POST-SECONDARY READING DEVELOPMENT AND PRINT EXPOSURE IN L1 AND L2 SPEAKERS OF ENGLISH

Sean Patrick McCarron

POST-SECONDARY READING DEVELOPMENT AND PRINT EXPOSURE IN L1 AND L2 SPEAKERS OF ENGLISH

By SEAN PATRICK McCARRON, B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2020) Hamilton, Ontario
(Cognitive Science of Language)

TITLE: POST-SECONDARY READING DEVELOPMENT AND PRINT EXPOSURE IN L1 AND L2 SPEAKERS OF ENGLISH

AUTHOR: Sean Patrick McCarron, B.A. (McMaster University)

SUPERVISOR: Professor Victor Kuperman

NUMBER OF PAGES: ix, 107

DECLARATION OF STUDENT CONTRIBUTION

Chapter 2 has been submitted to a journal for publication, and is now under revision, with myself (Sean Patrick McCarron) as first author and Dr. Victor Kuperman as second author. The writing for this chapter was primarily completed during the period of Summer and Autumn 2019. My contributions to this chapter were the literature review, parts of the data analysis and interpretation, coding, and conclusions. I was also the principal author for Chapter 3, working closely alongside Dr. Victor Kuperman as second author. The writing for this chapter was primarily completed during the Winter and Spring of 2020, and has been submitted to a journal for publication. My contribution was similar to that of Chapter 2.

Chapters 1 and 4, which respectively introduce and summarize my thoughts on this research, are entirely my own work.

Abstract

In this thesis, two studies are presented which examine reading development and proficiency in post-secondary education. The first study examines the utility of a common method for determining print exposure, the Author Recognition Test (ART), in populations less frequently examined—namely, college students (as opposed to university students), and individuals whose first language is not English. Item Response Theory analysis shows that ART is not informative for these populations, which suggests that the development of a novel test of print exposure for comparing different populations is necessary. The second study quantifies the impact of each year of post-secondary study on reading development, and the differential effects between native (L1) and non-native (L2) speakers of English. Findings show that each year of study itself is not a significant predictor of change, but rather improvement is explained by advancement in component skills of reading which develop over the course of the degree. Additionally, contrary to previous studies indicative of the Matthew Effect in college literacy development—which suggest that students improve by the end of their degree as a function of their ability at the beginning—this study demonstrates that L2 students generally benefit more from post-secondary education when compared to L1 peers, who start with a significant advantage. In this way, L2 students with sufficient mastery of component skills of reading emerge from post-secondary education with skills comparable to those of native English-speaking colleagues.

Acknowledgements

Thanks to my wonderful wife Lindsay, and our kids (in order of appearance)—Leif, Iver, Elna & Roan.

I am very grateful to the support of The Social Sciences and Humanities Research Council of Canada (SSHRC) for their CGS-M grant which contributed to this research.

Thanks as well to the additional members of my thesis defence committee, Drs. Catherine Anderson and Elisabet Service, for their time, consideration, and valued feedback.

And of course, I am extremely grateful to my supervisor, Dr. Victor Kuperman, for his patience and guidance—and without whom I would undoubtedly still be deciding on a topic for my thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Print Exposure and Reading Development

On the evolutionary scale, reading represents a fairly recent invention—one which has allowed us to extend our limited human memories beyond the oral traditions of our ancestors, record our thoughts, and reanimate the voices of the past. In her book *Proust and the Squid* (2007), neuroscientist and reading specialist Maryanne Wolf describes the ways in which our brains repurpose ancient neural connections to create the reading circuit, which must be forged anew in every developing reader. This "neuronal recycling" (Dehaene, 2009) employs parts of the brain used for visual specialization, object recognition and naming, and mathematical operations, among others, to create a mind exceptionally suited to the task of reading. Given that psychology is the study of the mind, the science of reading is a subdiscipline which offers an illuminating glimpse at its organization. To illustrate this point, Wolf quotes Sir Edmund Huey, one of the early cognitive psychologists who studied the science of reading, who wrote in 1908:

> And so to completely analyze what we do when we read would almost be the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history. (Huey, 1968, p. 6, as partially cited in Wolf, 2007, p. 19)

This "remarkable specific performance" is aptly described, because like any great feat, it requires much practice to develop. From an early age, the rehearsals begin in earnest, with teachers and caregivers sharing books with children, and encouraging them to read as much as possible. This shared reading early in life is an important component for learning to read independently, as it is a strong predictor of later reading ability in L1 (Bus, Van Ijzendoorn & Pellegrini, 1995). Adults learning a second language must also forge new neural connections, although the cerebral scaffolding of an L1 remains ever-present. Linguistic transfer effects can thus be beneficial or detrimental, depending on the degree of similarity or "linguistic distance" between one's L1 and L2 (Grabe, 2014; Koda, 2007). As anyone who has learned a second language can attest, being accustomed to fluent reading in their L1, the initial difficulty of reading in another language can be discouraging. Even with more experience, second-language reading tends to be less automatic and more effortful, which can slow down the processing of information (Godfroid, Winke, & Rebuschat, 2015; Whitford, Pivneva, & Titone, 2016). Higher-skilled readers tend to read more, leading to a beneficially reciprocal relationship between reading ability and reading quantity, while lower-skilled readers read less and therefore benefit less as well (*"Matthew effect",* Kempe, Eriksson-Gustavsson & Samuelsson, 2011; Stanovich, 1986). Unsurprisingly then, L2 learners who do more free voluntary reading (FVR) are generally more proficient readers (Elley and

Mangubhai, 1983; Krashen, 2004; Yamashita, 2008). For these reasons, it is important for second language learners to read frequently in the target language. This takes a great deal of motivation and a positive disposition toward the pastime, given the virtually endless ways to otherwise occupy ourselves. Unfortunately, accurately determining how much an individual reads during their free time is anything but straightforward. Because of the almost universally positive associations with reading as a diversion, self-report measures of FVR are famously unreliable due to the inherent potential for social desirability bias (Stanovich & West, 1989). Consequently, researchers have had to devise ways to investigate reading behaviour more obliquely—typically by gathering information known to correlate with an individual's exposure to print. FVR is also called "reading for pleasure", and this is generally understood to mean fiction-reading (Krashen, 1995). Although many of us read non-fiction for enjoyment as well, it is thought that fiction reading is more likely to be undertaken solely as a leisure activity. Researchers are interested in understanding how much reading is taken on voluntarily, independent of studies—thus, fiction reading generally represents additional reading practice which sets apart higher and lower-skilled readers (Mol & Bus, 2011), and has been shown to predict verbal ability where non-fiction reading has not (Mar & Rain, 2015). As will be discussed further, the consequence of these facts are that language scientists seeking to measure print exposure often must determine an individual's familiarity with the world of fiction.

1.2 Critical Literacy and World Knowledge

As noted above, it is virtually axiomatic that reading is an essential and laudable pastime. Nevertheless, it is worth underscoring that the amount of reading we do

is not trivial, and in fact has important real-world implications—not only for literacy skills in an L1 or L2, but also for knowledge of the world around us. To demonstrate this latter point, Stanovich and Cunningham (1993) tried to understand the source of what they called the "cognitive anatomy of misinformation" (p. 221). The authors found that print and television exposure were both significant predictors of general knowledge—this relationship was simply inverted. Even when accounting for general ability, those who read more also tended to know more about the world around them, and those who consumed more television performed more poorly. Despite 40% of their sample belonging to "one of the most selective public institutions in North America", (i.e. UCLA Berkeley, as per Cunningham & Stanovich, 2001) participants demonstrated a startling lack of general knowledge. By and large, these were not esoteric or academic questions—examples provided in the paper included what kinds of fruit contained Vitamin C, who their U.S. Senator was, which part of the body is affected by pneumonia, and who the United States fought against in the Second World War. Reflecting on this discrepancy between the knowledge of readers and non-readers, Stanovich and Cunningham quote Neil Postman's 1988 book, *Conscientious Objections*, where he states:

> [most people] know of many things; but *about* very little. To provide some verification of this, I conducted a survey a few years back on the subject of the Iranian hostage crisis. I chose this subject because it was alluded to on television every day for more than a *year*... The questions I asked were simple and did not require deep knowledge. For example, Where is Iran? What language do the Iranians speak? Where did the Shah come from? What does 'Ayatollah' mean? I found that almost everybody knew practically nothing about Iran.

> And those who did know something said they had learned it from Newsweek or Time or The New York Times. Television, in other words, is not the great information machine. (Postman, 1988, pp. 171-172, as cited in Stanovich & Cunningham, 1993, p. 224)

If infrequent readers are generally less informed about current events than more consistent readers, it is possible this is due to the "displacement effect", which describes how reading is often sacrificed in favour of more engrossing or passive forms of entertainment such as television. Often discussed in the context of children whose academic performance deteriorates when television watching exceeds a certain threshold (Neuman, 1988), it is less well-studied in adults, who are also prone to sacrificing reading time for other pastimes. In recent decades, of course, the Internet has largely usurped both the television set and the daily newspaper as the primary source of both entertainment and information. Unsurprisingly, this medium of rapid and on-demand information has led to changes in the way we engage with knowledge. For example, we tend to skim more when reading on the Internet when compared to the newspaper (Holmqvist, Holsanova, Barthelson & Lundqvist, 2003; Liu, 2005). Indeed, as far back as the 1990s, it was well-established that Internet readers had a tendency towards scanning information instead of reading it carefully, and web content creators were prescribed ways to render information concise and more easily digestible (Morkes & Nielsen, 1997). Many have raised concerns that this precipitous cultural transition from traditional to digital reading may have serious implications for the reading circuit and our ability to focus while reading (Baron, 2015). Academic literature is rife with reports of deleterious effects of technology use on academic performance, including associations between increased time spent on computers, cell phones, and the Internet on lower grades (Austin & Totaro, 2011; Dietz, & Henrich, 2014; Jacobsen

& Forste, 2011; Lepp, Barkley, & Karpinski, 2015; Wentworth & Middleton, 2014), instant messaging and distractibility (Levine, Waite, & Bowman, 2007), and the harms of media multitasking on task performance, fluid intelligence, impulse inhibition, and memory (Courage, Bakhtiar, Fitzpatrick, Kenny, & Brandeau, 2015; Minear, Brasher, McCurdy, Lewis, & Younggren, 2013; Sparrow, Liu, & Wegner, 2011; Uncapher, Thieu, & Wagner, 2016).

Among those preoccupied by the potential changes to the "reading brain" is the previously cited neuroscientist Maryanne Wolf, who details in her book *Reader, Come Home* (2018) her own personal struggles with returning to reading for pleasure after being immersed in our digital culture. The impetus for writing this book was born partially after being interviewed for the Washington Post in 2014, where Wolf described her difficulty concentrating on long passages of prose in a favourite author's book (Rosenwald, 2014). As per Wolf (2018), despite the enormous volume of reader feedback, the newspaper later informed her that of those who accessed the article online, only 30% had read it to completion. In addition to her own subjective experiences, Wolf summarizes numerous research publications on digital reading, including the increased tendency to skim for information (Liu, 2005), as well as evidence that digital readers show poorer sequential ordering for narrative events compared to peers reading a hard copy of the same book (Mangen & Van der Weel, 2016, 2017). Wolf's conclusion is that the reading brain is indeed declining considerably, and that we must be aware of these changes so that the ability of "deep reading" does not disappear:

> Kurt Vonnegut compared the role of the artist in society to that of the canary in the mines: both alert us to the presence of danger. The reading

> brain is the canary in our minds. We would be the worst of fools to ignore what it has to teach us. (2018, p. 18)

Wolf positions this change as a deterioration in what she terms "cognitive patience", which may have wide-ranging implications for how the average person processes and interprets information. If it is true that the general public is increasingly unable to become absorbed in a compelling but dense novel, it is reasonable to assume these same individuals may similarly struggle to critically examine a more intellectually demanding news story or non-fiction essay filled with nuanced opinions, conflicting reports, and dubious claims. Recently, there has been significant alarm that the general public may lack the critical literacy skills to effectively combat misinformation on the Internet (Cooke, 2018), a skill which is especially important in light of the way misinformation has been spread in response to the COVID-19 pandemic ("infodemiology", Cuan-Baltazar, 2020). Arguably, this acceptance of false or misleading information has been exacerbated in recent years with an increase in populist movements in Western democracies (Budd, 2019, 2020). This loosely-defined ideology is often characterized by a distrust of "urban elites", "expert science", and "university intellectuals" (Merkley, 2019). Perhaps not coincidentally, most Americans obtain news through social media (Gottfried & Shearer, 2016). This trend is similar in Canada, with many reporting that they use social media posts from friends as a news "filter"—most notably the case among a plurality of students and retirees (Hermida, 2012). The current distrust of expert analysis has coincided with a pervasive popular sentiment in some strata of Canadian society that a cursory understanding of complex issues will suffice for important roles in governance and leadership; that complex problems must have simple solutions (e.g. Mike Harris' "common-sense revolution" electoral campaign and Doug Ford's "everyman" appeal, see Budd, 2020; Ferguson, 2018; Gollom,

2019). Nevertheless, if future post-secondary graduates aim to achieve important positions in government or as cultural leaders, they must develop the ability to read carefully, engage earnestly, and think critically about a great volume of material.

1.3 Literacy in Canadian Graduates

Unfortunately, the current state of literacy among post-secondary graduates in Canada is far from encouraging. In the previous decade, the Canadian Council on Learning published a report stating 20% of Canadian graduates were below minimum literacy expectations (CCL, 2009), which paralleled later findings from both Statistics Canada and the Higher Education Quality Council of Ontario, reporting on data from the Programme for the International Assessment of Adult Competencies (PIAAC) which indicated the same was true of a quarter of all university students and half of all college students nearing graduation (Hango, 2014). Despite the generally low level of literacy among post-secondary students, PIAAC data has shown that Canadian graduates still have significantly higher literacy rates than Canadian adults in general (Lane & Murray, 2018).

There is also reason to believe that L2 graduates will fare even more poorly in reading skill when compared to L1 peers. This is important to anticipate, as Canadian universities are home to one of the largest populations of international students in the world, and the percentage has increased steadily in recent years to as high as 14% in 2018 (Statistics Canada, 2020). Among these international students, three-quarters reported speaking a first language other than English, and many stated they had experienced difficulty with language and communication in the education system, including passing proficiency tests required for admission

(Humphries, Knight-Grofe & Klabunde, 2009). It has been established that language ability in the language of instruction is predictive of post-secondary grade performance (Chapman, Wan & Xu, 1988; Wu, Garza & Guzman, 2015). There also exists a known performance disadvantage for first- and second-generation immigrants in Canada, which manifests as lower reading scores on the Program for International Student Assessment (PISA) in all provinces except the Maritimes, where students from immigrant families perform similarly to non-migrant peers (Volante, Klinger, Bilgili, & Siegel, 2017).

1.4 Present Objectives

Because of the myriad benefits of reading, as well as the firmly established relationship between reading volume and reading proficiency, it is important to be able to have an effective instrument for estimating an individual's reading volume without relying solely on self-reports of reading behaviour, which are prone to social desirability biases. Various objective measures of print exposure have been proposed and developed, and one of the most common is the Author Recognition Test (ART, Stanovich & West, 1989). This simple test, explored in greater detail in chapters 2 and 3 of this thesis, asks participants to check off names on a list which they are certain belong to those of published authors. Half of the names presented are indeed real authors, the other half are not. This test has been shown to be effective in evaluating the print exposure of L1 English university students, as demonstrated in a study where ART scores predicted word-encoding times as measured through eye-tracking (Moore & Gordon, 2015), and another linking ART scores with verbal ability (Acheson, Wells, & MacDonald, 2008). However, ART has not been proven

to be equally useful for L2 English students, or for students enrolled in college programs.

Deciding whether or not to pursue post-secondary studies is an important decision, yet higher education can be a time-consuming and expensive proposition—the latter especially for international students, who are required to pay significantly more expensive tuition fees. Students who choose to pursue a college certificate or undergraduate degree typically delay entry into the workforce by two to four years, and often incur significant student debt along the way. Many believe these trade-offs can be worthwhile, as a college or university level education is thought to improve not only one's job prospects, but also one's cognitive skills, particularly in the domains of critical thinking and reading ability. While a post-secondary degree may be a prerequisite for many high-paying jobs, it is more difficult to quantify how it impacts an individual's ability to read, think, and reason. After all, university students are tasked with reading a large quantity of academic literature throughout the course of their studies, and are required to think critically, engage thoughtfully with the material, and expand their perspectives. Consequently, it would be valuable to be able to measure the impact of each year of the post-secondary intervention itself on reading ability, as this is not only an important skill in its own right, but a reasonable proxy for how successful a graduate may be at thinking critically and engaging with complex ideas in the future (for evidence on the relationship between critical thinking and reading ability, see Bobkina & Stefanova, 2016; Facione, 1990).

With the previous information in mind, the primary goals of the present research are to evaluate the following:

1) The degree to which the Author Recognition Test (ART) provides a useful measurement of print exposure in college and L2 English-speaking students as compared to university and L1 English students (Chapter 2)
2) The effect of each year of study on reading skills, and how this effect varies between L1 and L2 speakers of English (Chapter 3)

To address the first question, Item Response Theory (IRT) is employed. IRT is a psychometric method of assessing the informative value of each test item in discriminating the proficiency of respondents on some latent ability. In the case of ART, these individual test items are in the form of individual author names, and the latent ability to be measured is print exposure. As we will see in Chapter 2, these author names are of varying degrees of renown, such that participants with a standard degree of print exposure might reasonably be expected to recognize many but not all of the author names presented. Knowledge of lesser-known (i.e. more obscure) author names are thus understood to be indicative of an individual with greater print exposure, and by extension, an individual who reads more than others. For the second question, the same dataset is used but the relationship between component skills of reading and L1 and L2 English students over each year of study is examined.

