INDIVIDUAL DIFFERENCES IN READING PROFICIENCY: INVESTIGATING INFLUENCING FACTORS AND HOW THEY INTERACT

INDIVIDUAL DIFFERENCES IN READING PROFICIENCY: INVESTIGATING INFLUENCING FACTORS AND HOW THEY INTERACT

By:
KELLY A. NISBET, M.Sc., BSc

A Thesis Submitted to
the Department of Linguistics and Languages
and the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2021) Hamilton, Ontario (Cognitive Science of Language)
TITLE: Individual Differences in Reading Proficiency: Investigating Influencing Factors and How They Interact
AUTHOR: Kelly A. Nisbet, MSc (University of Western Ontario), BSc (University of Alberta)
SUPERVISOR: Dr. Victor Kuperman
NUMBER OF PAGES: xv, 105

Lay Abstract

This thesis investigates how individual differences influence reading proficiency. Specifically, it asks how the ways in which people differ on certain reading-related skills and cognitive abilities can determine how well they read. Using different measures of proficiency, a variety of data collection and statistical methods, and looking across different populations, the goal of this thesis was to examine the ways in which people differ in these skills and abilities, how these differences interact, and the resulting impact on reading proficiency. This thesis resulted in three significant contributions to the field. First, it made available a new application for collecting data on an important variable in reading research – cloze probability. In addition, it culminated in the development of a novel statistical method that demonstrates how an individual's linguistic background can influence their reading fluency. Finally, a new connection was found between two important cognitive factors that interact to influence reading comprehension.

Abstract

This thesis investigates individual differences and their impact on reading proficiency using different measures of proficiency, a variety of data collection and statistical methods, and different populations. The goal was to examine the impact that individual differences in certain reading-related skills and cognitive abilities have on reading proficiency and how these differences interact.

Through three key studies that make up this thesis, several important discoveries and contributions were made to the field. Chapter 2 introduces an easy-to-use application for measuring cloze probability. 'ClozApp', was created and made publicly available, along with a user manual and sample code for programming. Chapter 3 contributed through the development of a novel statistical method used to analyze variance between populations with different linguistic backgrounds. This method was used to demonstrate how an individual's linguistic background (i.e., whether they were first- or second-language speakers of English) impacted how individual differences in reading skills influence their reading fluency, as indicated through their eye-movements. This statistical prediction method is open source and was made widely available for use along with sample data and code. In Chapter 4, a new connection was found between two important cognitive factors that are well-known in the reading literature: statistical learning and motivation. Using mediation analyses, this project discovered an interaction between these factors that further highlights the ways they impact reading proficiency.

This thesis demonstrates a comprehensive approach to investigating individual differences in reading proficiency in the following ways: (i) both reading fluency and comprehension were investigated as measures of reading proficiency, (ii) data collection included a variety of reading-related skills, cognitive abilities, and group differences, and (iii) unique statistical analysis methods were utilized to investigate both individual and group differences. This thesis highlights important new discoveries and makes significant lasting contributions to the field of reading research.

Acknowledgements

First and foremost, I'd like to thank my supervisor Victor Kuperman. I could not ask for a more knowledgeable, patient, kind, and compassionate mentor throughout this process. I am forever grateful for all the feedback and time spent programming, brainstorming, hypothesizing, and simply thinking about science. I have learned so much from experiencing your approach to tackling big questions, and even bigger datasets, and feel so fortunate to have had the opportunity to work with you over the past four years. You truly made me feel like a colleague and helped me gain the confidence I needed to push projects forward and learn from the failures. As an added bonus this all came with a great sense of humor, a mutual fondness for road trips, and plenty of memories bonding with idols over drinks and karaoke. Truly, thank you.

I would also like to thank my committee members, John Connolly and Elizabet Service. I will never forget the many stories I have had the pleasure of hearing, and the insightful and often hilarious conversations we have had. John, thank you for promoting such a wonderful environment that combines science, storytelling, and mutual bonding over the good and the bad that academia has to offer. Lili, your assistance with grant writing, brainstorming research projects, and your insightful questions have allowed me to push my research into new areas. Your patience and thoughtfulness truly made this an enjoyable experience.

My co-authors have been so helpful and provided considerable insight and advice. Thank you to Michel Généreux, Blake Anderson, Raymond Bertram, Aleks Pieczykolan, and Charlotte Erlinghagen. I'd like to specifically thank Noam Siegelman for the guidance and teachings throughout my graduate studies which have been instrumental in my data analysis and writing development - I have never looked forward to seeing so many red edits on a document. And my previous mentors, Thomas Spalding and Christina Gagne, thank you for your continuing support and guidance, and for providing me the opportunity and tools needed to step into academia.

I am very privileged to have been financially supported throughout my graduate studies. I would like to express my gratitude to McMaster University and the department of Linguistics and Languages for my teaching positions and support in travelling to disseminate my research, ARiEAL for the financial support for workshops and conferences, and the Ontario Government for their financial support through the Ontario Graduate Scholarship program.

I consider myself incredibly fortunate for the amazing line-up of lab mates over these past few years. You have all made this journey not only possible, but thoroughly enjoyable. To all the wonderful research assistants, thank you all for your hard work and attention to detail in data collection – this thesis would not exist without you! A special thanks to my lab mates and friends Davide Gentile, Melda Coskun, Nadia Lana, Sean McCarron, Gaisha Oralova, Kaitlyn Battershill, and Heather Wild for all the editing, practice talks, experiment brainstorming, code debugging, and all around great times. I specifically want to thank Connie Imbault for being the most amazing higher-up I could have asked for. Thank you for showing me the ropes and being an instrumental support during my time at McMaster. To the wise PostDocs who have been here before, Aki Kyrolainen, Dan Schmidke, and Vincent Poretta, thank you for showing that it is possible to get through this demanding process and giving us all hope for an optimistic future, even when things look bleak. And an extra thanks to Aki for sharing your genius towards programming and stats help whenever I got stuck. It was a pleasure working with everyone in the Reading lab and I'm so happy to have built such great friendships along the way.

Of course, my wonderful network of support extended well past the lab. A huge thanks to the ever-knowledgeable Chia-Yu Lin, for always being there to help with what feels like everything and lending an ear at the exact right times. You have such a knack for ensuring those around you feel heard and cared for and I hope you know just how appreciated you are. Chelsea Whitwell, I'm so happy you adopted me into your collection of introverts. Your loyalty, honesty, and fierce caring for those close to you are enviable and I know we've built a friendship for life. To all the wonderful friends and colleagues I got to spend time with at ARiEAL, including Nathalee Ewers, Rober Boshra, Olivier Mercier, Frederico Prado, Mariam Bekhet, Fareeha Rana, Kyle (and Danielle) Ruiter, Narcisse Torshizi, Richard Mah, and many more – thank you for making this time so special.

Moving across the country to a new city can be hard, so I'm thankful to have had the opportunity to meet Sheereen Harris, the kindest, most inspiring and uplifting gem of a human being. I am so happy to be able to call you my friend (and former roommate). I also need to thank the wonderful Harris family for welcoming me in and being my surrogate family for all the holidays and get-togethers ensuring I was never alone. Moving away also means leaving cherished friends at a distance as well. To Sofia Varzaru, thank you for always being there for me no matter what. As one of my longest known friends, I value our friendship immensely, and always appreciate your perspective and your drive. To Nicole Sullivan, thank you for all you did when I needed you most and for making some amazing memories along the way. I look forward to many more wonderfully stimulating conversations, debates, and trips through the mountains.

I'm beyond grateful to have the support of my partner, Shobal. Over this past year especially you have been instrumental in me pursuing my goals. Thank you for believing in me, gifting me the space and time needed to finish this thesis in the craziest of times, for always finding ways to make me laugh everyday, and above all else accepting and loving me as my true authentic self. I also have to thank my furry companion through it all, Monty. For reminding me ever so persistently (as only a cat can) to put down my laptop, book, or literally anything else in his way, to make time for some love and cuddles.

And finally, to my family! I know how unique and rare it is to have such a close family dynamic like ours and I wouldn't trade it for the world. I love you all so very much and can confidently say I would not be here without you. Geoff, thank you for being there for me through the hard times and always hyping me up when things go my way. Jamie, I feel so lucky to get you as both a sister and a best friend. You are the most thoughtful, caring, and resilient person I know, and I hope to continue learning from you always. Thank you for being my safe place, my travel companion, and for all of our seemingly endless chats. Dad, I'm so happy you passed along the ability to work under pressure, and am less happy about our shared fondness for procrastination. Through all the good times and the hard times 'I Byd It'. And of course, Mom, you are the strongest person I know, and I am, and always will be, in awe of your positive outlook and ability to love so incredibly hard. Thank you for teaching me to be me, to follow my gut, and to trust that I can accomplish anything I put my mind to. Your unwavering support through all life has thrown my way means more to me than I could possibly put into words. Lastly, to my biggest supporter, Grandy. I'm still in disbelief that he isn't here to see me reach this milestone, but I know he would be so proud and would inevitably be calling me Dr. Pook.

It's been a long four years, with more than a few unexpected twists and turns (including a global pandemic), but I made it to the end and am proud of what I've accomplished. To everyone mentioned here and everyone else who had a hand in helping me get to this point, thank you.

