German	German
Hong Kong	English	-
Ireland	English	English
Israel	Arabic	Arabic
Italy	German, Italian	German, Italian
Jordan	Arabic	-
Kazakhstan	Russian	Russian
Korea	Korean	-
Latvia	Russian	Russian
Lebanon	English, French	-
Lithuania	Russian	Russian
Luxembourg	English, French, German	-

Country	PISA (2018)	PIRLS (2016)
Macao	English, Portuguese	English, Portuguese
Malaysia	English	-
Malta	English	-
Mexico	Spanish	-
Moldova	Russian	-
Morocco	Arabic	Arabic
Moscow Region (RUS)	Russian	-
Netherlands	Dutch	Dutch
New Zealand	English	English
Oman	-	English, Arabic
Panama	English, Spanish	-
Peru	Spanish	-
Philippines	English	-
Portugal	Portuguese	Portuguese
Qatar	Arabic, English	Arabic, English
Russian Federation	Russian	Russian
Saudi Arabia	Arabic, English	Arabic, English
Serbia	Serbian	-
Singapore	English	English
South Africa	-	English
Spain	-	Spanish
Switzerland	French, German, Italian	-
Tatarstan (RUS)	Russian	-
Trinidad & Tobago	-	English
Turkey	Turkish	-
Ukraine	Russian	-
United Arab Emirates	Arabic, English	Arabic, English
United Kingdom	English	English
United States	English	English
Uruguay	Spanish	-