Functioning and succeeding in the modern labour market requires a confident level of literacy, and better reading skills can even increase earning potential (Green & Riddell, 2001; Hanushek, Schwerdt, Wiederhold, & Woessmann, 2013). In addition to this economic advantage, there is a social prestige associated with being "well-read". As previously mentioned, one consequence of this is that surveys about reading habits can be confounded by biased self-reporting, with many participants having an inflated concept of how much and how often they read. It

is important to recognize that as researchers, we bear many of the same predispositions towards reading behaviour, and these can be reflected in our reporting.

In this thesis, the findings are presented as accurately and objectively as possible. However, regression tables and mathematical models do not exist in a vacuum. As is often the case in science, the hard numbers require some degree of subjective interpretation. To this end, I have attempted to provide a reasonable analysis of the results in the context of the importance and value of literacy skill in our society. By way of analogy, consider the distinction between "story" and "plot" as defined by the English novelist E.M. Forster:

> We have defined a story as a narrative of events arranged in their time-sequence. A plot is also a narrative of events, the emphasis falling on causality. 'The king died and then the queen died' is a story. 'The king died, and then the queen died of grief' is a plot. [...] Consider the death of the queen. If it is in a story we say: 'And then?' If it is in a plot we ask: 'Why?' (1985, p. 86)

Put simply, I have tried to be a reliable and responsible narrator who has built a convincing plot in service of the arguments presented. I hope that the reader may ask themselves, "why?", and that my interpretation may offer some answers. As is often the case, the facts are essential, but left alone they do not tell the whole story.

Chapter 2

Is the Author Recognition Test a Useful Metric for Native and Non-Native English Speakers? An Item Response Theory Analysis

2.1 Abstract

Studies of reading have shown the "Matthew Effect" of exposure to print on reading skill: poor readers avoid reading and ability develops more slowly compared to peers, while good readers improve more quickly through increased exposure. Yet it is difficult to determine just how much an individual reads. The Author Recognition Test (ART, Stanovich & West, 1989) and its multilingual adaptations are often used for quantifying exposure to print and have shown high validity and reliability in proficient readers in their dominant language (L1). When studying bilingualism and second language acquisition, it is ideal to have a single test which is equally reliable for all cohorts for comparison, but it is unclear if ART is effective

for speakers of English as foreign language (L2). This study assesses the reliability of ART in an English-medium university and college students with different language backgrounds. Following Moore & Gordon (2015), we use Item Response Theory (IRT) to determine how informative the test and its items are. Results showed an expected gradience in ART performance, with L1 speakers showing higher scores than L2 speakers of English, university students showing higher scores than college students, and both cohorts performing better than students in an English as a second language (ESL) university pre-admission program. IRT analyses further revealed that ART is not an informative measure for L2 speakers of English, as most L2 participants show a floor effect. Reasons for this unreliability are discussed, as are alternative measures of print exposure.

2.2 Introduction

It is a long-standing observation that reading proficiency stands in a reciprocal causal relation to the amount of reading an individual undertakes in their free time (McQuillan & Au, 2001; Mol & Bus, 2011; Paulson, 2006). Naturally, how much voluntary reading a person does is at least partially influenced by their attitude toward reading as a pastime (Kush, Watkins, & Brookhart, 2005). At the earliest stages of reading development, the ability to decode new words using phonological knowledge is a particularly important skill which sets apart higher and lower-skilled readers (Perfetti, Beck, Bell, & Hughes, 1987; Tunmer & Nesdale, 1985). This ability gap leads to a "rich-get-richer and poor-get-poorer" phenomenon, whereby lower-skilled readers, discouraged or unable to derive enjoyment from reading, are slow to develop reading skills and expanded vocabulary knowledge,

whereas skilled readers consume literary material voraciously and thus reap the benefits. This bidirectional relationship is often referred to as a "Matthew Effect" for reading (Kempe, Eriksson-Gustavsson & Samuelsson, 2011; Stanovich, 1986; however, for a discussion on possible mediating factors see also Bast and Reitsma, 1998; Pfost, Hattie, Dörfler & Artelt, 2014; and, for additional discussion on the issues surrounding psychometric analyses of a possible Matthew effect for reading see also Protopapas, Parrila, and Simos (2016) and Protopapas, Sideridis, Mouzaki, and Simos (2011)). Understanding this relationship makes the operationalization and measurement of the amount of reading by an individual (their "exposure to print") an important goal for the study of reading.

One possibility for assessing an individual's exposure to print is to administer questionnaires which collect subjective judgments from respondents on the amount, genre diversity, or complexity of reading that they do, as well as an evaluation of their own reading proficiency. An example of this is the Reading Habits questionnaire developed by Acheson et al. (2008). In this self-evaluation, participants are asked to report how much time they spend reading and writing in an average week, as well as whether they thought they read more or less than their peers. Similar questionnaires have been administered for developmental college students (Sheorey & Mokhtari, 1994), young readers in England (Cain & Oakhill, 2011), teaching candidates (Benevides & Peterson, 2010), and a comparison between White and Asian Americans (Scales & Rhee, 2001). Self-reported evaluations of this kind, however, run the risk of interference from social desirability factors, which may influence respondents to overstate the breadth of their reading habits.

A complementary method of establishing how much an individual reads is through proxy tests designed to determine their level of print exposure. Perhaps

the best-known test of exposure to print is the Author Recognition Test (ART), first developed by Stanovich and West (1989) and which has been demonstrated to predict orthographic processing. This test presents participants with a list of author names and distractor ("foil") names, but which could nevertheless plausibly be believed to be names of real-world authors. Participants are asked to identify only which names belong to real authors by indicating with a checkmark, and to ignore any names which are not believed to be those of published authors. The resulting score is calculated by subtracting the number of foils incorrectly selected from the number of correct responses. Similarly, a magazine recognition test (MRT) has also been used for measuring one's knowledge of magazine titles (Stanovich & West, 1989) as a means of assessing print exposure through more popular media. Another related measure called the Title Recognition Test (TRT) assesses knowledge of book titles (Cunningham & Stanovich, 1990). This method has been used to compare disabled vs. non-disabled grade school readers (McBride-Chang, Manis, Seidenberg, & Custodio, 1993).

The linking hypothesis of such tests is that the quantity of reading materials (e.g., books or magazines) that one is exposed to correlates strongly with one's reading skill (Mol & Bus, 2011; Weinberger, 1996). Importantly, the tests do not assume that respondents have read the specific authors, magazines or books about which they are queried. Instead, the assumption is that a greater amount of reading leads to a greater awareness of the existing literature and reading sources, which translates into higher recognition scores. This greater awareness and its ancillary benefits have been referred to as "cultural capital" (Bourdieu & Richardson, 1986; Tunmer & Chapman, 2006), discussed in more detail below.

A large body of research has confirmed the ART's usefulness as a predictor of proficiency in lexical tasks: as will become important below, most of this work

involved university-level native readers of English. For example, scores on the ART have been shown to correlate positively with vocabulary size (e.g., Krashen & Kim, 1998; Lee, Krashen, & Tse, 1997; Martin-Chang & Gould, 2008; Rodrigo, McQuillan, & Krashen, 1996; West & Stanovich, 1991), speed or accuracy of reading words, sentences or passages (e.g., Acheson, Wells, & MacDonald, 2008; Choi, Lowder, Ferreira & Henderson, 2015; Kuperman, Matsuki, & Van Dyke, 2019; Martin-Chang & Gould, 2008; Moore & Gordon, 2015), as well as reading comprehension (Cipielewski & Stanovich, 1992; Landi, 2010). Furthermore, Unsworth & Pexman (2003) found that those who scored higher on the ART did not show regularity effects in lexical decision and phonological lexical decision tasks, suggesting they had better mental access to phonological information when reading. The ART is also popular because it can be administered in only a matter of minutes, making it one of the fastest ways to ascertain an individual's approximate level of reading proficiency.

Given the popularity and demonstrated validity of the ART for proficient native readers of English, much work has been conducted to establish its reliability in this population. For example, Acheson, Wells, & MacDonald (2008) used the student population of the University of Wisconsin to determine that the original ART developed twenty years prior contained many names of authors who were no longer well-known to college students. In response, they developed a new version of the ART which consisted of 130 items (65 author names and 65 foils) and kept only 15 of the original author names. A more recent psychometric study of 1,012 students at the University of North Carolina by Moore and Gordon (2015) used Item Response Theory (IRT) to evaluate the discriminative value of each item on the ART, i.e. how much a correct or incorrect response to each author name and each foil distinguishes readers with different levels of exposure to print (see detailed

discussion below). The outcome of this study was the determination of a discriminative value of each item and a proposed reduction of the ART item list from 130 (65 author names and 65 foils) to 100 of the most discriminative items (50 author names and 50 foils). Moore and Gordon (2015) evaluated the reliability of the abridged version of ART and confirmed its validity as a predictor of eye-movements registered during reading for comprehension (N=789, all correlations between mean gaze duration and the 100-item ART score highly significant at $p <$ .001), see also Choi et al., 2015 and Kuperman et al. (2019).

Furthermore, researchers have created ART versions for Hebrew (Shatil, Share, & Levin, 2000), Dutch (Vander Beken & Brysbaert, 2018), Korean (Kim & Krashen, 1998a; Lee et al., 2018) and Chinese (Chen & Fang, 2015), as well as those intended specifically for readers in the United Kingdom (Masterson & Hayes, 2007) and Canada (Chateau & Jared, 2000; Sénéchal et al., 1996). Additionally, ARTs for English-speaking children (sometimes called a Children's Author Recognition Test or CART) have been developed and implemented (Cipielewski & Stanovich, 1992; Ricketts, Nation, & Bishop, 2007; Stainthorp, 1997). Self-administered versions of the ART have even been created which contain no foils and were shown to still be a strong predictor of vocabulary size despite participants being informed that all names listed were of published authors (Krashen & Kim, 1998).

It can be confidently stated that the ART is a reliable and valid method for examining native speakers of English at the university level, and that it is adaptable to other languages and other populations. What is less certain, however, is whether or not the ART can be used *without adaptation* as a reliable tool in studies that involve comparisons between native and non-native readers of English, or between individuals widely different in their English reading proficiency – from university

students to college students[1] to students enrolled in ESL classes. Such comparisons are essential for answering a number of critical theoretical and practical questions in psychology and education. Thus, these fields need a tool that can be used as a uniform instrument, for comparing exposure to print across populations with variability in reading skills in their L1 and L2. The efficacy of the ART in evaluating these additional populations compared to native English university students will be the primary question of interest for the present paper.

There are indeed examples of studies which have used the ART and similar methods of evaluating print exposure in native English-speaking cohorts with different proficiency levels. These include unskilled vs. skilled child readers (Ricketts, Nation, & Bishop, 2007), older vs. college-age readers (Stanovich, West, & Harrison, 1995), and high vs. low-skilled postsecondary readers (Lewellen, Goldinger, Pisoni, & Greene, 1993). Moreover, the English version of the ART has also been used to predict literacy skill in second-language learners of English (Kim & Krashen, 1998b; McQuillan, 2006; Miller Guron & Lundberg, 2003; Stuart, 2004). Importantly, this comparative research effort may be jeopardized if the tool used for assessment is unreliable for use in at least some populations under comparison. To our knowledge, no systematic psychometric analysis has been conducted to test the ART's reliability in native English speakers with a below-university level of

[1] College is defined in Canada as non-degree-granting postsecondary institution with one- or two-year programs of studies.

reading proficiency, or non-native readers of English. Our study aims to fill this gap.

It is sensible to expect that the ART's reliability will vary across populations. One reason for this is that the ART item selection is most representative of fiction authors writing in English (e.g. James Joyce, Ernest Hemingway, etc.) and more generally authors belonging to the Western literary tradition, if not necessarily those writing in English (e.g. Umberto Eco). In the case of non-native speakers of English, it is a distinct possibility that the materials they predominantly read are either not in English, or – if they are in English – are not represented in the item selection of the ART. Second, individuals' reported time spent reading academic textbooks and fiction are known to be negatively correlated (Acheson, Wells, & MacDonald, 2008): for non-native speakers enrolled in an English-medium educational program, their assigned reading may be largely academic and their exposure to fiction may be comparatively limited. This generally results in lower performance on a test like ART due to the specificity of the English reading material to which L2 readers are exposed. In this case, differences in the ART scores may emerge because the English-language ART may not equally tap into the cultural capital of diverse populations.

The goal of this paper is to assess the reliability of the ART in five samples: (i) native English university-level readers; (ii) native English college-level readers; (iii)-(iv) non-native university- and college-level readers; and (v) non-native English readers enrolled in an ESL year-long pre-admission program at a university. This coverage enables us to assess how the ART's reliability is influenced by both the language background and variability in educational level, roughly equivalent to variability in reading proficiency.

2.3 Methods

2.3.1 Participants

A total of 1,107 students participated in the study between September 2017 and April 2019: 656 university students recruited from a convenience pool of McMaster University and 451 college students from a similar pool of Mohawk College of Applied Arts and Technology (both institutions located in Hamilton, Ontario, Canada). Participants represented five cohorts defined by varying levels of English proficiency and education, defined as (i)-(v) above. Table 2.4 reports sample sizes of all cohorts.

Table 2.1: Average ages at which cohorts learned English.

Level of Education	English	Median	Mean	Minimum	Maximum	Range
College	L1	0.00	0.19	0.00	9	9.00
	L2	7.00	8.63	0.00	39	39.00
University	ESL	8.00	8.08	0.00	18	18.00
	L1	0.00	0.03	0.00	6	6.00
	L2	6.00	6.05	0.00	22	22.00

Table 2.2: Average ages at which cohorts arrived in Canada.

Level of Education	English	Median	Mean	Minimum	Maximum	Range
College	L1	0.00	1.59	0.00	46	46.00
	L2	20.00	19.31	0.00	48	48.00
University	ESL	16.00	17.00	0.00	50	50.00

	L1	0.00	0.34	0.00	17	17.00
	L2	14.50	10.92	0.00	41	41.00

As part of the collection of demographic data, we asked participants to provide information about their language experience, including their first language, country of birth, and number of years spent in Canada. Mohawk College students were compensated by participating in a lottery that randomly distributed twenty $50 gift cards for the college's bookstore, and McMaster students were given a partial course credit. The study received the ethics clearance from the McMaster REB (2018-033) and Mohawk REB (18-003).

Table 2.3: Most frequently spoken first languages in university and college samples.

College			University	
Number	Language	Frequency	Language	Frequency
1	English	317	Chinese	268
2	Punjabi	90	English	191
3	Gujarati	64	Urdu	11
4	Chinese	29	Russian	7
5	Hindi	23	Arabic	5
6	Vietnamese	21	Punjabi	5
7	Arabic	20	Korean	4
8	Spanish	18	Spanish	4
9	Malayalam	16	French	3
10	Portuguese	15	Gujarati	3

2.3.2 Materials

All university samples (L1, L2 and ESL) completed the 130-item version of the ART published in Acheson et al. (2008), whereas the college samples (L1 and L2) were administered the 100-item version of the ART from Moore & Gordon (2015). To make the outcomes comparable, we later re-scored the 130-item ART to match the outcomes of the 100-item test version as follows. First, we only considered those 50 fiction authors (out of 65) that were included in Moore and Gordon's abridged ART due to their informativity. Second, the foils were different in the 130-item and the 100-item ARTs. We randomly selected 50 foils (out of 65 in the 130-item ART) and only considered participants' responses to those foils. The scores obtained by each university-based sample in the full 130-item ART and its 100-item subset defined above correlated at $r > 0.9$ (all $ps < 0.001$). We conclude that our reduction of the 130 items to 100 is highly representative of exposure to print in all university-based samples.

2.3.3 Procedure

Participants began with providing informed consent and then responded to a demographic questionnaire, including information about their age, education, first language, as well as their subjective estimate of reading, writing, listening, and speaking proficiency in English. In all cohorts, ART was part of a bigger battery, which we do not report here. The entire experimental session did not exceed 15 minutes in the college samples and 1 hour in the university samples.

For the ART component of the battery of tests, participants were presented with a checklist of names and were asked to check off only those names which they were certain belonged to a published author. Instructions to participants were the

same as in Acheson et al. (2008). An individual ART score was calculated as follows: every correct indication of an author increased the score by 1 point; every incorrect indication decreased it by 1; no penalty was incurred for not indicating an existing author. At-chance performance yields a score of 0; negative scores are possible as well.

All analyses were performed in statistical software R 3.6.1 (2019) and the IRT analysis was performed using the package ltm (Rizopoulos, 2006).

2.4 Results

Table 2.4 summarizes the distribution of the ART score per cohort. Results showed that native speakers in college perform more poorly in the ART task than those in university, and ESL students perform more poorly than non-native cohorts in both university and college (all ps < 0.05 in two-sample t-tests after Bonferroni correction for multiple comparisons). All cohorts showed ART scores that were reliably different from the chance level of 0 (all ps < 0.05 in one-sample directional t-tests). We also note that the ART score registered in our cohort of L1 university students is significantly weaker than that reported by Moore and Gordon (2015) in a cohort of 1,102 students of University of North Carolina at Chapel Hill (9.61 vs 13.57, t = - 7.57, p < 0.001). Thus, regional differences between Canada and the USA in ART performance are possible as well.

Table 2.4: Descriptive statistics of performance of all cohorts on ART, including cohort size N.

Level of Education	English	N	Mean	SD	Minimum	Maximum	Range
College	L1	283	7.63	9.76	-8	47	55
	L2	219	2.92	7.73	-9	47	56
University	ESL	154	1.18	6.53	-8	40	48
	L1	338	9.61	8.78	-10	45	55
	L2	113	5.38	7.12	-5	31	36

The raw scores in Table 2.4 clearly demonstrate that average performance on the ART varies by cohort, however it is also important to understand the degree of similarity between cohorts with respect to the authors they tend to recognize. Table 2.5 reports accuracy of recognition (percent correct) for every author in each cohort (Columns %), sorted in the decreasing order of accuracy in the L1 University cohort. We conducted a correlational analysis that compared the rank-order of accuracy of author recognition across cohorts. A high correlation means that in the two cohorts under comparison, the individual participants tend to more frequently recognize the same authors, as well as less frequently recognize others. A low correlation means that readers from different cohorts also vary in the authors they know, and not only the overall levels of familiarity with the authors.