Table of Contents

Abstract . . . . *iv*
Acknowledgments . . . . *v*
List of Figures . . . . *ix*
List of Tables . . . . *xi*
Declaration of Academic Achievement . . . . *xiv*

CHAPTER 1: Introduction 1

Motivation 1

Background 2

Objectives 7

Approach 9

Introduction to Chapters 9

References 14

CHAPTER 2: Clozapp: A Java Application for Collecting and Recording Cloze Probability Norms 23

Background 24

Clozapp Overview 25

Experiment 33

Methods 33

Results and Discussion 34

Availability 35

References 36

CHAPTER 3: Quantifying the Difference in Reading Fluency Between L1 and L2 Readers of English 37

Background 38

Current Study 40

Methods 42

Statistical Prediction Procedure. 45

Results 46

General Discussion 59

References. 62

Supplementary Materials 67

CHAPTER 4: The Validity of the Link Between Statistical Learning and Reading Comprehension: Testing for Mediating Variables 69

Introduction 70

Methods 72

Statistical Considerations 74

Results 75

General Discussion 82

Appendix A 85

Appendix B 86

References 88

CHAPTER 5: Discussion and Conclusion 93

Overview 93

Implications and Significance 94

Limitations and Future Directions 97

Conclusion 102

References 103

List of Figures

CHAPTER 2

Figure 1 26
Clozapp upon opening the Java file.
Figure 2 29
After loading the experiment markup file, participants will be prompted to fill in the demographic information requested by the experimenter. Pressing 'Start' begins the experiment.
Figure 3 30
After pressing 'Start', the participants are given the instructions for the experiment as well as the 'in progress' text reminding them of the task. Buttons to 'Pause', ask for 'Help', or to 'Quit' or 'Save' the experiment are also available.
Figure 4 31
A blank '____' indicates the word(s) to guess. Users type in their response and press <ENTER>.
Figure 5 32
The default format of recording a session.

CHAPTER 3

Figure 1 48
Boxplots of the three component skill tests by L1.
Figure 2 51
Predicted reading rates for hypothetical L2 readers with median and maximum performance in component skills (in grey) with prediction intervals (error bars), and observed median (and for the Canadian cohort, the 10th and 90th percentile) reading rates from the three cohorts (in black).
Figure 3 53
Total fixation time as a function of length, for each language and skill level.
Figure 4 54
Total fixation time as a function of frequency for each language and skill level.
Figure 5 55
Observed median total fixation times for Finnish (black open circles), and English (filled black triangles), and predicted total fixation times for hypothetical Finnish (small grey circles) for all combinations of length and frequency using median and top skills. For legibility purposes, prediction intervals are not shown.
Figure 6 56
Observed median total fixation times for German (black open diamonds), and English (filled black triangles), and predicted total fixation times for hypothetical German (small grey diamonds) for all combinations of length and frequency using median and top skills. For legibility purposes, prediction intervals are not shown.

CHAPTER 4

Figure 1 77
Mediation analysis of the effect of VSL on reading comprehension with motivation as the mediator (N = 188, bootstrap simulations = 10000). The significance values for the individual connections in the triangle come from the estimates of each connection in the base models found in the Appendix (B1 and B2).
Figure 2 79
Mediation analysis of the effect of VSL on reading comprehension with ART as the mediator (N = 188, bootstrap simulations = 10000). The significance values for the individual connections in the triangle come from the estimates of each connection in the base models (Appendix B3 and B4).
Figure 3 80
Mediation analysis of the effect of VSL on reading comprehension with motivation and ART as parallel mediators (N = 188, bootstrap simulations = 1000). Significance values for each connection in the figure can be found in Appendix B1-B6.
Figure 4 82
Plot of the interaction between VSL and motivation in their influence on Reading Comprehension. VSL score has been broken down into 3 bins representing accuracy at the mean – 1 standard deviation, the mean, and the mean + 1 standard deviation (N = 188).

List of Tables

CHAPTER 2

Table 1 34
Descriptive statistics for Clozapp, Schnoebelen and Kuperman (2010), Reichle et al. (1998), and Clozapp.
Table 2 34
Descriptive statistics for Clozapp, Schnoebelen and Kuperman (2010), and Reichle et al. (1998), when words are removed if they did not elicit a single correct guess in all three corpora (in other words, "full misses" were excluded).

CHAPTER 3

Table 1 46
Descriptive statistics for the three component skills tests by L1.
Table 2 49
Descriptive statistics for the main eye-tracking variables for each cohort. RR = reading rate. ToFD = total fixation duration.
Table 3 50
Linear regression model of reading rate as a function of language with German as baseline.
Table 4 50
Reading rate as a function of component skills: German. SD of the by-items random intercepts is 27.99, SD of the by-subject random intercepts is 52.10; the model residual SD is 49.20; N = 1618.
Table 5 50
Reading rate as a function of component skills: Finnish. SD of the by-items random intercepts is 24.48; SD of the by-subject random intercepts is 51.32; the model residual SD is 40.69; N = 1920.
Table 6 58
Top predicted combinations for German hypothetical readers. Possible values for ART (median 7.5, high 41), Spelling (median 77, high 87), and vocabulary (median 72.5, high 100).
Table 7 58
Top predicted combinations for Finnish hypothetical readers. Possible values for ART (median 7.5, high 41), Spelling (median 77, high 87), and Vocabulary (median 72.5, high 100).
Table S1 67
Fixed effects of the model fitted to exposure to print (ART) scores with German as the baseline.
Table S2 67
Fixed effects of the model fitted to the spelling test scores with German as the baseline.
Table S3 67
Fixed effects of the model fitted to vocabulary scores with German as the baseline.

Table S4 67
Fixed effects of language on total fixation time with German as the reference language. By-subject random intercepts SD = 0.25, by-item random intercepts SD = 0.17, residual SD = 0.46, N = 26416.
Table S5 67
Regression model for ToFD: German cohort. SD of the by-items random intercepts is 0.0690; SD of the by-subject random intercepts is 0.2811; the model residual SD is 0.4388; N = 9259.
Table S6 68
Regression model for ToFD: Finnish cohort. SD of the by-items random intercepts is 0.0164; SD of the by-subjects random intercepts is 0.1274; the model residual SD is 0.4521; N = 7881.

CHAPTER 4

Table 1 75
Descriptive statistics for variables of interest (n=188).
Table 2 76
Correlations between the variables of interest (N = 188). Pearson Correlation Coefficients are reported below the diagonal and p-values above the diagonal. The strength of the correlation is shown in greyscale from weak (in white) to strong (in dark grey).
Table 3 77
Mediation analysis of the relationship between VSL and reading comprehension mediated by motivation (N = 188). The 'Average Causal Mediation Effect' (ACME), 'Average Direct Effect' (ADE), 'Total Effect', and the 'Proportion Mediated' are displayed. Nonparametric bootstrap confidence intervals with the percentile method simulated in 10000 iterations.
Table 4 78
Mediation analysis of relationship between VSL and reading comprehension mediated by ART (N = 188). The 'Average Causal Mediation Effect' (ACME), 'Average Direct Effect' (ADE), 'Total Effect', and the 'Proportion Mediated' are displayed. Nonparametric bootstrap confidence intervals with the percentile method simulated in 10000 iterations.
Table 5 79
Mediation analysis of relationship between VSL and reading comprehension mediated by both motivation and ART (N = 188). The Total Effect, Direct Effect and the Indirect Effects of both the first (M1) and the second mediating variable (M2) are displayed. Nonparametric bootstrap for standard errors simulated in 1000 iterations.
Table 6 81
Linear regression model fitted to reading comprehension with a critical interaction VSL and motivation. The model residual standard error is 5.43 on 182 degrees of freedom; $R^2 = 0.2183$.
Table 7 81
Estimates for the VSL scores are the different levels of accuracy of interest: the mean – 1 standard deviation (0.49), the mean (0.66), and the mean + 1 standard deviation (0.82). The confidence level used is 95% on 182 degrees of freedom.
Table B1 86
Regression model for motivation as explained by VSL. The model residual standard error

is 0.972 on 186 degrees of freedom; R^2 = 0.060.
Table B2 86
Regression model for reading comprehension as explained by VSL and motivation. The model residual standard error is 5.686 on 185 degrees of freedom; R^2 = 0.129.
Table B3 86
Regression model for ART as explained by VSL. The model residual standard error is 7.205 on 186 degrees of freedom; R^2 = 0.070.
Table B4 86
Regression model for reading comprehension as explained by VSL and ART. The model residual standard error is 5.759 on 185 degrees of freedom; R^2 = 0.106.
Table B5 86
Regression model for Motivation as explained by VSL and ART. The model residual standard error is 0.3881 on 185 degrees of freedom; R^2 = 0.100.
Table B6 87
Regression model for reading comprehension as explained by VSL, motivation, and ART. The model residual standard error is 5.643 on 184 degrees of freedom; R^2 = 0.147.
Table B7 87
Regression model for reading comprehension as explained by VSL, ART, and the interaction between VSL and ART. The model residual standard error is 5.775 on 184 degrees of freedom; R^2 = 0.106.
Table B8 87
Regression model for reading comprehension as explained by VSL, spelling, and the interaction between these VSL and spelling. The model residual standard error is 5.540 on 184 degrees of freedom; R^2 = 0.177.

Declaration of Academic Achievement

This dissertation constitutes a 'sandwich' thesis according to the School of Graduate Studies, McMaster University. The three articles presented within my thesis constitute much of my doctoral work throughout my graduate studies and are therefore included in my thesis. I am the primary author of these works and was responsible for the literature reviews, write-up of the manuscripts, design of the studies, and analysis of the data. The publication status for each paper, along with the roles of the co-authors are outlined below.

Chapter 2: Reprint of an article published in *The Mental Lexicon* December 2019
"Nisbet, K., Généreux, M., Anderson, B., & Kuperman, V. (2019). Clozapp: A Java application for collecting and recording Cloze probability norms. *The Mental Lexicon*, *14*(3), 399-414. doi: 10.1075/ml.20004.nis"

- Nisbet, K.
    - Literature review, study design, application configuration/design, data collection, data analysis, R code write-up, user manual write-up, and manuscript write-up, preparation, and revisions
- Généreux, M.
    - Application design, application configuration
- Anderson, B.
    - Application configuration, application editing
- Kuperman, V.
    - Study design, data analysis, manuscript and user manual editing/revisions

Chapter 3: Reprint of an article published in *Studies in Second Language Acquisition* April 2021
"Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2021). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Manuscript accepted to Studies in Second Language Acquisition*, 1-28. doi: 10.1017/S0272263121000279"

- Nisbet, K.
    - Literature review, study design, data analysis, R code write-up, and manuscript write-up, preparation, and revisions
- Bertram, R.
    - Study design, manuscript editing/revisions
- Erlinghagen, C.
    - Data collection, manuscript editing
- Pieczykolan, A.
    - Data collection, manuscript editing
- Kuperman, V.
    - Study design, data analysis, manuscript editing/revisions

Chapter 4: Reprint of an article submitted for publication to *Cognitive Science* March 2021
"Nisbet, K., Siegelman, N., & Kuperman, V. (2021). The Validity of the Link Between Statistical Learning and Reading Comprehension: Testing for Mediating Variables. *Submitted March 2021 to Cognitive Science*."

- Nisbet, K.
    - Literature review, study design, data analysis, R code write-up, and manuscript write-up and preparation
- Siegelman, N.
    - Study design, manuscript editing/revisions
- Kuperman, V.
    - Study design, manuscript editing/revisions

Additional Achievements

In addition to the projects listed above that are included in my thesis, I was a contributor on three additional projects that were published or submitted for publication during my graduate studies. See details below.

Kuperman, V., Siegelman, N., Schroeder, S., Alexeeva, A., Acartürk, C., Amenta, S., Bertram, S., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Marelli, M., **Nisbet, K.**, Papadopoulos, T. C., Protopapas, A., Savo, S., Shalom, D., Slioussar, N., Stein, R., Sui, L., Taboh, A., Tønnesen, V., and Usual, K. A. (2021). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus (MECO). *Manuscript submitted to 'Studies in Second Language Acquisition'* March 2021.

Spalding, T. L., Gagné, C. L., **Nisbet, K. A.**, Chamberlain, J. M., & Libben, G. (2019). If birds have sesamoid bones, do blackbirds have sesamoid bones? The modification effect with known compound words. Frontiers in psychology, 10, 1570. DOI: https://doi.org/10.3389/fpsyg.2019.01570

Gagné, C. L., Spalding, T. L., **Nisbet, K. A.**, & Armstrong, C. (2018). Pseudo-morphemic structure inhibits, but morphemic structure facilitates, processing of a repeated free morpheme. Language, Cognition and Neuroscience, 33(10), 1252-1274. DOI: https://doi.org/10.1080/23273798.2018.1470250

CHAPTER 1.

Introduction

Motivation

For many, reading is a natural straightforward process requiring very little effort. In fact, people often engage in this behaviour throughout much of their everyday lives, whether this is reading the bus schedule, checking emails at work, or deciding on the appropriate dosage for cough medicine, without giving it much thought. However, the ability to utilize this skill while navigating daily life is not the same for everyone. This is an important issue, since difficulty in reading proficiency severely limits one's employability, health literacy, as well as social and professional well-being (Dejardins et al., 2005; Green & Riddell, 2001; Grenier et al., 2008).

Literacy is defined as "the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts" (United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics, 2018). Conversely, functional illiteracy is the inability to comprehend simple written texts such as passages in textbooks, emails from coworkers, or drug prescriptions. Globally, approximately 20 percent of the population is illiterate, with a disproportionate number of those individuals being women and girls as a result of limited access to education in many parts of the world (UNESCO, 2015). Low literacy rates can also be found in countries that have highly rated education systems and good access to these systems. Looking domestically, the International Adult Literacy and Skills Survey (2005) demonstrated that a staggering number of Canadians struggle with the critical skill of reading, with over 15 million Canadians being functionally illiterate (OECD and Statistics Canada, 2005). A more recent assessment suggests that the literacy skills of 48 percent of Canadians fall below high school level education, and 17 percent of Canadians function at the lowest level of literacy skill (OECD Programme for the International Assessment of Adult Competencies (PIAAC), 2013). Even in educational settings that typically see high levels of literacy, such as at the post-secondary education level, many students underperform in areas related to literacy skills. Recent provincial, national, and international surveys find that on average one in four university-educated and one in two college-educated Canadians do not possess literacy skills adequate for successfully performing in the modern labour market (Grayson et al., 2019; Hango, 2014; Weingarten & Hicks, 2018). This leads to important questions as to why some students are not attaining appropriate reading skills even after completing the highest level of public education.

These difficulties are even more pronounced for non-English and/or non-French speaking residents of Canada. It has been reported that approximately 21.9% of Canadians have immigrated from another country, and in the past five years alone, there have been over one million new immigrants to Canada, making up approximately 3.5% of Canada's total population (Statistics Canada, 2017). Upon arrival, second language (L2) acquisition is of critical importance for the thousands of immigrants who need to function in Canadian society who speak a first language (L1) that is not English. According to the 2016 census, over 7.7 million

Canadians report having an L1 other than English or French. This number has significantly increased by 13.3% over the past 5 years, as reported in the most recent census (Statistics Canada, 2017).

Literacy is an essential skill that has the ability to impact all aspects of an individual's life. Therefore, determining what factors contribute to high levels of literacy is a critically important area of investigation. Increasing knowledge of how to improve overall reading proficiency of Canadians, and the quality of English language learning for new Canadians, has tangible societal and economic impacts for our country.

Background

Individual Differences in Reading

While functional literacy requires the competence of both reading and writing, this thesis is primarily concerned with the foundational skill of reading. The ability to easily process text and understand the intended thoughts and ideas of the author is at the core of reading proficiency. This process importantly addresses both the accuracy (or comprehension ability) of a reader, and the speed (or fluency) with which the process can be completed. Breaking this process down into the required building blocks illustrates just how intricate this seemingly easy behaviour truly is. Reading is a complex, multi-dimensional process, and demonstrating capability in any area of reading proficiency requires simultaneous mastery of numerous reading-related skills and having certain general cognitive abilities. These skills and abilities and the reading process in general can additionally be heavily influenced by language background and other environmental factors. The substantial list of variables that demonstrably impact reading proficiency is impressive and continuously growing.

Given the vast number of variables that are known to impact reading proficiency, it is not surprising that understanding what constitutes reading ability is a complex problem. This is further compounded by the fact that every individual varies in their proficiency across the ever-growing list of related skills and cognitive factors. One possible approach taken by many in the field is to aggregate the data across individuals to better study the variables independently of individual variation. However, in taking this approach, key differences that may be at the root of how individuals become more or less proficient in reading are lost. The study of these individual differences in the realm of reading is crucial when determining which skills and cognitive factors are the most impactful when it comes to reading proficiency. This important information can go on to inform literacy education and reform, further highlighting the motivation behind this thesis.

In order to contribute beneficial information that can be utilized to improve literacy rates in the future, it is of critical importance to know which skills, abilities, and cognitive factors are at the forefront for impacting reading proficiency. Examining important areas in which individuals differ, and determining which factors show the greatest association with varying levels of reading proficiency is an impactful first step. Given the importance of literacy and reading and the societal implications, research on individual differences in reading has been undertaken by many researchers with great success (see review by Afflerbach, 2015). The following section will provide an overview of the current literature regarding some core factors known to impact reading proficiency.

Reading Skills and Abilities

Reading requires a number of different skills and abilities (also referred to as component skills) that exist at all levels of language, from letter identification to understanding the complex rules of grammar (Oakhill et al., 2003; Vellutino et al., 2007). These skills and abilities are central to the question of which factors are most involved in reading proficiency and how they interact. This section will highlight the complex component skills utilized by individuals during the process of reading text.

For any given language that is represented through written text, there is an accompanying script and alphabet denoting the language. The nature of the organization of these scripts has an impact on how individuals go about decoding this information to engage in reading. In other words, before any type of processing can occur, familiarity with the script of a given language is critical (Ellis et al., 2004; Taylor & Olson, 1995). The ability to quickly identify the components of the alphabet of a given language - that is, one's speed and ability in letter/symbol identification in a particular script - is a necessary skill that correlates with both reading acquisition and proficiency (Perfetti, 1992; Wiley et al., 2016). In addition, knowing and being able to reproduce the sounds of a language (even without saying them out loud), including phonological and rhyme awareness, also aids in the process of reading (Melby-Lervag et al., 2012; Share, 1995; Wood et al., 2009). The skill of 'decoding' written text, which is the ability to map the language-specific sounds to the letters in the process of identifying words, has also been proven influential in reading proficiency. Reading proficiency requires both the knowledge of the sounds of the letters (or symbols) on their own, as well how these letters sound when they combine to form a word (Garcia & Cain, 2014; Perfetti et al., 1987; Tunmer & Nesdale, 1985). This understanding of specific letter combinations not only aids in decoding ability, but can also provide important cues regarding how words break down into meaningful elements, known as morphological awareness. Demonstrating high morphological awareness has been shown to correlate with improved reading comprehension and general reading ability (Deacon et al., 2014; Perfetti et al., 2005). In addition to knowledge regarding units of meaning, an understanding of the syntactic rules that govern sentence structure is also beneficial to developing high reading proficiency (Brimo & Apel, 2017; Deacon & Kieffer, 2018).