Table 2.5: Comparison of response accuracy and estimated IRT parameters across five cohorts. Column % indicates how frequently each name was correctly selected as an author, a indicates the level of *discrimination* of individual ability each name provides, while β indicates the *difficulty* of correctly selecting each name.

#	Author Name	University, English L1			University, English L2			University, ESL			College, English L1			College, English L2		
		%	α	β	%	α	β	%	α	β	%	α	β	%	α	β
1	Stephen King	84.45	1.71	-1.55	49.70	1.86	-0.01	23.57	1.15	1.10	78.80	0.74	-2.15	42.47	0.88	0.33
2	F. Scott Fitzgerald	81.50	1.25	-1.63	53.25	1.72	-0.13	29.30	0.96	0.92	54.77	1.48	-0.38	27.85	1.49	0.81
3	Ernest Hemingway	80.97	0.95	-1.90	60.95	1.43	-0.43	39.49	0.84	0.42	42.40	1.71	0.05	17.35	1.67	1.28
4	Margaret Atwood	75.87	0.96	-1.52	44.97	3.55	0.09	19.11	2.16	0.92	43.11	2.19	0.00	16.89	1.42	1.44
5	Harper Lee	75.07	1.34	-1.19	50.89	2.14	-0.05	26.75	1.57	0.73	45.94	1.23	-0.03	25.11	1.14	1.13
6	E. B. White	63.54	1.11	-0.73	31.36	1.94	0.62	31.21	1.37	0.60	34.98	1.64	0.33	21.46	1.59	1.09
7	Virginia Woolf	63.54	1.23	-0.69	41.42	1.98	0.26	37.58	1.13	0.39	31.45	2.35	0.35	11.87	1.76	1.61
8	J. R. R. Tolkien	61.93	1.13	-0.65	42.01	2.21	0.22	36.31	1.42	0.37	60.78	1.23	-0.65	31.05	1.49	0.67
9	George Orwell	60.32	1.66	-0.47	39.64	2.32	0.29	21.02	2.54	0.78	50.18	1.77	-0.21	36.53	0.91	0.65
10	Maya Angelou	56.03	0.83	-0.45	42.60	1.19	0.32	18.47	2.26	0.93	19.79	2.07	0.85	15.07	1.92	1.32
11	James Patterson	55.23	1.58	-0.29	26.63	1.85	0.83	16.56	1.51	1.29	41.70	1.54	0.10	19.63	1.30	1.35
12	T.S. Elliot	55.23	1.17	-0.34	33.73	1.80	0.55	25.48	1.17	0.98	40.28	2.03	0.10	30.59	1.70	0.64
13	J. D. Salinger	53.62	1.70	-0.23	31.36	1.91	0.62	35.03	1.20	0.49	39.58	1.86	0.13	21.92	1.64	1.04
14	William Faulkner	44.24	2.47	0.08	23.08	2.84	0.81	23.57	2.05	0.75	23.32	1.92	0.73	14.61	1.56	1.52
15	Tom Clancy	42.90	1.26	0.18	26.04	2.46	0.74	21.02	2.10	0.85	49.82	1.16	-0.19	18.26	1.62	1.25
16	Danielle Steel	39.68	1.59	0.26	18.93	2.38	1.07	17.83	1.94	1.04	34.28	1.69	0.34	19.63	1.95	1.06
17	Ray Bradbury	39.14	1.73	0.27	25.44	1.66	0.94	22.29	2.17	0.78	19.08	2.61	0.77	10.96	2.55	1.43
18	Thomas Wolfe	38.87	1.42	0.32	21.89	1.47	1.20	17.83	2.07	1.01	27.92	2.54	0.45	12.33	1.78	1.57
19	John Grisham	37.80	1.14	0.43	26.63	1.07	1.18	23.57	1.29	1.01	26.86	2.40	0.50	17.81	1.91	1.17
20	Samuel Beckett	36.46	2.07	0.33	18.34	2.46	1.09	14.01	2.29	1.15	20.85	2.29	0.75	13.70	1.72	1.50
21	Kurt Vonnegut	36.19	1.91	0.35	24.85	2.52	0.78	20.38	2.04	0.89	14.84	2.57	0.98	9.59	2.54	1.53
22	James Joyce	34.85	1.70	0.43	26.04	1.80	0.87	23.57	2.09	0.74	19.08	2.21	0.84	10.96	1.94	1.60

23	Gabriel Garcia Marquez	34.32	1.76	0.44	24.85	1.68	0.96	21.02	1.21	1.21	9.54	5.40	1.01	13.70	1.36	1.72
24	Sue Grafton	32.17	0.83	0.90	24.26	1.03	1.37	18.47	1.97	1.00	14.13	2.66	1.00	10.50	2.09	1.59
25	Toni Morrison	30.83	1.35	0.67	23.08	1.84	1.00	24.84	1.77	0.76	16.25	3.00	0.84	15.53	1.69	1.39
26	Anne McCaffrey	29.22	1.32	0.75	18.34	2.32	1.12	24.20	1.11	1.10	17.31	2.14	0.95	13.24	2.70	1.25
27	Vladimir Nabokov	28.42	1.61	0.69	25.44	2.12	0.82	22.29	2.20	0.78	17.31	2.54	0.86	14.61	2.14	1.28
28	Isaac Asimov	27.08	2.26	0.63	20.71	3.34	0.87	16.56	1.80	1.16	16.96	3.13	0.79	14.16	2.36	1.26
29	Nora Ephron	25.47	2.42	0.66	14.20	3.23	1.22	14.65	1.92	1.22	13.07	3.12	0.99	11.42	1.67	1.70
30	Ralph Ellison	25.20	2.32	0.69	13.61	2.35	1.40	17.83	1.48	1.23	11.66	2.20	1.28	13.70	1.67	1.52
31	Ayn Rand	24.66	1.90	0.78	16.57	2.46	1.19	15.29	3.22	0.95	19.79	2.66	0.73	10.05	3.04	1.41
32	Judith Krantz	24.40	2.07	0.75	13.02	1.40	1.88	21.02	1.11	1.29	9.54	3.02	1.25	10.05	2.08	1.63
33	Robert Ludlum	24.13	1.25	1.04	23.08	1.04	1.43	17.83	2.15	0.99	15.90	3.67	0.79	10.96	2.79	1.38
34	Michael Ondaatje	23.32	2.34	0.75	17.16	3.26	1.05	19.75	1.76	1.00	10.60	3.93	1.05	10.50	2.62	1.44
35	Jack London	23.06	1.76	0.88	17.75	1.32	1.56	25.48	1.65	0.77	16.61	3.40	0.78	12.33	1.77	1.57
36	Raymond Chandler	22.79	1.91	0.85	10.65	3.35	1.46	17.83	1.86	1.06	14.84	3.37	0.87	12.79	1.97	1.46
37	Salman Rushdie	22.79	1.40	1.03	15.98	1.70	1.46	26.11	1.44	0.81	11.66	3.33	1.05	12.33	1.62	1.65
38	Kazuo Ishiguro	22.52	1.52	0.99	15.98	1.77	1.43	14.65	1.65	1.34	9.89	4.00	1.09	7.76	1.83	1.96
39	Clive Cussler	21.98	1.21	1.18	21.89	1.14	1.42	24.84	1.79	0.75	19.08	2.95	0.72	9.59	3.95	1.35
40	Joyce Carol Oates	21.72	2.22	0.83	17.16	1.87	1.31	15.29	2.05	1.15	15.19	3.47	0.84	12.33	2.79	1.29
41	Isabel Allende	20.11	2.31	0.88	15.38	3.32	1.14	16.56	1.92	1.12	9.19	2.70	1.35	9.59	2.13	1.65
42	Willa Cather	19.30	2.70	0.86	15.38	1.48	1.63	15.92	2.32	1.05	8.13	2.93	1.41	11.42	2.02	1.54
43	Bernard Malamud	18.77	1.13	1.45	14.79	1.56	1.62	23.57	2.21	0.72	9.89	2.96	1.24	10.50	1.64	1.80
44	James Michener	17.96	1.81	1.10	17.16	2.01	1.27	15.92	1.18	1.58	8.83	4.39	1.14	9.13	1.56	1.98
45	Thomas Pynchon	17.43	2.70	0.94	13.61	2.48	1.37	22.93	2.12	0.76	12.72	3.51	0.96	11.87	2.22	1.44
46	Nelson DeMille	17.16	2.96	0.92	12.43	2.40	1.48	16.56	1.34	1.40	15.19	3.38	0.85	12.79	1.73	1.56
47	Jane Smiley	16.89	2.05	1.08	10.06	2.08	1.79	17.83	2.04	1.02	11.31	3.40	1.06	9.13	1.82	1.82
48	Saul Bellow	15.28	2.89	1.01	11.83	3.32	1.37	15.92	2.32	1.05	9.89	3.20	1.19	7.31	2.60	1.74
49	Herman Wouk	13.94	1.71	1.37	14.20	2.09	1.44	21.66	1.47	1.02	10.25	3.88	1.08	11.87	2.30	1.42
50	Umberto Eco	13.94	2.96	1.06	14.79	2.99	1.21	12.74	1.85	1.37	8.83	3.34	1.27	12.33	2.60	1.32

Table 2.6 shows rank-order correlations between percent correct across our five samples. Spearman correlation coefficients are reported above the diagonal, and p-values below the diagonal. The very high correlation between L1 and L2 English university students in Table 2.6, alongside the difference in average scores in Table 2.4, demonstrates that although L2 speakers of English perform lower overall on the ART, there is considerable overlap in the authors they are most familiar with. A similar conclusion can be drawn from the comparisons of L1 and L2 college-level speakers of English. In contrast, the much lower correlation between ESL students and all other cohorts suggests that their responses were more random, which is supported by their comparatively lower average ART score.

Table 2.6: Rank-order correlations between percent correct across five samples. Spearman correlation coefficients are reported above the diagonal, and p-values below the diagonal.

Level of Education	English	University			College	
		L1	L2	ESL	L1	L2
University	L1	*****	0.902	0.516	0.867	0.736
	L2	<0.001	*****	0.575	0.810	0.665
	ESL	<0.001	<0.001	*****	0.499	0.446
College	L1	<0.001	<0.001	<0.001	*****	0.771
	L2	<0.001	<0.001	0.001	<0.001	*****

To evaluate the validity of ART for different cohorts, we employed Item Response Theory (IRT), see Moore and Gordon (2015). IRT determines the informative value of each entry in a given test for discriminating the latent ability of participants (estimated as a function of the participant's test score) in some

dimension (here; exposure to print, and consequently, reading ability), as well as estimating the overall informativity and the measurement error of the test itself overall (see Embertson & Reise, 2013). To reiterate, our critical question was how valid a test ART is for cohorts widely differing in their exposure to print and reading ability.

When using IRT, up to three separate parameters can be selected—item difficulty (α), item discrimination (β), and the likelihood that an individual is guessing on a given item (An & Yung, 2014; Slinde & Linn, 1979). These parameters are estimated from statistics of responses to test items, and researchers can evaluate how many parameters to use to create the best-fitting model. In the present case, a two-parameter IRT model (accounting for both item difficulty and item discrimination) provided a better fit to the data than the one-parameter alternative and the Rasch model,[2] as indicated by the likelihood ratio model comparison test. This model returned the difficulty parameter (reflected in the left vs rightward shift along the x-axis representing latent ability) and the discrimination parameter of the Item Characteristic Curve for every item in the test. The metric of difficulty or the β parameter of the model for an individual item (author name) is the estimated level of ability at which an individual would have a greater than 50% probability of correctly responding to this item (recognizing this author). Smaller values of β

[2] The Rasch model is a dichotomous one-parameter IRT model which assumes that all items have the same discrimination or slope, and are differentiated only by the single parameter of item difficulty (An & Yung, 2014; Rasch, 1960; Slinde & Linn, 1979).

represent lower difficulty. The discrimination a parameter is a slope of the line fitted to the Item Characteristic curve: the steeper it is, the better this item discriminates between responders who give a correct vs incorrect response to this item. In other words, it creates a more clearly-defined boundary between those who do and do not know a particular author—all respondents below this level of latent ability (i.e. print exposure) will know the author, and those with lower levels will not. Additionally, with a steep slope, smaller changes in ability level create larger changes in the probability of selecting a given author name when compared to a flatter curve.[3]

We fitted the IRT two-parameter models to the ART data of each cohort separately. Table 2.5 reports the outcomes, comparing all five cohorts by author name for percentage correct, the a parameter (discrimination), and the β parameter (difficulty), sorted by the percentage each name was correctly selected by University English L1 speakers.

Unsurprisingly, in Table 2.5 we see that names of authors who are well-known to North American audiences top the list of those most likely to be selected by University English L1 speakers (e.g. Steven King, F. Scott Fitzgerald, Ernest Hemingway, or Harper Lee. These are frequently names of authors commonly found in North American public-school reading curricula). We also note a relatively high rank of Margaret Atwood in our current Canadian data from L1 university students (rank 4), as compared to that (rank 25) from University of North Carolina (Moore

[3] See Moore and Gordon (2015) for a detailed description of the two-parameter models.

& Gordon, 2015). Margaret Atwood is one of the best known contemporary Canadian fiction writers, and the better recognition of her name in Canada than in the USA aligns with earlier reports of the link between changes in author popularity and ART scores.

In fact, the β values show that the native-English speaking university students have an advantage over other cohorts as these values tend to be lower for them: e.g., the 13 most recognized authors show negative values of difficulty (β) for the University L1 cohort, while the β values for College L2 and English ESL students never drop to negative values. In concrete terms, what this means is that even a University L1 student with a lower latent ability level will still have a 50:50 chance of correctly selecting the names of these more commonly-known authors, whereas College L2 and English ESL students need a higher ability level in order to perform at the same level for the same authors.

For University English L1 speakers, names with higher a parameter values, such as Nelson DeMille, Umberto Eco, Saul Bellow, and Herman Wouk are more discriminative of latent ability. Put simply, a native-English university student who is familiar with these more obscure names can be inferred to have been exposed to more print in their lifetime, and consequently, is more likely to be a more proficient reader.

To get a better sense of the unequal distribution of difficulty of the test across cohorts, Figure 2.1 visualizes the item estimates of the IRT model for each cohort respective Item Characteristic Curves. Each line in each figure represents an individual author's name from the Author Recognition Test, and the dashed line shows the point at which a participant has a greater or lower than chance probability of correctly selecting the author's name according to the participant's latent ability (print exposure). These figures illustrate how a broad range of levels

of print exposure can be informative for Native English-speaking University students (top left), whereas in contrast, a very narrow band of name informativity is found for cohorts such as the University ESL Program (middle left). This low spread for the ESL readers shows all items are roughly equally difficult, leading to a floor effect. In general, a rightward shift of curves is observed for most non-native and college students, reflecting the relatively consistent difficulty of most ART items for these cohorts.

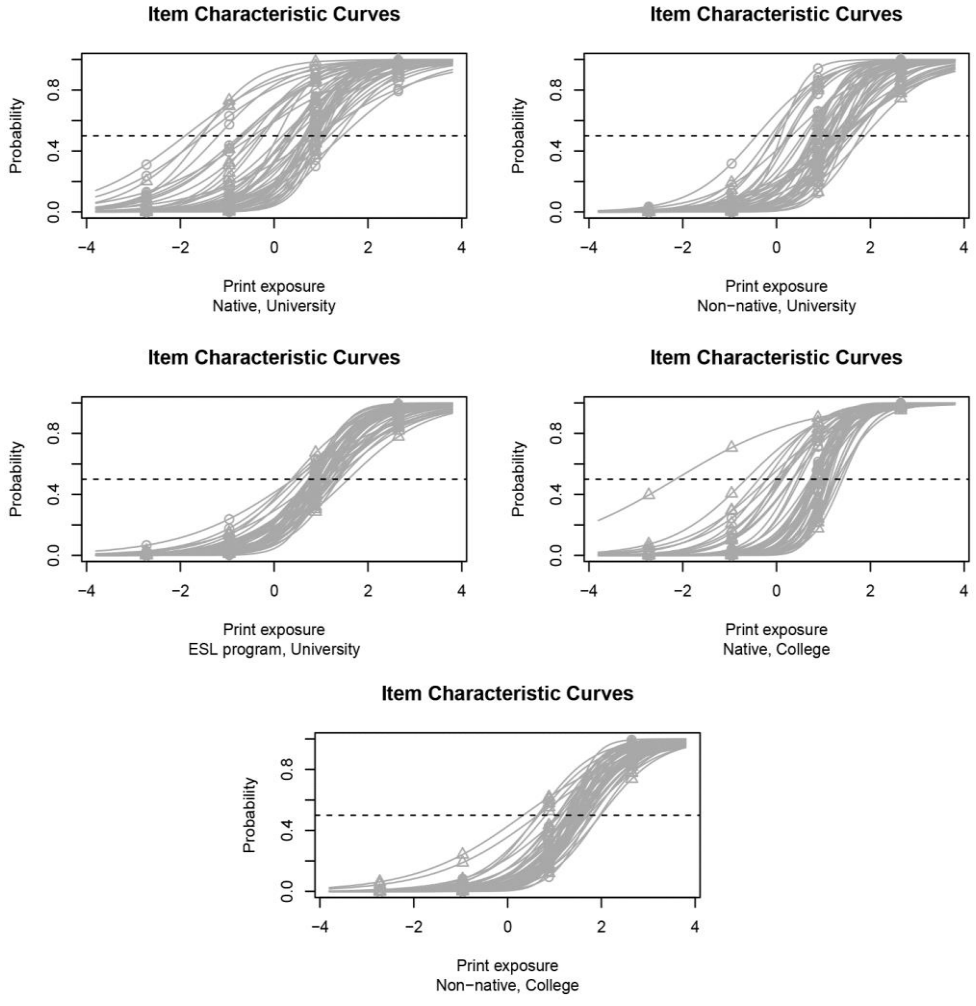

Figure 2.1: Item Characteristic Curves comparing all five cohorts.