These above-mentioned skills are highly inter-related, so demonstrating a high proficiency or knowledge in one area of expertise is likely to impact the proficiency in another. Further, it has been shown that these skills can work together to influence reading proficiency (Carlson et al., 2013; Deacon & Kirby, 2004; Plaza & Cohen, 2003). They are also well known to cooperate in impacting additional reading-related abilities, such as spelling (Apel et al., 2012; Beringer et al., 2010; Zhao et al., 2017).

Taken together, these skills are all necessary components of becoming a skilled reader, but they do not tell the whole story. While these skills have been researched thoroughly regarding their relationship to reading proficiency, additional factors have also been identified that impact the reading process. This thesis will primarily focus on a select subgroup of higher-level skills, that take into account a number of these previously stated foundational skills, and focus more on the process of integrating these skills towards the goal of efficiently processing and comprehending text. The central component skills that this thesis investigates are spelling ability and vocabulary size.

Spelling and reading have similar component skills and influences that lead to a person becoming highly proficient in both abilities, as mentioned previously. Additional research has shown that spelling ability can also directly impact reading proficiency. Spelling has been a highly studied component skill due to its association to reading acquisition and reading ability in general (Ehri, 2000; Graham & Santangelo, 2014; Moats, 2005; Ouellette, & Sénéchal, 2017). The commonly understood connection between spelling and reading lies in the ability for an individual to correctly identify the word, which leads to meaning retrieval. The more skilled an individual is at spelling, the faster this identification process can happen, allowing for an increased speed of reading. Previous studies have shown the impact of incorrect spelling on the word retrieval process, which further highlights the importance of maintaining a high spelling ability, given its impact on reading fluency (Falkauskas & Kuperman, 2015; Kuperman et al., 2021; Protopapas et al., 2013; Rahmanian & Kuperman, 2019). In general, poor spelling ability has been shown to correlate with slower reading (Martin-Chang et al., 2014; Oulette et al., 2017). Furthermore, Veldre et al., (2017) demonstrated that having a higher spelling ability contributed to maintaining high reading speed and fluency despite unusual word spacing in text. This well-connected and critical skill has led to important insights in reading proficiency and is an important variable to consider.

Vocabulary size is another critical variable that is known to influence reading proficiency. Individuals vary widely in the size of their vocabulary, and this variable correlates strongly with reading ability. Research has found that a larger vocabulary translates to higher reading proficiency (Moghadam et al., 2012; Nation, 2001; Qian, 2002). Strong ties have also been identified between the vocabulary of a given text and its difficulty, suggesting that individuals who have a larger vocabulary will have increased reading comprehension on difficult texts (Stahl, 2003). The study of individual differences in vocabulary and the relationship to reading proficiency is of high importance given these correlations. Further investigating how vocabulary size impacts both reading fluency and comprehension is also a valuable area of study since its impact has been demonstrated for both L1 and L2 speakers (Akbarian, 2010; Baleghizadeh & Golbin, 2011; Elgort & Nation, 2010; Schmitt et al., 2011).

Another reading-related factor that has drawn a lot of attention over the years is the influence an individual's exposure to printed material has on reading proficiency. The quantity of printed material that an individual has interacted with is thought to translate to an increased experience in reading, which can improve overall reading proficiency. This relationship between the amount of reading an individual does and their reading proficiency has been proven time and again (Au, 2001; McQuillan, 2006; Mol & Bus, 2011; Paulson, 2006). This variable is also well studied in reference to the well-known 'Matthew effect', which draws a mutual relationship between reading ability and time spent reading. This relationship outlines a situation in which better readers increase their exposure, and that in turn continues to improve their reading ability, whereas poor readers decrease their exposure making it difficult to improve their reading ability (Kempe et al., 2011; Stanovich, 1986). Exposure to print has been found to correlate with both comprehension and speed of reading (Acheson et al., 2008; Landi, 2010). Understanding the relationship between amount of reading and reading ability is therefore an important area of study when discussing literacy.

Finally, another important factor to consider when looking at an individual's reading proficiency relates to their ability to predict upcoming words when processing text. At the forefront of this research is a measure known as cloze probability which estimates an individual's ability in predicting upcoming text. This measure, typically accomplished by providing an incomplete text to participants and asking them to predict the missing information, has led to important insights in reading proficiency over the years. Cloze probability has been used as a measure of how accurately an individual can predict single words, an upcoming string of words in a sentence, and up to a whole paragraph of text (one word at a time). The variability in the probability of correctly guessing upcoming text has been shown to correlate with reading speed and comprehension (Gellert, & Elbro, 2013; Smith & Levy, 2013; Staub et al., 2015). Understanding the nature of the impact of this variable on reading proficiency is an ongoing and critical question in the literature, and more information is required to fully understand its impact.

Language Background

Many of the skills and abilities mentioned in the previous section are acquired throughout childhood and are largely informed by the environment of the individual as they are developing their language skills. A key environmental factor considers the languages spoken and taught during language acquisition. This may seem fairly straightforward if only one language is spoken, however this is not the case for the majority of the people in the world. It is estimated that over half of the population of the world speaks more than one language (Grosjean, 2010). When looking at the breakdown of speakers of English, it is estimated that 339 million people speak the language as an L1, and a staggering 1.6 billion people speak English as an L2 (Pecorari, 2018). In Canada it is estimated that approximately 7.7 million people speak English as their second or third language (Statistics Canada, 2017). This highlights the representation of a large number of different language backgrounds, which leads to important questions surrounding how exposure to multiple languages can impact language acquisition in general.

Learning to read in your L1 is a complex topic in and of itself, and it has been shown to rely upon a wide range of skills and abilities, as previously discussed. Models of reading in one's non-dominant language (L2) create a more complex picture even though some of the same skills are relied upon as when learning an L1. Understanding the differences between L1 and L2 learners, and how to approach teaching and learning English given a specific language background, is an important area of study. Reading ability in one's L2 is determined not only by their skills and abilities in this language, but also how well they read in their L1 (see Bernhardt, 2011). Furthermore, many accounts of L2 acquisition argue that the degree of similarity between one's L1 and their L2, known as L1-L2 distance, can impede or boost one's reading skill in English as well as proficiency in their component skills. Languages that have more similarities to English are predicted to facilitate easier transfer of L1 skills to L2 proficiency and are argued to provide an advantage in developing L2 reading proficiency and certain related component skills (Jeon & Yamashita, 2014; Yamashita & Shiotsu, 2017; Melby- Lervåg & Lervåg, 2014). For example, speakers of German may have an advantage over speakers of Chinese when learning English because, unlike Chinese speakers, German speakers benefit from a large overlap between vocabulary in their language compared to English (e.g., German "haus" means "house"), and in the similarity of their writing systems. Given the wide range of language backgrounds within Canada alone, it is critical to determine the role of this variable when discussing reading proficiency and more broadly, literacy.

Cognitive factors

In addition to direct components of reading, general cognitive abilities have also been found to impact reading ability. Factors such as memory and attention have been studied extensively regarding their influence on reading (Cain et al., 2004; Cutting & Scarborough 2006; Fedorenko, 2014; Schuette et al., 2008; Seigneuric & Ehrlich, 2005). While these factors are important when discussing an individual's reading proficiency, they are by no means the only factors involved. This thesis specifically looks at two cognitive factors - motivation and statistical learning.

Motivation

Motivation is a widely investigated and important concept when discussing learning and performance across a broad number of cognitive domains, including language-related skills. The concept of motivation is multifaceted in and of itself and has been broken down into numerous types and aspects of motivation. Many of which have demonstrated a strong association to reading. The link between motivation and language skills has been studied thoroughly with findings suggesting that higher motivation (both to learn in general and in acquiring a particular skill) is shown to correlate to more favorable outcomes in the skills and abilities being learned (Guthrie et al., 1999; Guthrie et al., 2006; Chapman & Tunmer, 1995; Pintrich, 2003; Pintrich et al., 1993; Wigfield et al., 2006). This relationship has been primarily demonstrated in L2 acquisition, but it has also been demonstrated in child language acquisition of the L1 (Chapman et al., 2000; Brown, 2000; Oroujlou & Vahedi, 2011). When it comes to reading research, the idea of motivation becomes harder to define, and is often thought to be multi-faceted. Several studies have investigated multiple aspects of reading motivation including intrinsic and extrinsic influences, an individual's attitude towards reading as a pastime, and even group differences such as gender (Guthrie & Coddington, 2009; Kush et al., 2005; Wigfield et al., 2016). While these are all beneficial to the field of language acquisition, this thesis is interested in the general motivation of an individual towards performing well regardless of their relationship to reading. This general task motivation is important, as it quantifies the individual's desire to do well, and may reflect the effort they apply to the tasks in the study (Van Iddekinge et al., 2018). I am primarily focused on this type of motivation in this thesis as it is important when discussing reading proficiency in the context of improving literacy. Given an individual's motivation to do well can vary widely, and accounting for this variance can help shed light on additional factors and interactions.

Statistical Learning

In recent years, there has been an increased focus on individual differences studies highlighting another general cognitive skill known as statistical learning (SL). SL refers to the brain's ability to notice and use re-occurring patterns in the environment to predict and process upcoming information. SL has been shown to influence a number of cognitive abilities across a wide variety of fields specifically within psychology, including implicit learning tasks, visual search, and contextual cuing (see, Frost et al., 2015). One area that has been a primary focus is its involvement within language acquisition and use (Erikson & Thiessen, 2015; Frost et al., 2019; Romberg & Saffran, 2010; Siegelman, 2020).

The interest in studying SL in relation to linguistic phenomena is growing, and the field is learning more and more about how SL is related to language. One approach is to think about how statistical probabilities are represented within a language. For example, at the orthographic level, certain co-occurrences of letters can be found with differing patterns. For example, the letters J and A are less likely to co-occur together in English words than the letters S and E (Chetail, 2007). Probabilistic patterns can also be seen in the mapping of letters to sounds (e.g., the likelihood of the letter combination 'gh' to sound like it does in the word 'rough' versus the word 'though'). Patterns of words that commonly co-occur can also be seen (e.g., the phrase '____ amok' can sensibly only be paired with the word 'run'). In morphology, an example of a non-adjacent dependence in the structure 'is __ing' is also evident. The word 'is' in this format predicts the ending '-ing' after a given verb. More complex linguistic fields such as syntax, prosody, and even gesturing also contain common patterns and regularities that can be useful in processing and learning language. Research has thus far demonstrated that humans have the capacity for extracting statistical regularities from the environment across modalities, different tasks, and in all age groups (Abla et al., 2008; Bulf et al., 2011; Conway et al., 2005; Kirkham et al., 2002; Neger et al., 2014; Rebuschat & Williams, 2012; Saffran et al., 1996). Given the many domains of statistical probabilities within language itself, it stands to reason that SL might play a role in language use, specifically in reading.

In recent years, there has been an increased focus on individual differences studies in statistical learning, some of which suggest the individual variation in SL performance predicts variability in some linguistic outcomes (see Misyak & Christiansen, 2012; Growns et al., 2020). Furthermore, SL has come into research focus in the recent years as a basis for successful acquisition of reading skills in both L1 and L2 (Arciuli & Simpson, 2012; Frost et al., 2013; Harm & Seidenberg, 2004). Much like other sub-domains of language, recent studies have started to examine whether there are correlations between SL and reading as a way to gain insights into their shared computations (Misyak et al., 2010; Singh et al., 2012; Siegelman, 2018; Spencer et al., 2015). However, additional research into the complex nature of the relationship between SL and reading is required.

Multi-dimensional Approach

Questions surrounding literacy and what factors impact an individual's reading proficiency have been thoroughly investigated over decades, and yet new variables are still being identified as key players that influence reading proficiency. With so many different skills and abilities shown to impact reading proficiency, there needs to be a concerted effort moving forward to disentangle how these aspects contribute. Furthermore, given the vast number and nature of these variables, spanning from general cognitive factors to very specific reading components, there is also a very real possibility that many of them overlap in their contribution. Combining the study of these above-mentioned skills, abilities, and cognitive factors can create a highly complex picture of reading proficiency that draws on numerous areas of expertise. Further complicating this is the presence of causal relations between the different variables found to be associated with reading proficiency. Further investigation into the nature of these inter-relationships, directionality of correlations between variables, and possible mediating factors underlying these relationships are worth inspecting.

This thesis approaches the complex topic of reading proficiency from a number of different perspectives to further understand the various factors that contribute and how they interact. The individual differences methodologies and approach are a necessary component of this thesis in identifying certain factors underlying the variability in reading proficiency seen among individuals. This will include investigating some core variables that are well-known in the literature as impacting reading proficiency, as well as introducing some newer components that are less well studied, to see how they fit into the broader picture. Additionally, investigating these questions among populations with different language backgrounds can further highlight the multitude of contributing factors in this area, and is a critical piece of this thesis.

Objectives

Given the complexity of reading proficiency and the expanding list of variables that are known to influence it, this thesis looks to approach this problem using a multi-dimensional approach to gain a more comprehensive picture of how these factors all contribute together to make up an individual's reading proficiency. This approach looks at the different components of proficiency (i.e., fluency and comprehension), studying both L1 and L2 speaking populations, using different methods of data collection, and different statistical methods to investigate important variables not only in isolation, but also the inter-relations among these variables. This can help to disentangle the contributions among variables and shed light on the consensus regarding the causal relationships between these correlated variables. By using this approach, this thesis demonstrates a novel look at the complex issue of individual variability in reading proficiency in order to help address gaps that exist in current research on this topic.

Objective One: How Do Different Measures of Proficiency Contribute to Understanding Individual Differences in Reading?

The first objective of this thesis was to employ a more holistic approach towards understanding reading proficiency by measuring both reading fluency and reading comprehension. Breaking down reading proficiency into these two different measures allows us to look at the bigger picture of how individuals differ in their reading ability. Reading is not only about being able to accurately comprehend the meaning underlying text, but also about how effortlessly one can go about doing this. That is, the fluency with which an individual reads also plays a role in overall reading ability and contributes a great deal to their literacy success in society. By measuring both comprehension ability and fluency in relation to measures of individual differences, this thesis looks to gain a better idea of the big picture of reading proficiency. Furthermore, inferences can be made as to how these different measures of reading proficiency work together to determine an individual's success in reading.

Objective Two: How Can Different Methods of Data Collection and Analysis be Used to Further Understand Individual Differences in Reading Proficiency?

The second objective of this thesis was to utilize a variety of data collection and data analysis methods and techniques to contribute to the bigger picture of reading proficiency. Using different methods of data collection can aid in creating a better idea of the issues surrounding individual variation when it comes to reading proficiency. In collecting data from multiple groups on a number of reading component skills tests as well as general cognitive factors, this thesis gives a unique view of the variety of factors contributing to reading proficiency. Through investigating different data collection methods of reading-related variables, this thesis also modifies and improves upon the data collection method for one commonly used behavioural measure of reading.

This thesis also sought to delve deeper into the data collected in order to gain new insight into how various factors impact reading proficiency. While traditional analysis tools have provided a great starting point, this thesis looked to probe further to better demonstrate the intricate complexities of how these variables are impacting the concept of interest. The primary goal of this objective was to examine the data in new and interesting ways to increase the understanding of how specific variables interact to impact reading proficiency.

Objective Three: Do Individual Differences Interact to Explain Variation in Reading Proficiency, and Does Language Background Play a Role?

The third objective of this thesis looked to answer whether there are any interactions among the different variables measured throughout the core studies. While there has been a lot of work investigating these factors individually, this thesis sought to answer whether there are overlaps between groups of variables known to influence reading proficiency. Specifically, it looked at whether these interactions spanned foundational component skills (spelling, vocabulary, exposure to print), language background, and cognitive factors (statistical learning and motivation). The primary goal of this objective was to answer whether any of the sampled variables interact or facilitate the influence of one another in impacting an individual's reading proficiency. This included investigating whether individual differences interact to influence reading proficiency in general, as well as investigating whether they interact differently based on group differences such as language background.

Overall Goal

Taken together, the overall goal of my thesis is to determine the critical components involved in becoming proficient in English, both in L1 readers of English, and in L2 learners of English. This thesis aimed to investigate how variation in component reading skills, general cognitive abilities, and the structure of one's L1 or L2, interact to facilitate or hinder reading fluency and comprehension in English. Literacy is an important global and societal issue, and it is therefore important to continue to take steps towards determining the factors underlying success in literacy skills such as reading. This thesis sets out to investigate what factors contribute to an individual's reading proficiency, how best to measure them, what statistical methods can be used to analyze the data, and whether these factors interact in their influence.

Approach

The first project in this endeavor, described in Chapter 2, was to create a more easily accessible measure of individual differences in word anticipation as reflected in cloze probability. This new application was designed to be easy to program, easy to run on any computer, and have easily manipulatable output for data analysis. The second project, described in Chapter 3 was designed to study reading in both L1 and L2 speakers of English, using eye-tracking as an additional behavioural measure. Furthermore, the analysis of this project led to the creation of a new statistical prediction method, which served as the core of the paper.

Seeing the differences between the L1 and L2 English groups, it was clear that additional languages should be studied to get a better picture of how linguistic distance impacted eye-movement behaviours during reading comprehension. This led to the undertaking of the Multilingual Eye-tracking COrpus (MECO) in which I was a collaborator. Working with MECO led to the development of another project combining the cloze probability application (from Chapter 2) with the reading proficiency measures utilized in the MECO corpus to see how linguistic distance factored into an individual's reading fluency and comprehension abilities and whether there are similarities in its influence on their cloze probability ratings. This project is not included in the current thesis due to the inability to test the required number of participants given lab shutdowns during the pandemic. This project will be further discussed in the 'Discussion and Conclusion'.

The final project of my thesis, described in Chapter 4, looks to expand the investigation beyond the individual differences seen in reading skills and abilities to examine additional cognitive factors, including SL and motivation. This project utilized mediation analysis to get a full picture of how some of these core variables interact with one another to influence reading comprehension. Further details as to how these projects seek to answer the main goals and objectives of the thesis will be presented in the next section.

Introduction to Chapters

The individual papers that constitute this thesis are related to one another on many levels. Broadly speaking, they all address the main questions that I set out to answer. Primarily, these were to investigate individual differences in certain skills and cognitive factors as they relate to reading ability. However, in approaching this goal, these papers take different perspectives, and employ different methods to answer this same overarching question. Additionally, there are certain aspects of each study that naturally link to one another to help form this broader picture. Below, I will introduce each of the main chapters by giving a brief overview of the chapter and how it relates to the overall objectives and theme of the thesis. Finally, I will also give a brief explanation as to how each of these projects relate to one another and any expected overlap.