An important dimension of a test's validity is its measurement error. Figure 2.2 shows a comparison of Standard Errors across all five cohorts, derived from respective IRT analyses. Item Response Theory enables us to estimate the standard error of measurement associated with each cohort: Figure 2.2 visualizes the estimates as a function of latent ability. These estimates are given on the original scale, i.e. points in the ART score. As expected for a test that appears to demonstrate a floor effect, it is more accurate (has a lower standard error) at the higher levels of latent ability than the lower ones. With the exception of the University L1 cohort where standard errors are equally low on the two extremes of ability, standard errors were much higher in the lower range of latent ability in all remaining cohorts.

The variability in the lower range of ability is massive. While the maximum standard error for L1 university students is on the order of 1 point (compare to the mean of 9.81), for L1 college students it reaches 2 points (with the mean of 7.63 points). Even more drastically, the maximum standard error for L2 college students is ± 4 points (compare to the mean score of 2.92). For ESL participants, the estimate of SE is ± 3 (and the mean is 1.98 points), and for L2 university students the standard error is ± 3.8 (with the mean of 5.38 points). In other words, the measurement is far less precise for these groups, as there is reduced confidence that the sample means are representative of their respective populations.

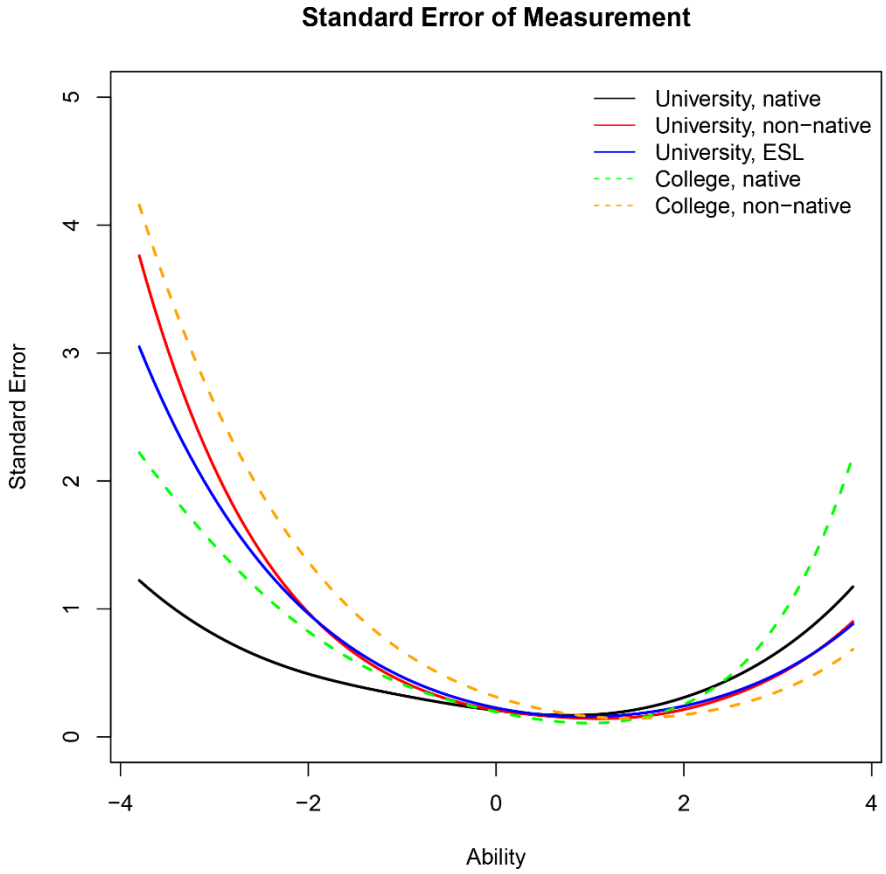

Figure 2.2: Comparison of Standard Errors across all five cohorts.

To sum up the findings, for many cohorts most author names on the ART are highly difficult to recognize, as reflected in the rightward shift of the Item Characteristic Curves in Figure 2.1. Additionally, for the L2 cohorts and the ESL cohort the spread of difficulty is extremely small, suggesting that most ART items are equally difficult, which indicates a floor effect. This is particularly the case for ESL and non-native college cohorts, who seem to be almost equally unfamiliar with all author names on the ART. Also, all cohorts except L1 university students demonstrated a very high standard error of measurement in the test, which makes the mean performance statistically indistinguishable from chance.

In accordance with multiple prior reports, we conclude that the ART is informative and accurate for L1 university students, and we add that it has a relatively high informativity and accuracy for L1 college students also. However, ART is not a reliable or informative tool for L2 students either at the university or college level, and it performs the worst for ESL students.

2.5 General Discussion

Exposure to print is one of the most robust predictors of reading performance and reading skill development (see meta-analysis of Mol & Bus, 2011). While primarily discussed for acquisition of literacy in one's dominant language it is equally evident that reading more in one's L2 increases the quality of one's L2 reading skill as well (Constantino, 1994; Constantino, Lee, Cho, & Krashen, 1997; Gradman & Hanania, 1991; Lao & Krashen, 2000; Mason & Krashen, 1997). Operationalization of exposure to print has a long history of eliciting subjective judgments of the quantity and quality of printed material that a person has access too (reviewed in Stanovich & West, 1989, among others). The introduction of the Author Recognition Test (ART) by Stanovich and West (1989) and derived checklist tests offered a measure that is objective, and one that has been demonstrated among proficient L1 readers as a valid and reliable estimate of one's reading experience. This paper examined validity of the ART (developed by Stanovich & West, 1989 and refined by Acheson et al., 2008 and Moore & Gordon, 2015) for readers of English as a first language (L1) and a second language (L2), across levels of English proficiency defined by the educational level (the English-medium university vs college vs ESL pre-admission program). The motivation was

to verify whether the same test can be meaningfully applied across all these groups as a comparator.

Our findings demonstrate an expected gradience in the ART performance across five cohorts (N = 1,107), with the decreasing order of performance being as follows: L1 university students, L1 college students, L2 university students, L2 college students, and ESL students. Importantly, all L2 speakers, and especially the groups with lower ART scores, performed virtually at chance level. The IRT analyses confirmed that ART is not informative for any L2 cohort and comes with a relatively high standard error of measurement. The measurement error was also high for L1 college students.

The conclusion is clear: in its current form, ART is not a meaningful test for L2 speakers of English. The factors which contribute to these results are less clear. Specifically, it is unknown if students with English as L2 are not reading a sufficient amount of material in any language for ART to be an effective measure of print exposure, or if instead these students are reading different kinds of authors (and in different kinds of languages) than the ones captured by ART. As mentioned in the Introduction, one potential weakness of ART is that the author names selected are all fiction writers generally belonging to the Western school of literature. If some L2 speakers read fiction representing a different tradition or do not read fiction at all, ART may not adequately tap into their "cultural capital" (Bourdieu & Richardson, 1986; Tunmer & Chapman, 2006). For instance, an adept of Chinese literature may not score a single point on ART because it does not include a single author from that tradition. As far as specific reading in English is concerned, L2 speakers of English enrolled in professional or academic programs may be focusing a greater deal of intellectual energy towards reading textbooks or technical manuals rather than consuming works of fiction. Again, ART would

underestimate their exposure to English print because it is primarily designed to capture reading for leisure (or "free voluntary reading") better reflected in reading fiction (Kim & Krashen, 1998b; Lee, Krashen, & Tse, 1997; West, Stanovich, & Mitchell, 1993).

In summary, these results underscore the importance of developing tests of English print exposure which are equally applicable to second language learners as they are to native speakers of English. The Author Recognition Test provides a helpful way of quickly determining exposure to print with a short checklist of items. However, given that the ART is of variable validity, new methods may be required which put native and non-native speakers of a given language on an equal footing. Preparing an author checklist that would be equally familiar to readers of many languages is not feasible. Adapting ART to a large number of languages is possible, as discussed in the introduction of this paper, but laborious. An additional possibility, which we explore in forthcoming studies, is to task participants with spontaneously naming as many authors as they can within a set period of time. This Author *Naming* Test will require recall and thus would differ from a recognition test like ART, but – with task instructions provided in a person's L1 – it will enable a speaker of any language tap into her cultural capital in a way unbiased by a specific extraneous literary, linguistic or cultural tradition. It may also provide a broader representation of a person's exposure to print by enabling her to identify both fiction authors and names of non-fiction authors, journalists, memoirists, and writers in additional genres. Granted, such a test would face additional challenges, including perhaps some form of external validation of the author names provided. Despite this, if an "author naming test" were to prove to be a reliable determiner of print exposure, it may represent a simpler way of accounting for the many different kinds of reading undertaken by an individual.

Chapter 3

Effects of Year of Post-Secondary Study on Reading Skills for L1 and L2 Speakers of English

3.1 Abstract

How does post-secondary education affect the development of proficiency in reading comprehension and related skills? How does this development vary between native (L1) and non-native (L2) speakers of English and how much change does each year of study bring about? The present cross-sectional study addressed these empirical questions by administering a battery of tests of reading proficiency to undergraduate students of a Canadian university in all years of study. The tests included measures of reading comprehension and multiple component skills of reading: vocabulary, spelling, print exposure, reading habits, and motivation. Results showed that year of study confers a direct effect on component skills of reading and an indirect effect on reading comprehension proficiency. Critically, skill

trajectories of L1 and L2 speakers of English varied widely between cohorts. While L2 speakers of English start in year 1 from a disadvantage in most skills, they close the performance gap with L1 speakers within 3-4 years. These findings quantify the effectiveness of the university intervention itself, but also its differential impact on student cohorts defined by their language background.

3.2 Introduction

Although Canadians are among the most highly educated people in the world, the Canadian Council on Learning (CCL, 2007) found that nearly half of the country's population lacks the minimum level of literacy skills for today's global labour market.[4] It is tempting to suggest that the problem is one of lack of education—for example, a U.S. literature review showed that those with a university degree have stronger verbal skills than those without, equivalent to approximately 10.3-12.6 percentile points on average (Pascarella & Terenzini, 2005). Yet even among university graduates, literacy performance is far from optimal. The CCL found that 20% of university graduates are below minimum functional literacy levels, and they anticipated this number would continue to climb

[4] That is, below level 3 on the five-level scale defined by the Adult Literacy and Life Skills (ALL) survey (Desjardins, Murray, Clermont & Werquin, 2005). Proficiency is classified into three separate literacy domains—prose, document, and quantitative. Respondents at Level 3 on the prose scale are expected to be able to make basic inferences and compare and contrast information across several different sentences and paragraphs.

(CCL, 2009). This finding was later echoed by the Higher Education Quality Council of Ontario, which reported that the literacy skills of a quarter of all university students and half of college students nearing graduation were below the absolute minimum necessary to succeed in their future careers (Weingarten & Hicks, 2018). Similarly, a report by Statistics Canada based on the Programme for the International Assessment of Adult Competencies (PIAAC) showed that 27% of Canadian university graduates only performed at level two or below in literacy on a five-level scale, i.e., below a level deemed minimal for successful participation in the modern-day labour market (Hango, 2014). Recently, the poor performance of Canadian graduates has even caught the attention of mainstream media, with a report highlighting that a majority of university students felt they lacked the necessary academic skills to succeed, particularly with respect to literacy and numeracy (Grayson, Côté, Chen, Kenedy & Roberts, 2019).

This state of affairs is not unique to Canada, nor is it a recent development. Barton and Lapointe (1995) summarized the findings of a large-scale literacy study carried out in the United States in 1992 by the Educational Testing Service (Kirsch, 1993), which surveyed adults over age 16 across 27,000 homes. In Barton and Lapointe's estimation, college graduates "...are certainly *more* literate, on average, than those who do not go to college, or do not graduate. But their *levels* of literateness range from a lot less than impressive to mediocre to near alarming, depending on who is making the judgment." (1995, p. 2). It is precisely this subjectivity of interpretation that has led to some confusion about exactly how discouraging the current picture really is, both in terms of levels of performance and how to quantify the expected amount of change in literacy over the course of one's degree (see Pascarella, Blaich, Martin, & Hanson, 2011).

These sobering statistics make understanding and quantifying the impact of each year of post-secondary education an important topic of study for the science of reading. If some students do not become sufficiently literate by the end of a post-secondary degree in the North American education system, what dynamics underlie this outcome? Specifically, the present paper asks what benefit is conferred by each year of study to the development of reading proficiency.

Research on this subject is rather sparse, and typically focuses on the difference between those with and without a college degree, rather than attempting to quantify the effect of each year of study on reading skill. In the book *"How college affects students",* Pascarella and Terenzini (2005) provide an exhaustive review of over 2,600 studies on the impacts of college education on students by looking at scores on a variety of tests of mathematics, science, social studies, liberal arts competencies, and most importantly for the purposes of this paper, English, i.e., reading comprehension and writing skill. In providing their analyses, the authors make an important distinction between changes which occur *during* college but which cannot definitively be stated to be the result of college intervention alone, and the *net effects* of college, which attempts to isolate the degree of change attributable to post-secondary education by adjusting for factors such as grades, race, age, and sex, among others. Additionally, the authors examine between- and within- college effects, exploring the differential effects of the college institution selected, as well as the conditional effects of college, that is, the effect that individual traits have on one's college experience. The authors assessed that on average, college students in their senior year demonstrate an improvement of .77 of a standard deviation in reading and writing skills, or 28 percentile points when compared to their first-year counterparts. In comparison, net effects of college for the same measure were estimated to be .59 of a standard deviation, or 22 percentile

points. This would suggest that the majority of the change which takes place during college is not simply the result of the natural maturation of a given individual, but rather may be ascribed to the post-secondary intervention itself. Another relevant finding comes from a longitudinal study by Bray, Pascarella and Pierson (2004) looking at how reading skill develops over college years. The researchers showed that comprehension improved relative to the degree of reading ability upon starting college—in other words, college students demonstrated Matthew Effects, in that those who were strong readers to begin with benefited the most from post-secondary intervention. Throughout their review, Pascarella & Terenzini (2005) show consistent evidence that the majority of this student academic improvement occurring over four-year college programs takes place between year one and year two, although the exact effect of each year of study is unclear.

The present cross-sectional study builds on this prior literature to evaluate changes in reading comprehension and reading component skills during a four-year undergraduate degree in a Canadian university with English as a medium of instruction, see below.

Although the meta-analysis of Pascarella & Terenzini (2005) is thorough, the influence of language background on reading development over post-secondary education is essentially left unexamined. L2 (or non-native) proficiency is an important factor predicting success in one's post-secondary career (Chapman, Wan & Xu, 1988; Wu, Garza & Guzman, 2015), and recent data estimate 14 percent of Canadian students are international, including 47% of students in the province of Ontario, where this research was conducted (Statistics Canada, 2020). Among all international students, 75% reported speaking a language other than English during childhood, and a large proportion of them report having language and

communication difficulties (Humphries, Knight-Grofe & Klabunde, 2009). Ability to process English-language text fluently and competently is of critical importance for L2 students in an English-medium academic environment, but this ability is often limited by lower overall proficiency in this L2, reduced reading experience in L2, as well as possible negative transfer effects from their L1 (Grabe, 2014; Koda, 2007). L2 reading tends to be slower and more effortful, with longer word fixations in eye-tracking studies (Godfroid, Winke, & Rebuschat, 2015; Whitford, Pivneva, & Titone, 2016), and L2 readers focus more of their attentional resources on bottom-up processing which may interfere with the automaticity necessary for comprehension (McLeod & McLaughlin, 1986). L2 students also perform more poorly on reading comprehension compared to L1 speakers of English, unless given extra time accommodations to account for their slower reading times (Miller, 2014). Consequently, many L2 students are ill-equipped to meet the reading requirements in a post-secondary environment (Roessingh & Douglas, 2012).

Some studies have investigated the role played by component skills of reading in higher vs. lower-skilled readers, which may offer a useful parallel to the differences between L1 and L2 readers. For instance, Grant, Wilson, and Gottardo (2007) examined component skills relating to reading comprehension, print exposure, and vocabulary in a cohort of post-secondary students diagnosed with reading disabilities (RD), and a non-RD control group. Results demonstrated that those in the RD group were not significantly different from the control group on any measures, possibly thanks to the compensatory strategies that enabled the RD group to qualify and be successful as post-secondary students in the first place.

3.2.1 The present study

We report a cross-sectional examination of reading performance in a large sample (N = 1,250) of students enrolled in a four-year undergraduate university program at a top-tier Canadian university. One goal of this study is to chart the development of reading comprehension as well as key component skills of reading (spelling, vocabulary size, exposure to print) in the course of university studies. Our focus is the direct and indirect effects that each year of study confers on changes in reading comprehension and related skills. Our second goal is to identify the role of language background (L1 vs L2 speakers of English) in reading skill development. We are interested both in (i) whether L2 speakers of English are at an initial (first-year) disadvantage in reading comprehension and component skills as compared to L1 speakers, (ii) whether the trajectories of skill development over the university years further enhance the L1 advantage or, conversely, enable L2 speakers to close the gap, and (iii) whether language background has a direct effect on reading comprehension or, conversely, most or all of its influence on reading comprehension is mediated by component skills of reading. Beyond tests of reading comprehension and component skills, our test battery included self-reported measures of perceived proficiency in English, demographic and language use information, and measures of motivation to perform well in the task. We study the influence of these additional factors as indices of perceived language development, relevant for attitudes and motivation toward language learning.

3.3 Methods

3.3.1 Participants

A total of 1,250 students (mean age= 20.26, sd= 2.96, 1,078 females) participated in the study between September 2017 and April 2020, recruited from a convenience participant pool at McMaster University in Hamilton, Ontario, Canada. All participants were enrolled in a four-year undergraduate degree program and were classified into groups based on their year of study: 1 to 4 and 5 (for those beyond the 4^{th} year of study). As part of the collection of demographic data, we asked participants to provide information about their language experience, including their first language, country of birth, and number of years spent in Canada. Based on the self-reported first language, participants were sub-divided into native English speakers (N=752) and non-native English-speakers (N=498). Table 3.3 reports the breakdown of participants by year of study and L1/L2 language background for the entire body of participants and for every test we use. The tables below provide additional demographic information about both cohorts included in this study: age of arrival to Canada and age at which English was learned (Table 3.1) and the breakdown of first languages for speakers of English as L2 (Table 3.2). Different numbers of participants contributed to individual tests: below we report sample sizes of native and non-native speakers of English for each test.