Chapter 2 Overview

Chapter 2 is a paper titled 'Clozapp: A Java application for collecting and recording Cloze probability norms' published in the *Mental Lexicon* journal. Cloze predictability refers to an individuals' ability to predict upcoming words while processing text. The more predictable a word is in context generally correlates with the speed with which it takes to process it (e.g., more predictable words tend to be processed faster). This can be a useful measure when choosing stimuli for experiments, and as a way to investigate individual differences in the ability to predict and comprehend text passages. In this paper, a new freely accessible application for collecting cloze probability norms is shared, and it is demonstrated that this method of data collection is on par with the accuracy of other available methods. This paper primarily addresses the second objective of this thesis in that it is an updated version of a data collection method and examines its comparability with other similar methods. This paper focuses on an important reading skill and investigates how individuals differ in their ability to predict upcoming words in text using the measure of cloze probability. Given the need for a fast and easy way to implement testing of this ability, this paper was published to highlight a new publicly available cloze probability application complete with a user manual and modifiable script for researchers that are new to the protocol.

Chapter 3 Overview

Chapter 3 is a paper accepted for publication in *Studies in Second Language Acquisition* and is titled 'Quantifying the difference in reading fluency between L1 and L2 readers of English.' This paper looks at the effect of language background and reading component skills on reading speed in English. It highlights a potential L1 advantage or disadvantage for an alignment or lack thereof between an individual's L1 and L2. This chapter addresses all three objectives of this thesis by looking at an important measure of reading proficiency, using different data collection methods and a novel statistical analysis tool, all while investigating interactions among these variables.

This study employs a comprehensive measure of reading fluency by using eye-tracking methodology. Eye-tracking has been an important tool used to investigate reading proficiency and is demonstrated to be an effective and advanced methodology for studying individual variability in reading (Ashby et al., 2005; Chace et al., 2005; Rayner, 2009). This method of data collection allows an ongoing online view of reading behaviours through an individual's eye-movements. Important reading-related skills and abilities have been tied to critical eye-movement behaviours such as fixation duration, saccade length, frequency and percentage of skips, regressions among others (Jarodzka, & Brand-Gruwel, 2017; Rayner, 1998; Rayner, 2009). Reading comprehension is one aspect of reading proficiency that has been investigated using eye-tracking methodology (Kaakinen et al., 2002; Kaakinen, & Hyönä, 2005; Rayner et al., 2006). Reading fluency is also an important and well-studied area of focus regarding eye-tracking and reading (Brysbaert, 2019; Ashby & Clifton, 2005). This study focused on reading rate (as measured in words per minute) as well as the total fixation duration on each individual word to get a measure of each individual's reading fluency.

Testing different populations contributes to the second objective by looking at how individual difference variables interact to influence reading proficiency across different groups. Here the main factor is the participants' language background (their L1) when being tested in English (their L2). Given that reading is a multi-faceted skill and numerous studies have shown spelling, vocabulary, and exposure to print to be related to reading performance (see 'Background'), these reading-related variables were included in this study. A standard spelling test developed by Andrews & Hersch (2010) was used, vocabulary size was measured using the resource LexTALE (Lemhöfer & Broersma, 2012), and exposure to text in English was measured using the Author Recognition Test (ART) (Stanovich & West, 1989, 1993).

In addition, a novel statistical prediction method was created to investigate the upper bounds of a group that otherwise proves difficult to study. Using this statistical prediction method, hypothetical L2 speakers were created based off models generated using data collected in the study. These hypothetical speakers were given L1-like test scores on all of the component tests (spelling, vocabulary, and exposure to print) in order to investigate how individual differences impact reading fluency. This allowed a comparison of the actual L2 speakers, to these hypothetical L2 speakers to see how much individual variation plays a role in their fluency and how much is determined based on their language background.

Finally, this study contributes to the third objective of the thesis by investigating the interactions between the different variables of interest. Its primary focus was investigating how language background interacts with the individual differences seen in the reading-related skills of spelling, vocabulary, and exposure to print in predicting reading fluency.

Chapter 4 Overview

Chapter 4 is a paper titled 'The Validity of the Link Between Statistical Learning and Reading Comprehension: Testing for Mediating Variables.' submitted for publication to the journal *Cognitive Science*. This project investigated the link between SL efficiency and reading comprehension by investigating potential variables that could underlie or facilitate this relationship, including motivation. This project contributes to all three objectives of this thesis by looking at another measure of reading proficiency, utilizing numerous data collection methods and a more complex data analysis method, and again looks at the interactions among variables of interest when predicting reading.

This study looks at reading proficiency by way of reading comprehension ability. The well-known Grey Oral Reading Test (GORT) was utilized in this study as it has been shown to be a reliable and valid measure of reading comprehension (Wiederholt & Bryant, 2001). In this study, the more advanced passage numbers 5-8 were used, as these are most associated with the reading level expected of the population of interest (undergraduate university students).

One important question about the relationship between SL and reading that has yet to be answered is the role of an individual's motivation. This study addresses the second thesis objective by focusing on data collection efforts for both SL (using the Visual Statistical Learning (VSL) paradigm created by Siegelman et al., (2018), and task motivation (using the Student Opinion Scale, or SOS) developed by Thelk et al., (2009). Given the nature of the relationship between SL and reading reported thus far, it is also reasonable to suggest that SL could affect either some components of reading ability, or all components, but to differing degrees. Therefore, it is also important to look further into the relationships between specific reading components and SL as they relate to reading comprehension.

In addition to SL and motivation, data was collected on the same three reading-related skills as in Chapter 2 (spelling, vocabulary, and exposure to print) using the same measures. The question concerned how these variables uniquely and together contribute to the relationship between SL and reading. In order to look at the data and answer the main question of interest, the data analysis method of mediation analysis was employed. This approach is important as it can demonstrate different ways in which variables can interact or work together to have an influence on one another (Hayes & Scharkow, 2013; Tingley et al., 2014).

In contributing to the third objective of the thesis, this study uniquely paired two cognitive factors known to influence reading proficiency (SL efficiency and motivation), with three common reading-related skills (spelling, vocabulary and exposure to print). Specifically, the study examined whether an individual's performance on a visual SL task predicts their reading comprehension score. In addition, the study aimed to determine whether there are additional variables that might be contributing to the relationship between SL and reading. One way to investigate this complex relationship is to look at the role of SL in relation to other factors that determine an individual's reading proficiency. There is clear support of the role of SL in language. By including individual differences in known reading-related variables in a more complex regression model, the hope is to better understand how this relationship comes about on a more targeted language ability, i.e., reading proficiency.

There is a wide range of skills and abilities that have been linked to reading comprehension over the years and it is critical that the field investigates how SL fits into the broader picture. In this study, there was a focus on motivation, spelling ability, and exposure to text, which show large variation among individuals and are all known to influence reading comprehension. Through this comprehensive analysis, this study looked at important interactions that are otherwise difficult to ascertain.

Links and Overlap Between Chapters

The papers presented within the core chapters of this thesis all address the main objectives, while taking different approaches. An additional link can also be found in the theme of statistical regularities seen within all three chapters. The component skills chosen for both chapters 3 and 4 are strongly connected to statistical regularities underlying language and language acquisition. Reading can be considered to be a primarily visual activity, and spelling, vocabulary and exposure to print all rely heavily on visual regularities of language. Both spelling and vocabulary rely on the frequency with which words as stimuli are presented to the individual throughout their lives, and spelling is known to be influenced by how often people see the correct versus incorrect spellings (Rahmanian & Kuperman, 2019). Vocabulary size can be thought of as distributional information, as some words are simply more common than others given the reading history of the individual. Both spelling and vocabulary abilities can therefore be influenced by the amount and type of reading material an individual is exposed to throughout their lives. Furthermore, while not directly investigated, it is possible that statistical pattern recognition could play a role on a semantic level as well during the prediction of upcoming words, which could possibly be connected to variability in cloze probability responses. Another common element throughout the three main chapters of this thesis is the idea that reading proficiency is multifaceted and reflects the importance of investigating this concept from a number of different perspectives.

While each of the following chapters relies on different fields of inquiry for the literature reviews, different methodology, and different statistical analyses, they are all concerned with similar objectives and, therefore, have some content overlap in their motivation and underlying goals, and descriptions of reading-related skills. Specifically, there will be significant overlap when discussing the importance of acquiring functional literacy, measuring components of reading, as well as a discussion of outcomes that relate to implications for social programs and classroom teaching of English.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952-964. doi: 10.1162/jocn.2008.20058

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. doi: 10.3758/BRM.40.1.278

Afflerbach, P. (Ed.). (2015). *Handbook of individual differences in reading: Reader, text, and context*. Routledge. doi: 10.4324/9780203075562

Akbarian, I. H. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System, 38*(3), 391-401. doi: 10.1016/j.system.2010.06.013

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299. doi: 10.1037/a0018366

Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and writing*, *25*(6), 1283-1305. doi: 10.1007/s11145-011-9317-8

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Ashby, J., & Clifton Jr, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89-B100. doi: 10.1016/j.cognition.2004.12.006

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. The Quarterly Journal of Experimental Psychology Section A, 58(6), 1065-1086. doi: 10.1080/02724980443000476

Au, J. M. J. (2001). The effect of print access on reading frequency. *Reading Psychology*, *22*(3), 225-248. doi: 10.1080/027027101753170638

Baleghizadeh S, Golbin M. The Effect of Vocabulary Size on Reading Comprehension of Iranian EFL Learners. LiBRI. Linguistic and Literary Broad Research and Innovation 2010; Vol. 1, Issue 2, 33-46.

Berninger, V. W., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6. *Journal of psycholinguistic research*, *39*(2), 141-163. doi: 10.1007/s10936-009-9130-6

Bernhardt, E. (2010). *Understanding advanced second-language reading*. Routledge.

Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, *40*(1), 57-74. doi: 10.1111/1467-9817.12050

Brown, H. D. (2000). *Principles of language learning and teaching* (Vol. 4). New York: Longman.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. doi: 10.1016/j.jml.2019.104047

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132. doi: 10.1016/j.cognition.2011.06.010

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, *96*(1), 31. doi: 10.1037/0022-0663.96.1.31

Carlson, E., Jenkins, F., Li, T., & Brownell, M. (2013). The interactions of vocabulary, phonemic awareness, decoding, and reading comprehension. *The Journal of Educational Research*, *106*(2), 120-131. doi: 10.1080/00220671.2012.687791

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(3), 209. doi: 10.1037/h0087476

Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, *87*(1), 154. doi: 10.1037/0022-0663.87.1.154

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2000). Early reading-related skills and performance, reading self-concept, and the development of academic self-concept: A longitudinal study. *Journal of educational psychology*, *92*(4), 703. doi: 10.1037/0022-0663.92.4.703

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103-120. doi: 10.1016/j.cognition.2017.02.015

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24. doi: 10.1037/0278-7393.31.1.24

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific studies of reading*, *10*(3), 277-299. doi: 10.1207/s1532799xssr1003_5

Deacon, S. H., & Kieffer, M. (2018). Understanding how syntactic awareness contributes to reading comprehension: Evidence from mediation and longitudinal models. *Journal of Educational Psychology*, *110*(1), 72. doi: 10.1037/edu0000198

Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, *18*(6), 432-451. doi: 10.1080/10888438.2014.926907

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just" more phonological"? The roles of morphological and phonological awareness in reading development. *Applied psycholinguistics*, *25*(2), 223. doi: 10.1017.S0124716404001117

Desjardins, R., Murray, T. S., Rubenson, K., Werquin, P., Recotillet, I., Dong, L., Veenhor, B., Sciadas, G., Dhawan-Biswal, U., Willms, J. D., & Kirsch, I. (2005). *Learning a living: First results of the adult literacy and life skills survey*.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in*

language Disorders. doi: 10.1097/00011363-200020030-00005
Elgort, I., & Nation, P. (2010). Vocabulary learning in a second language: Familiar answers to new questions. In *Conceptualising 'learning'in applied linguistics* (pp. 89-104). Palgrave Macmillan, London. doi: 10.1057/9780230289772_6
Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H., Polyzoe, N., ... & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading research quarterly*, *39*(4), 438-468. doi: 10.1598/RRQ.39.4.5
Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108. doi: 10.1016/j.dr.2015.05.002
Falkauskas, K., & Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1607. doi: 10.1037/xlm0000132
Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in psychology*, *5*, 335. doi: 10.3389/fpsyg.2014.00335
Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. Trends in cognitive sciences, 19(3), 117-125. doi: 10.1016/j.tics.2014.12.010
Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128. doi: 10.1037/bul0000210
Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language?. *Psychological science*, *24*(7), 1243-1252. doi: 10.1177/0956797612472207
García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74-111. doi: 10.3102/0034654313499616
Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, *31*(1), 16-28. doi: 10.1177/0734282912451971
Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, *27*(9), 1703-1743. doi: 10.1007/s11145-014-9517-0
Grayson, J. P., Côté, J., Chen, L., Kenedy, R., & Roberts, S. (2019). Academic skill deficiencies in four Ontario universities. *York University*.
Green, D. A., & Riddell, W. C. (2001). *Literacy, numeracy and labour market outcomes in Canada* (pp. 89-552). Canada: Statistics Canada.
Grenier, S., Jones, S., Strucker, J., Murray, T. S., Gervais, G., & Brink, S. (2008). Learning literacy in Canada: Evidence from the international survey of reading skills. *Ottawa: Statistics Canada. Recuperado de http://www. statcan. gc. ca/pub/89-552-m/89-552-m2008019-eng. pdf*, *6*.

Grosjean, F. (2010). *Bilingual*. Harvard University Press, Cambridge, Massachusetts.

Guthrie, J. T., & Coddington, C. S. (2009). *Reading motivation.* In K. R. Wenzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (p. 503–525). Routledge/Taylor & Francis Group.

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *The Journal of Educational Research*, *99*(4), 232-246. doi: 10.3200/JOER.99.4.232-246

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific studies of reading*, *3*(3), 231-256. doi: 10.1207/s1532799xssr0303_3

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

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, *111*(3), 662. doi: 10.1037/0033-295X.111.3.662

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological science*, *24*(10), 1918-1927. doi: 10.1177/0956797613480187

International Adult Literacy and Skills Survey (2005). Retrieved from the Statistics Canada website: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=15034

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. Journal of Computer Assisted Learning, 33(3), 193-201. doi: 10.1111/jcal.12189

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160-212. doi: 10.1111/lang.12034

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on online text processing. *Discourse processes*, *33*(2), 159-173. doi: 10.1207/S15326950DP3302_03

Kaakinen, J. K., & Hyona, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse processes*, *40*(3), 239-257. doi: 10.1207/s15326950dp4003_4

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. doi: 10.1080/00313831.2011.554699

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42. doi: 10.1016/S0010-0277(02)00004-5

Kuperman, V., Bar-On, A., Bertram, R., Boshra, R., Deutsch, A., Kyröläinen, A. J., ... & Protopapas, A. (2021). Prevalence of spelling errors affects reading behavior across languages. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0001038

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. doi: 10.1080/13803610500110141

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. doi: 10.1007/s11145-009-9180-z

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*(2), 325-343. doi: 10.3758/s13428-011-0146-0

Martin-Chang, S., Ouellette, G., & Madden, M. (2014). Does poor spelling equate to slow reading? The relationship between reading, spelling, and orthographic quality. *Reading and Writing*, *27*(8), 1485-1505. doi: 10.1007/s11145-014-9502-7

McQuillan, J. (2006). The effects of print access and print exposure on English vocabulary acquisition of language minority students. *Reading*, *6*(1), 41-51.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological bulletin*, *140*(2), 409. doi: 10.1037/a0033890

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological bulletin*, *138*(2), 322. doi: 10.1037/a0026744

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331. doi: 10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in psychology*, *1*, 31. doi: 10.3389/fpsyg.2010.00031

Moats, L. C. (2005). How spelling supports reading. *American Educator*, *6*(12-22), 42.

Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia-Social and Behavioral Sciences*, *66*, 555-563. doi: 10.1016/j.sbspro.2012.11.300

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. doi: 10.1037/a0021890

Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.

Neger, T. M., Rietveld, T., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in human neuroscience*, *8*, 628. doi: 10.3389/fnhum.2014.00628

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes*, *18*(4), 443-468. doi: 10.1080/01690960344000008

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes*, *18*(4), 443-468. doi: 10.1080/01690960344000008

OECD and Statistics Canada (2005), Learning a living: First results of the Adult Literacy and Life Skills Survey. Paris and Ottawa: Authors.

OECD Programme for the International Assessment of Adult Competencies (PIAAC), 2013

Ouellette, G., & Sénéchal, M. (2017). Invented spelling in kindergarten as a predictor of reading and spelling in Grade 1: A new pathway to literacy, or just the same road, less known?. *Developmental psychology*, *53*(1), 77. doi: 10.1037/dev0000179

Ouellette, G., Martin-Chang, S., & Rossi, M. (2017). Learning from our mistakes: Improvements in spelling lead to gains in reading speed. *Scientific Studies of reading*, *21*(4), 350-357. doi: 10.1080/10888438.2017.1306064

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. doi: 10.1080/10790195.2006.10850187

Perfetti, C. A. (1992). *The representation problem in reading acquisition.* In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (p. 145–174). Lawrence Erlbaum Associates, Inc.

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.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). *The Acquisition of Reading Comprehension Skill.* In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology*. *The science of reading: A handbook* (p. 227–247). Blackwell Publishing. doi: 10.1002/9780470757642.ch13

Pecorari, D. (2018). Teaching English to Speakers of Other Languaes (TESOL). In P. Seargeant, A. Hewings, & S. Pihlaja (Eds.), *The Routledge Handbook of English Language Studies* (pp. 199-211). London: Routledge.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of educational Psychology*, *95*(4), 667. doi: 10.1037/0022-0663.95.4.667

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational research*, *63*(2), 167-199. doi: 10.3102/00346543063002167

Plaza, M., & Cohen, H. (2003). The interaction between phonological processing, syntactic awareness, and naming speed in the reading and spelling performance of first-grade children. *Brain and cognition*, *53*(2), 287-292. doi: 10.1016/S0278-2626(03)00128-3

Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C., & Mouzaki, A. (2013). What do spelling errors tell us? Classification and analysis of errors made by Greek schoolchildren with and without dyslexia. *Reading and Writing*, *26*(5), 615-646. doi: 10.1007/s11145-012-9378-3

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language learning*, *52*(3), 513-536. doi: 10.1111/1467-9922.00193

Rahmanian, S., & Kuperman, V. (2019). Spelling errors impede recognition of correctly spelled word forms. *Scientific Studies of Reading*, *23*(1), 24-36. doi: 10.1080/10888438.2017.1359274