Students were given a partial course credit as a compensation for participation, and they were informed that participation was voluntary and that they could withdraw from the study at any time. The study received ethics clearance from the McMaster REB (2018-033).

Table 3.1: Demographic information about participants.

	Cohort	Median	Mean	SD	1st Qu.	3rd Qu.	Min	Max
Age of arrival in Canada	L1	0.00	.37	1.92	0.00	0.00	0.00	17.00
	L2	15.00	11.42	7.82	4.00	17.00	0.00	41.00
Age at which English was learned	L1	0.00	.02	0.31	0.00	0.00	0.00	6.00
	L2	6.00	.28	3.87	3.00	9.00	0.00	21.00

Table 3.2: Ten most common first languages for non-native speakers of English (total N= 498, 434 reported in table).

Order	Language	Frequency
1	Chinese	295
2	Urdu	32
3	Arabic	21
4	Korean	17
5	Russian	14
6	Punjabi	13
7	Spanish	12
8	Farsi	10
9	French	10
10	Polish	10

3.3.2 Materials

We administered a test of reading comprehension as well as tests of several component skills of reading: they are presented below.

Vocabulary Test: The vocabulary size test used for this study was adopted from Nation and Beglar (2007). The 14,000 word version of the original test of receptive vocabulary contains 140 multiple-choice items, with 10 items from each 1000 word family level. Each item contains a target word and a sentence in which this word is used (sometimes in a different morphological form). Four definitions for the target word follow, numbered a-d. The instruction is to "Circle the letter a-d with the closest meaning to the key word in the question".

Example:

SEE

They saw it.

a. cut

b. waited for

c. looked at

d. started

This original test was adapted in our study to enable more rapid assessment. The first ten items (representing 1000 most common words of English) were skipped, and all participants started from the 2000 word-family level. Also, a stop rule was established, such that after a subject had responded to ten items representing the same word-family level (i.e., 1000 words) they were only allowed to move to the next word-family level (next 1000) if they had committed less than

5 errors in those 10 items. The test was stopped if a subject had more than 5 errors in a given 1000 word-family level or if they had completed the test. The score was the number of correct responses with a possible range of 0 to 140 points. On this test's scale, a difference in 10 points roughly corresponds to a difference in 1000 words in vocabulary size.

Spelling Test: The spelling recognition test was adopted from Andrews and Hersch (2010; see also Andrews, Veldre, & Clarke, 2020). The original test consists of 88 English words, of which half are spelled correctly and half are not. The score is the number of correct responses, ranging from 0 to 88. The full list of words used in the spelling test is reported in Appendix A.

Author Recognition Test (ART): The Author Recognition Test, or ART, is an effective measure of an individual's print exposure (Acheson, Wells, & MacDonald, 2008; Stanovich & West, 1989). This test presents participants with a list of author names and distractor ("foil") names which could nevertheless plausibly be believed to be names of real-world authors. Participants are asked to identify only the names that belong to real authors by indicating them with a checkmark, and to ignore any names which are not believed to be those of published authors. The resulting score is calculated by subtracting the number of foils incorrectly selected from the number of correct responses.

Reading comprehension: The Gray Oral Reading Test (4th edition, GORT-4; Weiderholt & Bryant, 2001) is a well-established measure of reading comprehension, which tasks participants with reading short stories and answering

questions about the content therein. Stories are arranged in increasing order of lexical and syntactic complexity. We used stories 5-12 of the test in order to avoid a ceiling effect for responses. Each story contains 50-150 words and is followed by five comprehension questions. Accuracy of responses to these questions is our measure of reading comprehension skill. The possible score range was 0 – 40 points (8 stories x 5 questions).

Reading Habits Questionnaire: The reading habits questionnaire is a list of five questions to provide a subjective indication of the participant's own ability in reading and writing English, as well as the time spent each day on reading in different formats (print vs electronic). Specific questions are provided in greater detail in the Results section, and in full in Appendix B.

Motivation Questionnaire: Upon completion of the battery of reading skill tests, participants were asked ten questions related to their motivation on the test, e.g. how important they felt it was to do well, how motivated they felt to do well, and so on. For this self-reported measure of subject motivation, the 10-item Student Opinion Survey was used (Thelk, Sundre, Horst, & Finney, 2009; see also Finney, Mathers, & Myers, 2016). This questionnaire is reported in Appendix C.

3.3.3 Procedure

Participants began with providing informed consent and then responded to a demographic questionnaire, including information about their age, education, first language, as well as their subjective estimate of reading, writing, listening, and

speaking proficiency in English. Following this, participants were presented with a battery of tests of reading proficiency, which are detailed in the preceding Materials section. At the end, participants were asked about their motivation level to complete the battery of tests. The entire experimental session did not exceed one hour.

Variables and statistical considerations

Scores in the component skills of reading included in the test battery were each considered as dependent variables. We fitted a linear multiple regression model to each dependent variable and tested the effects of language background (native vs non-native), year of study and their interaction. Then we fitted a linear multiple regression model to reading comprehension scores with individual scores of participants obtained in all component-skill tests, as well as year of study and interactions between year of study and component skill scores.

After fitting each regression model, we removed outliers, i.e., data points with standardized residuals exceeding 2.5 SD from the predicted values. We report regression models fitted to these trimmed datasets. Analyses below used linear multiple regression models implemented as function lm in the statistical platform R v. 3.6.1 (R Core Team, 2019); for plotting, the effects library was used (Fox & Weisberg, 2019). Each model was tested for overfitting by using validate function with 200 bootstrapping iterations (library rms in R): estimates of optimism were small in every model. We report coefficient tables for each model and additionally provide ANOVA estimates of significance for critical categorical variables and interactions.

Where applicable, we also report differences between levels of categorical variables (e.g., L1 vs L2 speakers or years of study) in percentile points for interpretability. Each difference was calculated by finding the percentile in Group A that corresponded to the median of Group B and computing the difference between that percentile and 50%: thus, if the median score of L1 speakers corresponded to the 70th percentile of scores among L2 speakers, the difference was 20 percentile points (70% – 50%). We compare performance of the 1st and the 4th years of study. While we also consider students beyond the 4th year, these subsamples tend to be smaller and possibly partly composed of academically less efficient individuals who required additional time to complete a typical 4-year undergraduate degree.

3.4 Results

We begin with reporting analyses of individual tests of component skills of reading, followed up by an analysis of reading comprehension.

Vocabulary: A total of 254 participants (127 native and 127 non-native speakers of English) contributed to this test.

Table 3.3: Number of participants for component skill tests by year of study and language background.

Test	Year 1		Year 2		Year 3		Year 4		Year 5		Totals		
	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2	Total
TOTAL	130	190	217	117	184	101	164	61	57	29	752	498	1250
Vocabulary	18	44	47	34	31	24	26	18	5	7	127	127	254

Spelling	65	90	144	68	116	50	94	36	23	14	442	258	700
ART	80	92	134	53	118	47	106	32	35	13	473	237	710
Reading Habits	51	101	88	61	68	52	65	32	24	16	296	262	558
Motivation	52	102	85	62	66	55	66	32	24	16	293	267	560
GORT	49	92	84	56	65	50	61	28	21	13	280	239	519
GORT Joint	16	38	47	32	31	23	23	14	4	7	121	114	235

Figure 3.1 visualizes results of this analysis and Table 3.4 reports the regression model. An analysis of variance (ANOVA) based on the linear model showed a significant effect for year of study ($F(4,240)= 8.647$, $p < 0.001$) and for language background ($F(1,240)= 65.492$, $p < 0.001$). This demonstrates, perhaps unsurprisingly, that native English speakers outperform non-native L2 speakers, and that students tend to improve with each year of study. What is more interesting, however, is that there was a marginally significant interaction between year of study and native language ($F(4, 240) = 2.24$, $p = 0.066$). The improvement was somewhat greater in non-native English speakers, such that in years 4 and 5 they closed the performance gap with the native speakers that was in place in years 1-3.

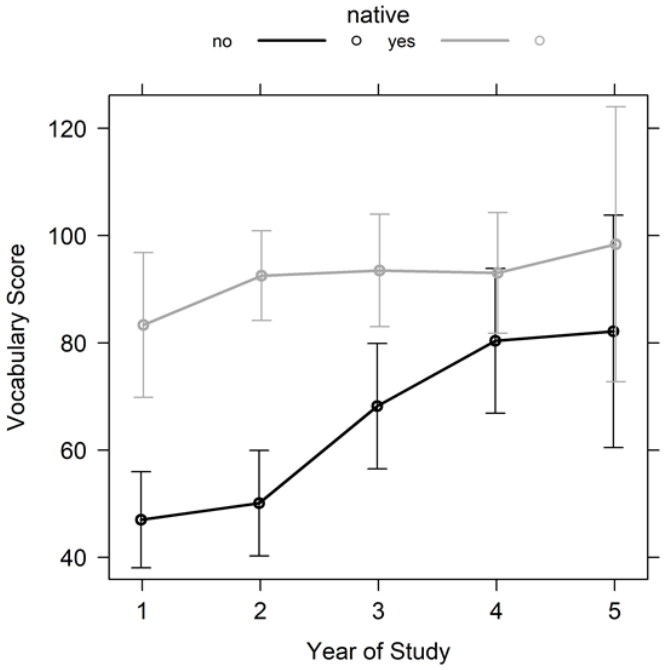

Figure 3.1: Partial effects of year of study on vocabulary score for L1 and L2 speakers of English. Error bars represent the standard error for each data point.

Table 3.4: Regression model for Vocabulary Test. N = 254 before trimming, 250 after trimming, Adjusted R-squared: 0.287. Reference levels: Year 1 and L2.

	Estimate	SE	t-value	p-value
Intercept	47.000	4.545	10.341	<0.001
Year 2	3.088	6.751	0.457	0.648
Year 3	21.208	7.480	2.835	0.005
Year 4	33.389	8.229	4.058	<0.001
Year 5	35.143	11.902	2.953	0.003
L1	36.333	8.229	4.415	<0.001
Year 2 : L1	6.110	10.519	0.581	0.562
Year 3 : L1	-11.042	11.456	-0.964	0.336
Year 4 : L1	-23.684	12.139	-1.951	0.052
Year 5: L1	-20.076	18.924	-1.061	0.290

To contextualize these findings, L1 English students showed an advantage of 25 percentile points over L2 peers. L2 English speakers improved a total of 27 percentile points from Year 1 to Year 4, and L1 students improved a total of 22 percentile points over the same period.

Spelling: A total of 700 participants (442 native and 258 non-native English speakers) contributed to this test: see Table 3.3 above for the breakdown of participants by year of study and language background. Table 3.5 reports the regression model and Figure 3.2 visualizes its critical results. ANOVA based on the linear model showed a significant effect for year of study ($F(4,681) = 7.195$, $p < 0.001$) and for language background ($F(1,681) = 49.127$, $p < 0.001$). Again, native English speakers outperformed non-native L2 speakers, and all students tended to improve on spelling with each year of study. Similar to vocabulary scores reported above, there was a reliable interaction between year of study and language background on spelling skill ($F(4, 681) = 3.00$, $p = 0.018$). In line with the vocabulary test, non-native speakers showed more improvement in their spelling skills than native speakers. While an advantage to native speakers existed in years 1-3 of undergraduate education, this gap was closed by year 4.

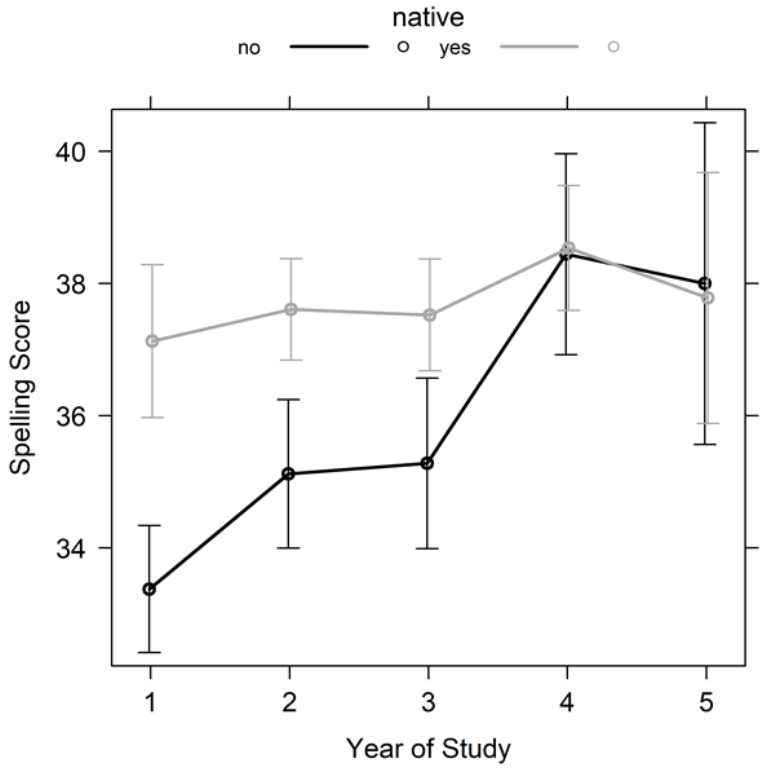

Figure 3.2: Partial effects of year of study on the spelling test score for L1 and L2 speakers of English. Error bars represent the standard error for each data point.

Table 3.5: Regression model for Spelling test. N = 700 before trimming, 694 after trimming. Adjusted R-squared: 0.105. Reference levels: Year 1 and L2.

	Estimate	SE	t-value	p-value
Intercept	33.378	0.489	68.230	<0.001
L1	3.751	0.766	4.897	<0.001
Year 2	1.743	0.752	2.318	0.021
Year 3	1.902	0.819	2.324	0.020
Year 4	5.067	0.915	5.536	<0.001
Year 5	4.622	1.333	3.467	0.001
L1 : Year 2	-1.263	1.032	-1.223	0.222
L1 : Year 3	-1.505	1.097	-1.372	0.170
L1 : Year 4	-3.658	1.190	-3.073	0.002
L1 : Year 5	-3.969	1.750	-2.268	0.024

L1 English students showed an overall advantage of 16 percentile points over L2 peers. L2 English speakers improved very strongly, gaining a total of 41 percentile points from Year 1 to Year 4, and L1 students improved a total of 7 percentile points over the same period.

Author Recognition Test (Print Exposure): A total of 710 participants (473 native and 237 non-native English speakers) contributed to the Author Recognition Test dataset, see Table 3.3 above for the breakdown of participants by year of study and language background.

Table 3.6 reports the regression model and Figure 3.3 visualizes critical findings. ANOVA based on the linear model showed a significant effect for both year of study ($F(4,685) = 3.689$, $p = 0.006$) and for language background ($F(1,685) = 74.243$, $p < 0.001$). The interaction between year of study and language background did not approach significance ($F(4,685) = 0.994$, $p = 0.410$). The null interactive effect suggests that the improvement linked to the greater exposure to print was equally strong in both L1 and L2 English speakers.

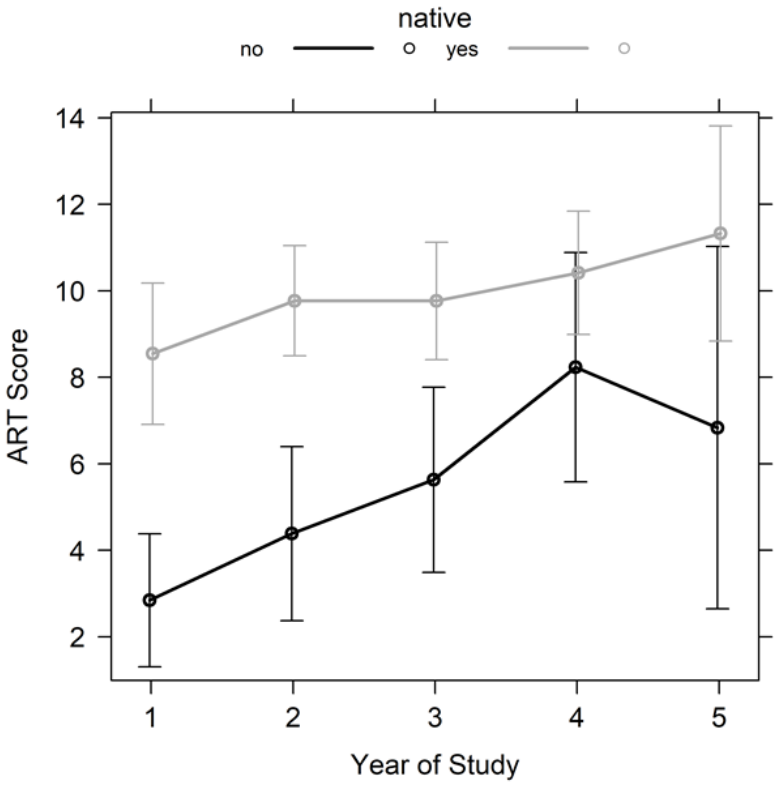

Figure 3.3: Partial effects of year of study on ART score by language background. Error bars represent the standard error for each data point.

Again, English L1 speakers outperformed L2 speakers on ART, and the print exposure score improved with each year of study. L1 English students showed an advantage of 26 percentile points over L2 peers. On average, participants improved a total of 20 percentile points from Year 1 to Year 4.

Table 3.6: Regression model for ART. N = 710 before trimming, 693 after trimming. Adjusted R-squared: 0.111.

	Estimate	SE	t-value	p-value
Intercept	2.843	0.784	3.628	<0.001
L1	5.702	1.143	4.990	<0.001
Year 2	1.542	1.290	1.195	0.233

Year 3	2.788	1.342	2.077	0.038
Year 4	5.391	1.561	3.454	0.001
Year 5	3.991	2.273	1.755	0.080
L1 : Year 2	-0.317	1.666	-0.190	0.849
L1 : Year 3	-1.569	1.724	-0.910	0.363
L1: Year 4	-3.521	1.911	-1.842	0.066
L1 : Year 5	-1.211	2.733	-0.443	0.658

Reading Habits: A total of 558 participants (296 L1 and 262 L2 speakers) contributed to this dataset: see Table 3.3 for the breakdown of participants by year of study and language background. Question 1 of the reading habits questionnaire asked participants about the quality of their writing in English, question 2 about their reading speed in English, question 3 about their reading comprehension, and question 4 about how many hours a day they spend reading and writing. Questions 1 through 4 demonstrated highly comparable results. Non-native English-speaking participants demonstrated a significant improvement on all these subjective reports over the course of undergraduate studies, while English native speakers effectively showed a flatline across five years of study. Regression models are reported in Appendix D.

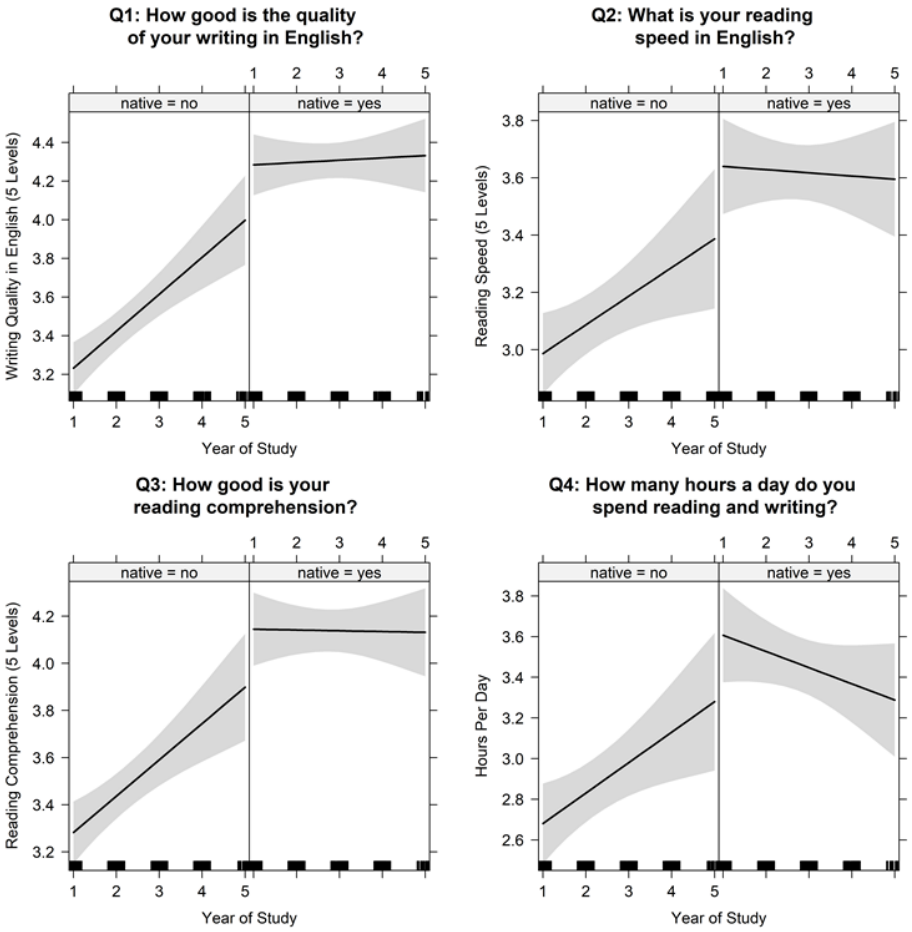

Figure 3.4: Partial effects of year of study and language background on self-report responses to reading habits questions. Error bars represent the standard error for each data point.

Neither year of study nor language background were significant predictors of responses to Question 5, which asked "What percentage of this time do you spend reading or writing texts on social media websites (0-100)?"

Motivation: A total of 560 participants (293 L1 and 267 L2 speakers of English) contributed to this task, see Table 3 for a breakdown of participants by year of study and language background. Table 7 reports the regression model and

Figure 3.5 visualizes critical effects. ANOVA based on the linear model showed a significant effect for language background (F(1,492)= 14.783, $p < 0.001$). L1 speakers of English reported a higher motivation for the tasks administered in the test battery compared to L2 speakers. Year of study was not a significant predictor of motivation (F(4,492)= 1.221, p= 0.301). Similar to the model for the Author Recognition Test, the best-fitting model did not include an interaction effect, as this was not significant (p > 0.1).

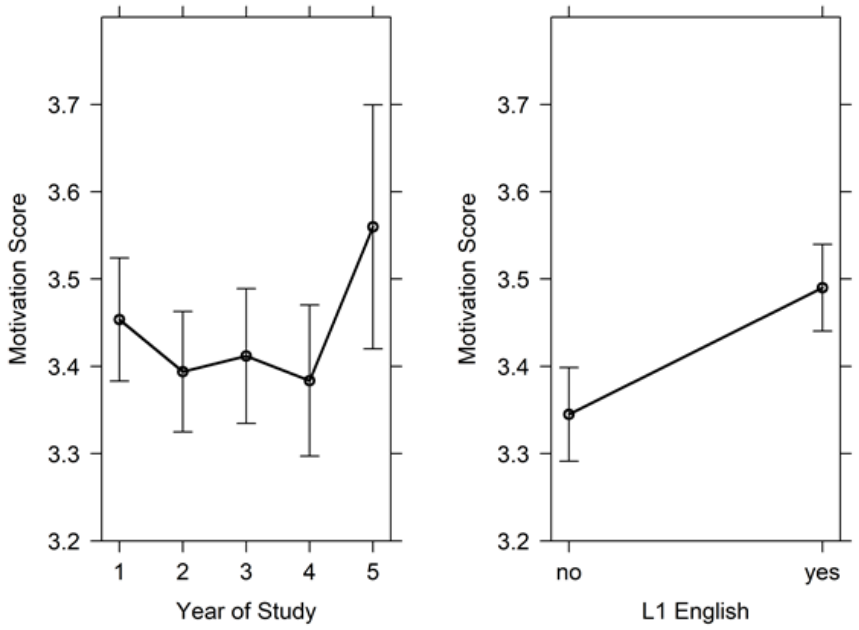

Figure 3.5: Partial effects of year of study (left) and language background (right) on scores in the motivation test.

Table 3.7: Regression model for Motivation. N = 560 before trimming, 500 after trimming, Adjusted R-squared: 0.029. Reference levels: Year 1 and L2.

	Estimate	SE	t-value	p-value
(Intercept)	3.376	0.038	89.356	< 0.001
Year 2	-0.060	0.051	-1.184	0.237
Year 3	-0.042	0.053	-0.783	0.434
Year 4	-0.070	0.057	-1.220	0.223

Year 5	0.106	0.080	1.328	0.185
L1	0.145	0.038	3.845	< 0.001

L1 English students showed an advantage of 19 percentile points over L2 peers. L2 English speakers improved a total of 12 percentile points from Year 1 to Year 4, whereas L1 students did not show any change over the same period.

Reading comprehension: The Gray Oral Reading Test (GORT) is a measure of reading comprehension. The analysis of GORT scores brings together all component skills of reading, relevant demographic information, and a test of reading comprehension. First, we analyze the effects of cohort (L1 vs L2) and year of study on reading comprehension scores in all available participants. As the next step, we analyze contributions of these predictors as well as specific component reading skills on reading comprehension scores for participants that have completed all relevant tests.

A linear multiple regression model was fitted to GORT scores (N = 519, with 280 L1 and 239 L2 speakers of English) with language background, year of study and their interaction as predictors. None of these effects were statistically significant (all $p > 0.15$; model not shown). Somewhat surprisingly, reading comprehension did not reveal a direct advantage to either L1 over L2 speakers nor did it show improvement over the course of undergraduate education.

We refined this analysis by adding scores on the other component skills of reading to the model. This restricted our sample size to N = 235 participants (121 L1 and 114 L2), see Table 3.3 above for a further breakdown of samples by year of study. Our best-fitting linear regression model (Table 8) was one which examined

the individual factors of year of study, language background, spelling, ART score, motivation, and vocabulary score. Figure 3.6 summarizes the findings.

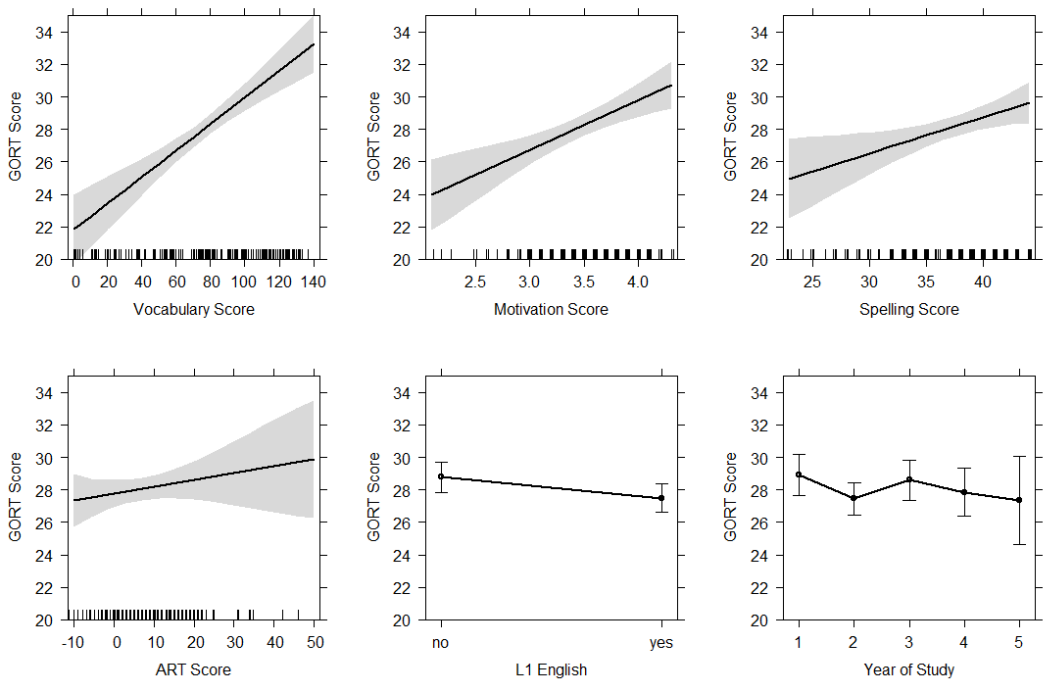

Figure 3.6: Partial effects of component skills of reading, year of study and language background on reading comprehension scores.

An individual's comprehension score was reliably predicted by their vocabulary score ($F(1,221) = 32.64$, $p < 0.001$), print exposure as measured by ART ($F(1,221) = 20.31$, $p < 0.001$), motivation ($F(1,221) = 27.07$, $p < 0.001$), and spelling ($F(1,221) = 117.10$, $p < 0.001$). As Figure 3.6 suggests (and the model coefficients confirm), effects of individual predictors on reading comprehension vary greatly in magnitude: we discuss them in decreasing order. Vocabulary size was the strongest predictor, positively correlated with comprehension accuracy. An increase in one unit of SD for vocabulary size predicted an increase in 2.85 points of comprehension accuracy, or 7% of the test scale. The second strongest effect was motivation:

individuals with a higher motivation to complete tasks show much stronger performance in reading comprehension. This effect is noteworthy because it is not often taken into account in studies of reading performance. The present data reveal, however, that it is sizable (an increase in 1 SD of motivation predicts an increase in 1.25 points of comprehension or 3% of the test scale). Spelling score correlated positively with comprehension accuracy, suggesting that the quality of orthographic representations affects higher-level cognitive tasks of understanding and integrating information in print even in very advanced readers. An increase in 1 SD of spelling translated into an increase of 1 point of the GORT score or 2.5% of the test scale. Exposure to print had a relatively weak positive effect on comprehension: an increase in 1 SD of the ART score came with an increase in 0.39 points in reading comprehension accuracy or 1% of the test scale. Language background and year of study did not have reliable effects on comprehension, nor did an interaction between language background and any other predictor (all $p > 0.1$).

Table 3.8: Regression model for GORT reading comprehension scores. N =235 before trimming, 231 after trimming, Adjusted R-squared: 0.461.

	Estimate	SE	t-value	p-value
Intercept	4.043	3.615	1.119	0.265
Spelling	0.223	0.085	2.628	0.009
ART	0.042	0.043	0.982	0.327
Motivation	3.072	0.794	3.871	<0.001
Vocabulary	0.082	0.013	6.090	<0.001
Year 2	-1.459	0.835	-1.747	0.082
Year 3	-0.301	0.901	-0.334	0.739
Year 4	-1.047	1.001	-1.046	0.296
Year 5	-1.560	1.521	-1.025	0.306
L1	-1.285	0.687	-1.872	0.063

3.5 Discussion

In evaluating the component skills of reading, a consistent picture begins to emerge. For some skills (exposure to print), L1 and L2 speakers of English improve their proficiency at approximately the same rate of growth throughout their post-secondary education. For others like vocabulary size and spelling, an increase is found in all students, but L2 speakers of English benefit much more from university education. Although L2 speakers may start from a disadvantaged position in these skills, they close the gap with L1 speakers towards the end of the undergraduate degree. The fact that this should be so is not a trivial one, as it is not well-established that the "intervention" of post-secondary education is equally or differentially effective across cohorts with different language backgrounds. In an ideal world, the educational environment should strive to provide a level playing field for all students. But with university-level reading material requiring a relatively high degree of proficiency to begin with, it is encouraging to witness that L2 speakers benefit in the development of reading component skills to approximately the same or even greater degree as L1. These findings are brought into sharp relief by examining percentile point differences between cohorts. Aggregating across all measures analyzed in this paper for illustrative purposes, L1 English students show a large average benefit of 24 percentile points compared with L2 peers across all years of study. Yet from Year 1 to Year 4, L1 students improve by an average of just 12.60 percentile points, whereas L2 peers improve by an average of 29.20 percentile points over the same period. These estimates are comparable with the effect of college education reported by Pascarella and Terenzini (2005), see above.

Importantly, the observed discrepancy between achievements of L1 and L2 speakers in component skills of reading runs counter to the longitudinal findings of Bray, Pascarella, and Pierson (2004). They discovered that an individual student's reading comprehension improved over the course of their post-secondary studies as a function of their initial reading ability as a first-year student. In other words, students showed a kind of Matthew Effect—the "rich" readers became "richer" and the "poor" readers became "poorer". In our data, the cohort starting at a disadvantage becomes "richer" and the advantaged one progresses less. This conclusion finds further support in our analysis of reading habits (Figure 3.4). It indicates that L2 speakers of English are aware of their progress, as their subjective estimates of different language faculties increase with year of study. The increase in self-reported proficiency is negligible in L1 speakers of English who indeed improve less. We believe that the discrepancy stems from differences between our cross-sectional data versus Bray et al.'s longitudinal results. Logically, systematic changes within-person that can only be assessed longitudinally do not necessarily coincide with the behavior of the group, where advances in some individuals may be watered down by the lack of change in others. It is thus possible that those L2 students who had a greater initial English proficiency showed more gains over the course of university studies than those who were initially less proficient. Yet as a group, on average, L2 speakers gained more in reading skills than L1 speakers did over the same period of time, showing an "anti-Matthew" effect. We believe this finding to be the central contribution of the present paper to existing accounts of reading development in higher education. Logically, this result could have originated from a scenario where L1 speakers are at the ceiling of reading proficiency on the respective test, and thus cannot show as much improvement on the test scale. We find this highly unlikely given that the L1 performance in all tests is far

from the ceiling, as well as repeated reports of prevalent functional illiteracy among university students, also in Canada (see the Introduction).

Another intriguing set of results reveals that (a) neither year of study nor the native language have a direct effect on reading comprehension but (b) reading comprehension is strongly predicted ($R^2 = 0.46$) by component skills of reading, which are co-determined by both language background and year of study. Taken together, these results suggest that going through the years of undergraduate education, both L1 and L2 English-speaking students improve their reading comprehension by virtue of improving their component skills of reading—e.g., vocabulary size, spelling, and exposure to print—which in turn engender progress in reading comprehension. To rephrase, we find evidence that the individual growth in component skills of reading over the years of study does improve one's reading comprehension. Yet, apparently, this tendency at the level of individual participants is not strong enough to show up when cohorts from different years of study are directly compared in reading comprehension to one another. Statistically speaking, component skills of reading absorb or fully mediate the effects of either year of study or language background on reading comprehension.

Another implication of this set of findings is that once these component skills are accounted for, the categorical distinction in comprehension scores between native and non-native English-speaking students is effectively erased. Thus, an individual demonstrating good mastery of the component skills of reading is able to overcome a group difference and possible academic disadvantages associated with being a non-native speaker of English. This set of findings also speaks to the greater effectiveness of university intervention for non-native speakers of English. An applied aspect of these findings is that explicit and direct instruction in the

component skills of reading—primarily, vocabulary size and, to a smaller degree, word decoding—may lead to important gains in reading comprehension over and above the natural gains that the daily routine of a student confers.

3.6 Limitations and Future Directions

This is a cross-sectional rather than a longitudinal study. Thus, each subsequent year after Year 1 only captures individuals who have successfully progressed to that year, and this success may be partly dependent on the status of their reading skills. As a result, part of the improvement that we observe over the undergraduate years in several component skills of reading may be due to selection bias due to academic achievement rather than a true improvement: a longitudinal study is necessary to disentangle these confounds. Also, the cross-sectional design does not enable us to separate changes that happen during university from those taking place due to university education (see discussion in Pascarella & Terenzini, 2015): above we consider them jointly.

Moreover, the nature of our study zooms in on the relative performance of student groups defined by their year of study and language background. We do not provide a comparison of student performance with normative data that would identify their position on the scale of functional literacy. Given independent reports (see the Introduction) it is likely that the overall distribution of functional literacy in our participant cohort reflects the relatively low levels observed among Canadian university students. Our data do not speak in favor or contra this notion. Instead, our data give an insight into different dynamics of reading development in different sizable groups of students.

An important venue for future research will be to determine if the present results are generalizable across other universities and colleges, as well as other or additional tests of reading comprehension and component skills of reading. Furthermore, it may be worthwhile to investigate if the method and environment of post-secondary instruction are important determiners of reading improvement. As many universities begin to offer more avenues for digital learning, it is not certain how this comparatively isolated form of education compares to the traditional university model, where frequent social interactions and a community environment form an essential part of post-secondary life. In either environment, of course, the majority of material which a student will read over the course of their academic career is done in solitude, ideally with enough time to absorb and reflect upon the information presented. However, the campus model offers opportunities to interact with others who are also engaging with the same or similar material, and this reciprocity may contribute to the reinforcement of knowledge in an important way.

In sum, the present findings offer a new insight into the progress in reading comprehension and related skills of undergraduate university students in the course of their studies, and the role of individual component skills of reading that engender this progress. Another point of novelty of the study is in showing how L2 speakers of English can emerge from post-secondary education with reading skills to rival even students who have been speaking English from a very young age.

Chapter 4

General Discussion

4.1 Findings and Implications

In the first study presented in this thesis, Item Response Theory (IRT) analysis of scores on the Author Recognition Test (ART) showed that individual author names on the test did not provide discriminating evidence between higher- and lower-skilled readers among L2 English speakers, college students, and students in a university ESL pre-admission program. Indeed, for the latter group, virtually all author names were approximately equal in difficulty. In other words, ART is not informative for determining print exposure in these populations. This in turn makes it difficult to compare scores across populations with different linguistic and educational backgrounds. The reasons for these findings are still unclear—it may be that these populations are not reading enough, or that they are reading other authors not included in ART due to its primary focus on Western fiction writers. It is equally uncertain what proportion of reading is done in each of a multilingual's languages. Consequently, ART is ill-equipped to understand the volume and kinds of reading done by these students.

In the second study, findings showed that each year of post-secondary study directly benefited component reading skills, and indirectly benefited reading comprehension, which was mediated by the improvement in component skills. Additionally, although L2 university students start from a significant disadvantage, they also tend to improve significantly more than L1 peers through each year of their undergraduate degree on most measures of reading ability, and at least as well in another measure (print exposure). What this demonstrates is that with sufficient ability in component skills of reading, the clear distinction between L1 and L2 speakers essentially vanishes, as L2 speakers are able to perform with native-like proficiency.

4.2 Author Naming Test

The difficulties associated with comparing scores for the Author Recognition Test (ART) between different populations has led me directly to the proposal of a new Author Naming Test (ANT), which will be investigated further in later studies as an alternative index of print exposure. The design of ANT is relatively straightforward. Participants are given a set amount of time to spontaneously name as many authors as they can, which are recorded and tallied by the researchers. This is broadly in line with other similar psychological tests of verbal fluency such as the semantic category fluency or the initial- letter fluency tests (e.g. Hurks et al., 2006). Essentially, the reasoning is that the more readily accessible some information is to memory, the greater is the individual's familiarity with the subject.

Another potential utility of ANT is that multilingual participants can be asked in which of their languages (L1, L2, etc.) they read the majority of each author's works. From this, researchers may evaluate what proportion of each language spoken is represented by each of the author names provided, and it will be possible to determine if this proportion correlates with the literacy skill in the language tested. This may be likely, given that it is well-established that the quantity of free voluntary reading (FVR) done in one's L2 has been shown to predict L2 test performance (Constantino, Lee, Cho, & Krashen, 1997; Gradman & Hanania, 1991; Krashen, 2004). Granted, it is not a certainty that those who read more in their second or third language will necessarily provide more author names from these languages, but it is an interesting additional empirical question which is worth considering.

An initial pilot study of ANT has suggested it may have some utility for measuring print exposure. This experiment was carried out online using Amazon Mechanical Turk, in which participants were asked to complete a battery of tests. This included a brief survey about reading habits (including the number of books kept at home), ANT, and as a last test (to avoid priming for ANT), ART. In this simplified version of ANT, participants were asked to only name fiction authors, and were awarded one point for each entry provided. Results showed that while ART and ANT do not correlate, ANT *does* correlate with the number of books at home—a well-established predictor of literacy (Elley, 1992; Sikora, Evans, & Kelley, 2019). It thus stands to reason that further inquiry may demonstrate that reading skill does correlate with ANT. To determine if ANT is viable as a proxy measure of print exposure with validity equal to that of ART for L1 English-speaking university students, it will be necessary to prepare a more elaborate study, where ANT scores are collected in addition to measures of component skills of reading to

determine if these tests are correlated. If successful with this population, ANT would then be tested with other groups such as L2 speakers of English and college students, to see if the same kind of relationship with reading skill is observed.

Additionally, ANT may provide independent verification of how many books are in the home, given that there is typically no simple way of substantiating survey responses to this question at a large scale. Given the increasing prevalence of eBooks and tablets, of course, the number of books on one's physical shelves may soon no longer accurately predict an individual's exposure to "print"—and an Author Naming Test, which takes into account virtually all of the written media consumed by an individual, may be a more robust indicator.

One important consideration in the design and implementation of an Author Naming Test is that the ability to spontaneously produce names of authors is one which is largely contingent on memory capacity and recall. For this reason, a baseline of ability in these domains will need to be controlled for in the statistical analysis of the relationship between ANT and component skills of reading. Due to this fundamental difference between ART and ANT, it is perhaps unsurprising that these two tests have not been shown to correlate thus far. After all, ART relies on a kind of "signal detection logic", in which the participant tries to pick out only those names which trigger a recognition response, separate from the "noise" of the distractor names (Stanovich & West, 1989). This is a far more passive task in comparison to a "naming fluency" skill, which requires the retrieval of a set of stored memory items. Because of this, it is also possible that an *Author* Naming Test may be too narrow a constraint for many of today's readers, who may be more familiar with book titles which they have read. Increasingly we consume books on tablets and e-Readers, which allow the reader to immediately return to where they left off,

bypassing entirely the physical process of picking up a book off a shelf, examining its cover, and imperceptibly absorbing the name of the author. The importance of a measure such as the number of books in one's home as a predictor of reading skill suggests that this kind of ambient awareness of, and tactile interaction with books themselves as physical objects may be an important element of a culture of readership, and something which may be lost in the transition to digital reading. Taking this into account, perhaps a combination of whichever items the participant can summon from memory—author or title—will provide a better understanding of the quantity and kind of reading undertaken by an individual.

4.3 Additional Future Directions

A common refrain in psychological research which nevertheless bears repeating: more longitudinal research is needed to better understand these findings, as well as more replication. As mentioned in the previous chapter, because the research on the effects of year of study on reading skill was cross-sectional rather than longitudinal, it is difficult to say with certainty the extent to which these findings would be generalizable to individuals. There is unquestionably the possibility of attrition among university students who are unable to continue on to later years of university study, a kind of undergraduate "literacy selection" (Reder, 1998).

Additionally, the displacement effect, commonly examined with respect to television and video game consumption among children which usurps time spent reading, is surprisingly not well-studied among adults and particularly in college students. One study did attempt to examine this question by inviting college students to complete a time-diary survey, and students mostly reported that

Internet and technology use did not displace leisure reading (Mokhtari, Reichard, & Gardner, 2009). However, like any subjective self-report measure, these results should be interpreted with caution as they are prone to social desirability bias (for example, mean self-report scores for reading ability were 7.9/10, doubtfully accurate given the previous discussion about undergraduate reading skill). Although there is little doubt about the substantial use of technology and the Internet among undergraduate students, nevertheless it would be important to understand the extent to which they are displacing leisure time which might otherwise be spent reading with time engaged in these technological pursuits. In Chapter 3, the research conducted on the effects of year of study included another self-reported reading habits questionnaire which asked students about how much time they spent reading and writing each day, as well as a separate question asking how much of this time was spent on social media. A linear model revealed a significant relationship between these two questions, such that as the hours of time spent reading increased, time spent on social media decreased. However, this was only found to be significant for native speakers of English. Non-native speakers, curiously, did not show any kind of displacement relationship between social media and overall reading time. Because the displacement effect was not the focus of this study, this finding was not reported in this chapter. Additionally, since participants were only surveyed about the percentage of overall reading time spent on social media, it was unclear how much time is spent on other forms of social media, for example video sharing platforms. A future study might try to better understand the potential displacement effects of all social media use on reading without relying solely on self-report measures.

Another question which will be worth confronting will be the effects of the current digital shift in education. Like the digital shift in reading, this change may

lead to unforeseen and in some cases, potentially adverse consequences. At the time of writing, the world is suffering under a global pandemic of the kind not seen in a hundred years. As a result, universities around the world are shifting their resources to online learning to help slow the spread of disease. As learning moves off-campus, and videoconferencing replaces lecture halls, the character of the university as an institution may fundamentally change. At present, it is not possible to state with any certainty what changes this will engender in the ways in which students read and consider academic material. From their inception, universities have generally sought to create a physical space for students to congregate and share ideas, and this foundational element will be profoundly altered. This change should not go unremarked.

Lastly, a note about the value of fiction reading. It is important to reiterate that ART specifically tries to capture the volume of "free voluntary reading", or reading for pleasure—which is more often fiction (Kim & Krashen, 1998b; Lee, Krashen, & Tse, 1997; West, Stanovich, & Mitchell, 1993). Far from being a waste of time, reading fiction has been shown to contribute to verbal ability, social skills, and interpersonal sensitivity (Fong, Mullin, & Mar, 2013; Mar & Rain, 2015; Mumper & Gerrig, 2017). Fiction books are also especially useful for reading development as the wide range of styles and registers, even within a single book (e.g. dialogue between characters, descriptions of locales or physical features, narration of events, etc.) serves as a link between academic and conversational language (Krashen, 1995; Paulson, 2006).

Furthermore, unlike biographies, which may exclude certain details about a public figure's life—either due to not conforming to the "non-fiction narrative" being developed, or not being publicly disclosed until after publication—fiction represents

a "complete world" where all of the relevant information is contained. In other words, a comprehensive record of all of the protagonist's relevant thoughts and actions are contained within the book in which they live. Accordingly, fiction stories provide a concise and elucidating worldview from the perspective of the author, which can fundamentally alter our own. To illustrate this point, Strange (2002) discusses the reasons why a book such as *Uncle Tom's Cabin* (Stowe, 2017, originally published in 1852) was able to change the public discourse on slavery in the United States so quickly and persuasively, when the knowledge alone of its injustice ought to have been sufficient. Strange cites Charles Dudley Warner (1896), who wrote to the effect that prior to the novel's release, the American public had generally become inured to the grim realities about slavery through the abundant spread of abolitionist information. As Strange explains, in Warner's estimation, the difference was that *Uncle Tom's Cabin* was able to portray the immorality and cruelty of oppression in a stark and personal way. For this reason, Warner credits the book for contributing to an awakening of "the public conscience", leading ultimately to abolition. This example speaks to the profound power of fiction to promote empathy and perspective-taking, and to encounter lives far removed from our own.

In the previous chapter, findings were presented which showed how improving component skills of reading can enhance reading comprehension. One implication from these findings was that an explicit focus on developing these skills during university may lead to even further improvements. Another way these skills might be improved is through fiction reading, in particular among university faculties which might generally have less exposure to fiction as part of their program requirements. As important as it is for graduates and citizens alike to know

objective facts, it must also be made clear to students that the pursuit of facts alone is insufficient for a well-rounded academic career.

Bibliography

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*(1), 278-289.

An, X., & Yung, Y. F. (2014). Item response theory: What it is and how you can use the IRT procedure to apply it. *SAS Institute Inc. SAS364-2014*, *10*(4).

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299.

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior research methods*.

Austin, W., & Totaro, M. W. (2011). High School Students' Academic Performance and Internet Usage. *Journal of Economics and Economic Education Research*, *12*(1), 41.

Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press, USA.

Barton, P. E., & Lapointe, A. (1995). *Learning by Degrees: Indicators of Performance in Higher Education.* Princeton, NJ: Educational Testing Service.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental psychology*, *34*(6), 1373.

Benevides, T., & Peterson, S. S. (2010). Literacy attitudes, habits and achievements of future teachers. *Journal of Education for Teaching*, *36*(3), 291-302.

Bobkina, J., & Stefanova, S. (2016). Literature and critical literacy pedagogy in the EFL classroom: Towards a model of teaching critical thinking skills. *Studies in Second Language Learning and Teaching*, *6*(4).

Bourdieu, P., & Richardson, J. G. (1986). Handbook of Theory and Research for the Sociology of Education. *The forms of capital*, 241-258.

Bray, G. B., Pascarella, E. T., & Pierson, C. T. (2004). *Postsecondary education and some dimensions of literacy development: An exploration of longitudinal evidence.* Reading Research Quarterly, 39(3), 306-330.

Budd, B. (2019). The populist radical right goes Canadian: An analysis of Kellie Leitch's failed 2016–2017 Conservative Party of Canada leadership campaign. In *Populism and World Politics* (pp. 137-163). Palgrave Macmillan, Cham.

Budd, B. (2020). The people's champ: Doug Ford and neoliberal right-wing populism in the 2018 Ontario provincial election. *Politics and Governance*, *8*(1), 171-181.

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of educational research*, *65*(1), 1-21.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of learning disabilities*, *44*(5), 431-443.

Canadian Council on Learning. (2007). *State of learning in Canada: no time for complacency.* Retrieved June 20, 2020 from https://www.dartmouthlearning.net/wp-content/uploads/2013/02/State-of-Learning-in-Canada-No-Time-for-Compacency-2007.pdf

Canadian Council on Learning. (2009). *Post-secondary Education in Canada: Meeting Our Needs?* Ottawa: Canadian Council on Learning. Retrieved June 20, 2020 from https://files.eric.ed.gov/fulltext/ED525254.pdf

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & cognition*, *28*(1), 143-153.

Chapman, D. W., Wan, T., & Xu, M. (1988). The academic adjustment of international students in American universities. *Education and Society*, *6*, 96-103.

Chen, S. Y., & Fang, S. P. (2015). Developing a Chinese version of an Author Recognition Test for college students in Taiwan. *Journal of Research in Reading*, *38*(4), 344-360.

Choi, W., Lowder, M. W., Ferreira, F., & Henderson, J. M. (2015). Individual differences in the perceptual span during reading: Evidence from the moving window technique. *Attention, Perception, & Psychophysics*, *77*(7), 2463-2475.

Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, *54*(1), 74-89.

Constantino, R. (1994). Pleasure Reading Helps, Even If Readers Don't Believe It (Open to Suggestion). *Journal of Reading, 37*(6), 504-5.

Constantino, R., Lee, S. Y., Cho, K. S., & Krashen, S. (1997). Free Voluntary Reading as a Predictor of TOEFL Scores. *Applied Language Learning, 8*(1), 111-18.

Constantino, R., Lee, S. Y., Cho, K. S., & Krashen, S. (1997). Free Voluntary Reading as a Predictor of TOEFL Scores. *Applied Language Learning, 8*(1), 111-18.

Cooke, N. A. (2018). Critical literacy as an approach to combating cultural misinformation/disinformation on the Internet. *Information literacy and libraries in the age of fake news*, 36-51.

Courage, M. L., Bakhtiar, A., Fitzpatrick, C., Kenny, S., & Brandeau, K. (2015). Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, *35*, 5-41.

Cuan-Baltazar, J. Y., Muñoz-Perez, M. J., Robledo-Vega, C., Pérez-Zepeda, M. F., & Soto-Vega, E. (2020). Misinformation of COVID-19 on the internet: infodemiology study. *JMIR public health and surveillance*, *6*(2), e18444.

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, *82*(4), 733.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*. Penguin.

Desjardins, R., Murray, S., Clermont, Y., & Werquin, P. (2005). Learning a living: First results of the Adult Literacy and Life Skills survey. *Statistics Canada catalogue*, (89-603).

Dietz, S., & Henrich, C. (2014). Texting as a distraction to learning in college students. *Computers in Human behavior*, *36*, 163-167.

Elley, W. B. (1992). *How in the world do students read? IEA study of reading literacy.* International Association for the Evaluation of Educational Achievement Tests.

Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading research quarterly*, 53-67.

Facione, P. A. (1990). The California Critical Thinking Skills Test--College Level. Technical Report# 2. Factors Predictive of CT Skills.

Ferguson, R. (2018, June 12). What made Ontarians vote for Doug Ford? *Toronto Star*. Retrieved August 19, 2020, from

https://www.thestar.com/news/queenspark/2018/06/12/what-made-ontarians-vote-for-doug-ford.html

Finney, S. J., Mathers, C. E., & Myers, A. J. (2016). Investigating the Dimensionality of Examinee Motivation across Instruction Conditions in Low-Stakes Testing Contexts. *Research & Practice in Assessment*, *11*, 5-17.

Fong, K., Mullin, J. B., & Mar, R. A. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(4), 370.

Fox, J. and S. Weisberg (2019) *An R Companion to Applied Regression, Third Edition* Sage Publications.

Godfroid, A., Winke, P., & Rebuschat, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. *Implicit and explicit learning of languages*, *48*, 325.

Gollom, M. (2019, July 1). Canadians say country split between ordinary folks and elites. But what is an elite? *CBC News*. Retrieved August 19, 2020, from https://www.cbc.ca/news/elites-canadians-politics-word-negative-1.5182816

Gottfried, J., & Shearer, E. (2016). *News Use Across Social Medial Platforms 2016*. Pew Research Center. Retrieved June 30, 2020 from https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

Grabe, W. (2014). Key issues in L2 reading development. In *CELC Symposium Bridging Research and Pedagogy* (pp. 8-18).

Gradman, H. L., & Hanania, E. (1991). Language learning background factors and ESL proficiency. *The Modern Language Journal, 75*(1), 39-51.

Grant, A., Wilson, A. M., & Gottardo, A. (2007). *The Role of Print Exposure in Reading Skills of Postsecondary Students With and Without Reading Disabilities.* Exceptionality Education International, 17(2), 175-194.

Grayson, J. P., Côté, J., Chen, L., Kenedy, R., & Roberts, S. (2019). A Call to Action: Academic Skill Deficiencies in Four Ontario Universities. Retrieved June 21, 2020 from https://skillsforuniversitysuccess.info.yorku.ca/files/2019/04/04-26-2019-AcademicSkills.pdf

Green, D. A., & Riddell, W. C. (2001). *Literacy, numeracy and labour market outcomes in Canada* (pp. 89-552). Ottawa: Statistics Canada.

Hango, D. (2014). *University graduates with lower levels of literacy and numeracy skills.* Statistics Canada / Statistique Canada.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2013). *Returns to Skills around the World: Evidence from PIAAC* (No. w19762). National Bureau of Economic Research.

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism studies*, *13*(5-6), 815-824.

Holmqvist, K., Holsanova, J., Barthelson, M., & Lundqvist, D. (2003). Reading or scanning? A study of newspaper and net paper reading. In *The Mind's Eye* (pp. 657-670). North-Holland.

Huey, E. B. (1968). *The Psychology and Pedagogy of Reading: With a review of the history of reading and writing and of methods, texts, and hygiene in reading.* Cambridge, MA: MIT Press, p. 6.

Humphries, J., Knight-Grofe, J., & Klabunde, N. (2009). *Canada First: The 2009 Survey of International Students*. Canadian Bureau for International Education (CBIE)/Bureau canadien de l'éducation internationale (BCEI).

Hurks, P. P. M., Vles, J. S. H., Hendriksen, J. G. M., Kalff, A. C., Feron, F. J. M., Kroes, M., ... & Jolles, J. (2006). Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *Journal of Clinical and Experimental Neuropsychology*, *28*(5), 684-695.

Jacobsen, W. C., & Forste, R. (2011). The wired generation: Academic and social outcomes of electronic media use among university students. *Cyberpsychology, Behavior, and Social Networking*, *14*(5), 275-280.

Kempe, C., Eriksson-Gustavsson, A. L., & Samuelsson, S. (2011). Are there any Matthew effects in literacy and cognitive development? *Scandinavian Journal of Educational Research*, *55*(2), 181-196.

Kim, H., & Krashen, S. (1998a). The author and magazine recognition tests as predictors of literacy development in Korean. *Perceptual and motor skills, 87*(3_suppl), 1376-1378.

Kim, H., & Krashen, S. (1998b). The author recognition and magazine recognition tests, and free voluntary reading as predictors of vocabulary

development in English as a foreign language for Korean high school students. *System*, *26*(4), 515-523.

Kirsch, I. S. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. US Government Printing Office, Superintendent of Documents, Washington, DC 20402 (Stock No. 065-000-00588-3).

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language learning*, *57*, 1-44.

Krashen, S. D. (1995). Free voluntary reading: Linguistic and affective arguments and some new applications. *Second language acquisition theory and pedagogy*, 187-202.

Krashen, S. D. (2004). *The power of reading: Insights from the research: Insights from the research.* ABC-CLIO.

Krashen, S., & Kim, H. (1998). The Author Recognition Test without foils as a predictor of vocabulary and cultural literacy test scores. *Perceptual and motor skills*, *87*(2), 544-546.

Kush, J. C., Watkins, M. W., & Brookhart, S. M. (2005). The temporal-interactive influence of reading achievement and reading attitude. *Educational Research and Evaluation, 11*(1), 29-44.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, *23*(6), 701-717.

Lane, J., & Murray, T. S. (2018). *Literacy Lost: Canada's Basic Skills Shortfall*. Canada West Foundation.

Lao, C. Y., & Krashen, S. (2000). The impact of popular literature study on literacy development in EFL: More evidence for the power of reading. *System, 28*(2), 261-270.

Lee, H., Seong, E., Choi, W., & Lowder, M. W. (2018). Development and assessment of the Korean Author Recognition Test. *Quarterly Journal of Experimental Psychology*, 1747021818814461

Lee, S. Y., Krashen, S., & Tse, L. (1997). The author recognition test and vocabulary knowledge: a replication. *Perceptual and motor skills*, *85*(3_suppl), 1428-1430.

Lepp, A., Barkley, J. E., & Karpinski, A. C. (2015). The relationship between cell phone use and academic performance in a sample of US college students. *Sage Open*, *5*(1), 2158244015573169.

Levine, L. E., Waite, B. M., & Bowman, L. L. (2007). Electronic media use, reading, and academic distractibility in college youth. *CyberPsychology & Behavior*, *10*(4), 560-566.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, *122*(3), 316.

Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation*, *61*(6), 700-712.

Mangen, A., & Van der Weel, A. (2016). The evolution of reading in the age of digitisation: an integrative framework for reading research. *Literacy*, *50*(3), 116-124.

Mangen, A., & Van der Weel, A. (2017). Why don't we read hypertext novels? *Convergence*, *23*(2), 166-181.

Mar, R. A., & Rain, M. (2015). Narrative fiction and expository nonfiction differentially predict verbal ability. *Scientific Studies of Reading*, *19*(6), 419-433.

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, *31*(3), 273-284.

Mason, B., & Krashen, S. D. (1997). Can extensive reading help unmotivated students of EFL improve? *ITL-International Journal of Applied Linguistics, 117*(1), 79-84.

Masterson, J., & Hayes, M. (2007). Development and data for UK versions of an author and title recognition test for adults. *Journal of Research in Reading*, *30*(2), 212-219.

McBride-Chang, C., Manis, F. R., Seidenberg, M. S., & Custodio, R. G. (1993). Print exposure as a predictor of word reading and reading comprehension in disabled and nondisabled readers. *Journal of Educational Psychology*, *85*(2), 230.

McLeod, B., & McLaughlin, B. (1986). Restructuring or automaticity? Reading in a second language. *Language learning*, *36*(2), 109-123.

McQuillan, J. (2006). The effects of print access and print exposure on English vocabulary acquisition of language minority students. *The Reading Matrix*, *6*(1).

McQuillan, J., & Au, J. (2001). The Effect of Print Access on Reading Frequency. *Reading Psychology*, *22*(3), 225-48.

Merkley, E. (2019). Anti-intellectualism, populism, and motivated resistance to expert consensus. *Public Opinion Quarterly.*

Miller Guron, L., & Lundberg, I. (2003). Identifying dyslexia in multilingual students: Can phonological awareness be assessed in the majority language? *Journal of research in Reading*, *26*(1), 69-82.

Miller, L. A. (2014). *The effects of extended time on reading comprehension performance for English as a Second Language college students: Is there a need for accommodations?* (Doctoral dissertation, Syracuse University).

Mokhtari, K., Reichard, C. A., & Gardner, A. (2009). The impact of internet and television use on the reading habits and practices of college students. *Journal of Adolescent & Adult Literacy, 52*(7), 609-619.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological bulletin*, *137*(2), 267.

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior research methods*, *47*(4), 1095-1109.

Morkes, J., & Nielsen, J. (1997). Concise, scannable, and objective: How to write for the Web. Retrieved June 20, 2020 from https://www.nngroup.com/articles/concise-scannable-and-objective-how-to-write-for-the-web/

Mumper, M. L., & Gerrig, R. J. (2017). Leisure reading and social cognition: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(1), 109.

Nation, I.S.P. & Beglar, D. (2007) A vocabulary size test. *The Language Teacher, 31(7)*, 9-13.

Neuman, S. B. (1988). The displacement effect: Assessing the relation between television viewing and reading performance. *Reading research quarterly*, 414-440.

Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. *Review of general psychology*, *3*(2), 101-117.

Pascarella, E. T., & Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research. Volume 2*. Jossey-Bass, An Imprint of Wiley. 10475 Crosspoint Blvd, Indianapolis, IN 46256.

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of academically adrift? *Change: The Magazine of Higher Learning*, *43*(3), 20-24.

Paulson, E. J. (2006). Self-selected reading for enjoyment as a college developmental reading approach. *Journal of College Reading and Learning, 36*(2), 51-58.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly (1982-)*, 283-319.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*, *84*(2), 203-244.

Postman, Neil (1988). *Conscientious objections: Stirring up trouble about language, technology and education.* New York: Vintage Books.

Protopapas, A., Parrila, R., & Simos, P. G. (2016). In search of Matthew effects in reading. *Journal of learning disabilities, 49*(5), 499-514.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2011). Matthew effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities, 44*(5), 402-420.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests: Danish institute for Educational Research. *Denmark Paedogiska, Copenhagen.*

Reder, S. (1998). Literacy selection and literacy development: Structural equation models of the reciprocal effects of education and literacy. *Literacy for the twenty-first century: Research policy, practices and the national adult literacy survey*, 139-157.

Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*(3), 235-257.

Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child development*, *86*(1), 23-36.

Rizopoulos, D. (2006) ltm: An R package for latent variable modelling and item response theory analyses. Journal of Statistical Software, 17(5), 1–25. URL http://www.jstatsoft.org/v17/i05/

Rodrigo, V., McQuillan, J., & Krashen, S. (1996). Free voluntary reading and vocabulary knowledge in native speakers of Spanish. *Perceptual and motor skills*, *83*(2), 648-650.

Roessingh, H., & Douglas, S. (2012). English language learners' transitional needs from high school to university: An exploratory study. *Journal of International Migration and Integration*, *13*(3), 285-301.

Rosenwald, M. S. (2014). *Serious reading takes a hit from online scanning and skimming, researchers say. The Washington Post, 6.* Retrieved June 20, 2020 from https://www.washingtonpost.com/local/serious-reading-takes-a-hit-from-online-scanning-and-skimming-researchers-say/2014/04/06/088028d2-b5d2-11e3-b899-20667de76985_story.html

Scales, A. M., & Rhee, O. (2001). Adult Reading Habits and Patterns. *Reading Psychology*, *22*(3), 175-203.

Sénéchal, M., LeFevre, J. A., Hudson, E., & Lawson, E. P. (1996). Knowledge of storybooks as a predictor of young children's vocabulary. *Journal of Educational Psychology*, *88*(3), 520.

Shatil, E., Share, D. L., & Levin, I. (2000). On the contribution of kindergarten writing to grade 1 literacy: A longitudinal study in Hebrew. *Applied Psycholinguistics*, *21*(1), 1-21.

Sheorey, R., & Mokhtari, K. (1994). The Reading Habits of Developmental College Students at Different Levels of Reading Proficiency. *Reading Improvement*, *31*(3), 156.

Sikora, J., Evans, M. D. R., & Kelley, J. (2019). Scholarly culture: How books in adolescence enhance adult literacy, numeracy and technology skills in 31 societies. *Social science research, 77, 1*-15.

Slinde, J. A., & Linn, R. L. (1979). The Rasch model, objective measurement, equating, and robustness. *Applied Psychological Measurement*, *3*(4), 437-452.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning*, *62*(2), 473-505.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, *333*(6043), 776-778.

Stainthorp, R. (1997). A children's author recognition test: A useful tool in reading research. *Journal of Research in Reading*, *20*(2), 148-158.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of education*, *189*(1-2), 23-55.

Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of educational psychology*, *85*(2), 211.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly,* 402-433.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811.

Statistics Canada, Authors. (2020). International student enrolments at Canadian public colleges and universities, 2017/2018. *Component of Statistics Canada catalogue no. 11-627-M.* Retrieved June 23, 2020 from https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2020006-eng.htm

Stowe, H.B. (2017). *Uncle Tom's Cabin.* Amazon Classics.

Strange, J. J. (2002). How fictional tales wag real-world beliefs. *Narrative impact: Social and cognitive foundations*, 263-286.

Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of Key Stage 1. *British Journal of Educational Psychology*, *74*(1), 15-36.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the student opinion scale to make valid inferences about student performance. *The Journal of General Education, 58,* 129-151.

Tunmer, W. E., & Chapman, J. W. (2006). Literate Cultural Capital at School Entry Predicts Later Reading. *New Zealand Journal of Educational Studies*, *41*(2), 183.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of educational Psychology*, *77*(4), 417.

Uncapher, M. R., Thieu, M. K., & Wagner, A. D. (2016). Media multitasking and memory: Differences in working memory and long-term memory. *Psychonomic bulletin & review*, *23*(2), 483-490.

Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *56*(1), 63-81.

Vander Beken, H., & Brysbaert, M. (2018). Studying texts in a second language: The importance of test type. *Bilingualism: Language and Cognition*, *21*(5), 1062-1074.

Volante, L., Klinger, D., Bilgili, Ö., & Siegel, M. (2017). Making sense of the performance (dis) advantage for immigrant students across Canada. *Canadian Journal of Education/Revue canadienne de l'éducation, 40*(3), 329-361.

Warner, C.D. (1896). The story of Uncle Tom's Cabin. *Atlantic Monthly, 78*(467), 311-322. Retrieved July 2, 2020 from http://utc.iath.virginia.edu/articles/n2es31at.html

Weiderholt, J. L., & Bryant, B. R. (2001). Gray oral reading test. *Vol. 4th. Austin, TX: Pro-Ed.*

Weinberger, J. (1996). A longitudinal study of children's early literacy experiences at home and later literacy development at home and school. *Journal of research in Reading*, *19*(1), 14-24.

Weingarten, H. P., & Hicks, M. (2018). *On Test: Skills Summary of Findings from HEQCO's Skills Assessment Pilot Studies*. Higher Education Quality Council of Ontario.

Wentworth, D. K., & Middleton, J. H. (2014). Technology use and academic performance. *Computers & Education*, *78*, 306-311.

West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*(5), 325-330.

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly,* 35-50.

Whitford, V., Pivneva, I., & Titone, D. (2016). Eye movement methods to investigate bilingual reading. In *Methods in bilingual reading comprehension research* (pp. 183-211). Springer, New York, NY.

Wolf, M. (2007). *Proust and the squid: The story and science of the reading brain*. New York: HarperCollins.

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. New York, NY: HarperCollins.

Wu, H. P., Garza, E., & Guzman, N. (2015). International student's challenge and adjustment to college. *Education Research International*, *2015*.

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, *36*(4), 661-672.

Appendices

Appendix A: Spelling Test

Word	Incorrect
attitude	
critisism	x
benafit	x
refrences	x
misary	x
psycology	x
political	
glamourous	x
reciept	x
available	
addmission	x
tounge	x

appreciate	
materilistic	x
independent	
chronicle	
seperate	x
senior	
behaviour	
atterney	x
sufficient	
efficiency	
implie	x
courtesy	
mortgage	
govenment	x
basicly	x
privalege	x
consequence	
sieze	x
suspicious	
prosedure	x

conveinient	x
insurance	
imminant	x
guitar	
elementary	
sacrifice	
commitment	
decrepit	
jeapordise	x
forfeit	
fulcrum	
annihlate	x
distinguish	
inquirey	x
sincirely	x
equivical	x
gaurantee	x
delecate	x
bachelor	
annual	

necesscarily	x
favourate	x
announcment	x
severe	
occurence	x
insatiable	
partitionining	x
asure	x
exhibition	
warrent	x
interrogate	
havoc	
conscientious	
parallel	
interpretation	
bureaucracy	
importent	x
negotiate	
proliferate	
vigilent	x

missellaneous	x
curiculum	x
plagarism	x
acomplice	x
pollution	
permanent	
aplause	x
subpoena	
accommodation	
attentsion	x
rendezvous	
subtlety	
honerable	x
inhibition	
classafied	x
assessor	

Appendix B: Reading Habits Questionnaire

(1) How good is the quality of your writing in English?

Very Poor - Poor - Neutral - Good - Very Good

(2) What is your reading speed in English?

Very Slow - Slow - Average - Fast - Very Fast

(3) How good is your reading comprehension?

Very Poor Poor Neutral Good Very Good

(4) How many hours a day do you spend reading and writing?

0-0.5 hours 1-2 hours 2-3 hours 3-4 hours 4+ hours

(5) What percentage of this time do you spend reading or writing texts on social media websites (0-100)?

Appendix C: Motivation Questionnaire

Please think about all of the tests that you just completed.
Mark the answer that best represents how you feel about each of the statements below.

* Required Field

Question	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Doing well on this test was important to me.	1	2	3	4	5
I engaged in good effort throughout this test.	1	2	3	4	5
I am not curious about how I did on this test relative to others.	1	2	3	4	5
I am not concerned about the score I receive on this test.	1	2	3	4	5
This was an important test to me.	1	2	3	4	5
I gave my best effort on this test.	1	2	3	4	5
While taking this test, I could have worked harder on it.	1	2	3	4	5
I would like to know how well I did on this test.	1	2	3	4	5
I did not give this test my full attention while completing it.	1	2	3	4	5
While taking this test, I was able to persist to completion of the task.	1	2	3	4	5

Appendix D: ANOVA outputs of regression models for analyses of effects of year of study and language background on reading habits.

Table D1. ANOVA for Reading Habits Questions 1. *"How good is the quality of your writing in English?"* N = 558 before trimming, 538 after trimming, Adjusted R-squared: 0.249.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Language Background	1	81.517	81.517	132.953	< 0.001
Year of Study	1	22.607	22.607	36.871	< 0.001
Year of Study : Language Background	1	6.446	6.446	10.514	0.001
Residuals	531	325.569	0.613	NA	NA

Table D2. ANOVA for Reading Habits Questions 2. *"What is your reading speed in English?"* N = 558 before trimming, 538 after trimming, Adjusted R-squared: 0.093.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Language Background	1	30.339	30.339	44.001	< 0.001
Year of Study	1	6.958	6.958	10.091	0.002

Year of Study : Language Background	1	2.550	2.550	3.698	0.055
Residuals	531	366.128	0.690	NA	NA

Table D3. ANOVA for Reading Habits Question 3: *"How good is your reading comprehension?"* N = 558 before trimming, 538 after trimming, Adjusted R-squared: 0.171.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Language Background	1	50.758	50.758	83.715	< 0.001
Year of Study	1	13.301	13.301	21.937	< 0.001
Year of Study : Language Background	1	4.664	4.664	7.692	0.006
Residuals	531	321.958	0.606	NA	NA

Table D4. ANOVA for Reading Habits Question 4: *"How many hours a day do you spend reading and writing?"* N = 558 before trimming, 538 after trimming, Adjusted R-squared: 0.081.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Language Background	1	48.474	48.474	37.197	< 0.001
Year of Study	1	9.875	9.875	7.578	0.006
Year of Study : Language Background	1	6.678	6.678	5.125	0.024
Residuals	531	691.971	1.303	NA	NA