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372. doi: 10.1037/0033-2909.124.3.372
Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*, *62*(8), 1457-1506. doi: 10.1080/17470210902816461
Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, *10*(3), 241-255. doi: 10.1207/s1532799xssr1003_3
Rebuschat, P., & Williams, J. (Eds.). (2012). *Statistical learning and language acquisition* (Vol. 1). Berlin, Germany: De Gruyter Mouton.
Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914. doi: 10.1002/wcs.78
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. doi: 10.1126/science.274.5294.1926
Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26-43. doi: 10.1111/j.1540-4781.2011.01146.x
Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151-218. doi: 10.1016/0010-0277(94)00645-2
Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and writing*, *18*(7), 617-656. doi: 10.1007/s11145-005-2038-0
Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities?. *Cognitive science*, *42*, 692-727. 10.1111/cogs.12556
Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*(3), e12365. doi: 10.1111/lnc3.12365
Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental science*, *15*(4), 482-495. doi: 10.1111/j.1467-7687.2012.01141.x
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319. doi: 10.1016/j.cognition.2013.02.013
Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and writing*, *28*(4), 467-490. doi: 10.1007/s11145-014-9533-0
Stahl, S. A. (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in language disorders*, *23*(3), 241-247.
Stanovich, K. (1986). 1986: Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly 21, 360–406.
Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading research quarterly*, 402-433. doi: 10.2307/747605
Statistics Canada, 2017. Linguistic Diversity and Bilingualism in Canadian Homes. Retrieved from the Statistics Canada website: https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016010/98-200-x2016010-eng.cfm

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, 35-50. doi: 10.2307/747815

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1-17. doi: 10.1016/j.jml.2015.02.004

Taylor, I., & Olson, D. R. (Eds.). (1995). *Scripts and literacy: Reading and learning to read alphabets, syllabaries, and characters* (Vol. 7). Springer Science & Business Media.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Examining inferences about test-taking motivation: The Student Opinion Scale (SOS). *Journal of General Education*.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of educational Psychology*, *77*(4), 417. doi: 10.1037/0022-0663.77.4.417

UNESCO (2015). Education for all 2000–2015: Achievements and challenges. Education for All GlobalMonitoring Report 2015. Paris: UNESCO.

UNESCO Institute for Statistics. (2018, October). *Defining literacy*. GAML Fifth Meeting, Hamburg, Germany. Retrieved from the United Nations Educational, Scientific, and Cultural Organization website: http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/12/4.6.1_07_4.6-defining-literacy.pdf

Veldre, A., Drieghe, D., & Andrews, S. (2017). Spelling ability selectively predicts the magnitude of disruption in unspaced text reading. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1612. doi: 10.1037/xhp0000425

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*, *44*(1), 249-279. doi: 10.1177/0149206317702220

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific studies of reading*, *11*(1), 3-32.

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

Wiederholt, J. L., & Bryant, B. R. (2001). GORT 4 Gray Oral Reading Tests Examiner's Manual. *Austin, TX: Pro-Ed*.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2007). Development of achievement motivation. *Handbook of child psychology*, *3*. doi: 10.1002/9780470147658.chpsy0315

Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child development perspectives*, *10*(3), 190-195. doi: 10.1111/cdep.12184

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1186. doi: 10.1037/xhp0000213

Wood, C., Wade-Woolley, L., & Holliman, A. J. (2009). Phonological awareness. *Contemporary perspectives on reading and spelling*, 7-23.

Yamashita, J., & Shiotsu, T. (2017). Comprehension and knowledge components that predict L2 reading: A latent-trait approach. *Applied linguistics*, *38*(1), 43-67. doi: 10.1093/applin/amu079

Zhao, J., Joshi, R. M., Dixon, L. Q., & Chen, S. (2017). Contribution of phonological, morphological and orthographic awareness to English word spelling: A comparison of EL1 and EFL models. *Contemporary Educational Psychology*, *49*, 185-194. doi: 10.1016/j.cedpsych.2017.01.007

CHAPTER 2.

Clozapp: A Java Application for Collecting and Recording Cloze Probability Norms

Nisbet, K., Généreux, M., Anderson, B., & Kuperman, V. (2019). Clozapp: A Java application for collecting and recording Cloze probability norms. *The Mental Lexicon*, *14*(3), 399-414. doi: 10.1075/ml.20004.nis

Abstract

This paper introduces a freely available and easy to use Java application for the collection and recording of Cloze probability ratings. Clozapp presents participants with text fragments of the researchers' choice and collects guesses regarding upcoming words. It can also collect basic demographic information about participants. Available modes of data collection include elicitation of responses to a limited number of omitted words in a text or to all words in a text. Clozeapp can be customized to present instructions and experimental stimuli in any given language and to collect multiple types of demographic data. This paper presents the application by detailing the states and actions available, as well as descriptions of how to customize the app to fit different experimental needs including possible input and output details. The application manual is provided. As a proof of concept, we used Clozapp to conduct a replication study of two existing collections of Cloze probability norms. The Clozapp norms showed strong reliable correlations ($r > 0.7$) with both existing data sets, suggesting a high convergence between modes of data collection. The application provides an efficient and customizable way of collecting predictability norms for language research.

Background

The effects of contextual constraint on the prediction and processing of upcoming text have been an important topic of study in the past few decades, specifically for the field of reading. Our ability to predict upcoming information is a critical skill that facilitates seamless communication and quick processing of written text. We may know for certain the upcoming word in the following saying that someone is "*fit as a* ______", however we may only guess the ending to "*they ran* ____." Many different factors are at play when looking at whether a word is highly predictable or not, however the measurement of this factor has proven to be quite useful in the design and implementation of psycholinguistic experiments. We first provide a brief description of the task commonly used for measuring contextual predictability experimentally, i.e., the Cloze completion task (Taylor, 1956; also see: Bickley, Ellington, & Bickley, 1970; Jonz, 1976; Klare, Sinaiko, & Stolurow, 1972). Then we briefly review major findings linking Cloze probability norms to reading behavior. Finally, we motivate the present study.

A common way to estimate how predictable a word is in its context is to ask participants to guess the upcoming word based on available context. These estimates are typically acquired via a Cloze completion task. In a typical implementation of this task, text is presented to a participant a word at a time. For example, a participant would start with just receiving the first word of a text (e.g., *'The ____'*) and make a guess as to what will come next. After the participant has made a guess (either correct or incorrect), the actual word is presented, and the participant then guesses the next word (e.g., *'The boy ____'*). This continues word by word until the end of the sentence (e.g., *'The boy went outside to play with his toys*'). Another common use of the Cloze task is to delete only one word, normally at the end of a sentence, and collect guesses about this single word. The outcome of the Cloze completion task is the probability estimate for each completed word, calculated as a percent of participants who guess the word correctly based on prior context. We now know that words can range over a few orders of magnitude in their predictability, from highly constrained or predictable (for example, we know that the text "*We wish you a merry*" is likely followed by "*Christmas*") to virtually unconstrained: this variability has a demonstrable effect on word processing time and effort (see Smith & Levy, 2013; Staub, Grant, Astheimer, & Cohen, 2015).

Research has shown that more predictable words elicit shorter fixation times, are skipped more often and receive fewer regressions that less predictable words (Altarriba, Kroll, Sholl, & Rayner 1996; Balota, Pollatsek, & Rayner 1985; Ehrlich & Rayner, 1981; Inhoff, 1984; Schustack, Ehrlich, & Rayner, 1987; Staub, 2015; Zola, 1984). Rayner and Well (1996) and much further research has demonstrated that the effect of predictability on the reading effort is gradient rather than categorical (e.g., a greater degree of predictability leads to shorter reading times and higher accuracy). Smith and Levy (2013) have further reported that the functional form of the effect is logarithmic. These critical findings over the years have led to two outcomes. First, researchers routinely include predictability estimates in studies of word reading in context as a "benchmark" predictor of reading behavior, along with such factors as word frequency of occurrence or length (Rayner, 1998). Second, the finding that the functional relationship between probability and reading time is logarithmic (Smith & Levy, 2013) suggests that even subtle differences in Cloze probability norms (a word can be predicted by 1% of participants versus 0.1%) translate into substantial differences in response times. Estimation at this scale requires collection of Cloze norms from a large number of participants. Thus, acquisition of predictability judgments is and will continue to be an important logistic factor to account for when designing experiments, gathering stimuli, and analyzing studies. Therefore, having a simple, efficient and transferrable tool to collect Cloze norms for texts of any length is beneficial.

The data-collection application described in this paper grew out of this experimental need. To our knowledge, only one application offers desired capabilities for collecting probability norms, the Java-based Cloze Test Project (Jones & McRoy, 2003). As the application is not currently available either publicly or upon request from the team, we set out to build our own application with the desired functionality, Clozapp.

Clozapp is a Java class and is open source so that other researchers with access to computer programming resources may improve its capabilities to fit their experimental needs. Java has many advantages as a programming framework. It is an object-oriented language with a safe and common syntax, which makes it extensible and platform-independent. The Java Runtime Environment (JRE) necessary for executing Java applications can be downloaded for free and is easily installed. Clozapp has been developed with JRE version 8 and runs as a standalone program, without need for internet access. This application is compatible with versions of Java from JRE 8 through 13.

Clozapp Overview

All materials related to the Clozapp application – including the user manual (Clozapp User Manual), the Java applications, and a template for experimentation – are available for download on Open Science Framework (OSF), see Availability below. This section describes the individual parts of the manual in order to explain how to interact with Clozapp and use all of its features. Additional details and examples can be found in the Clozapp User Manual itself as well as in the Experiment Template.

Clozapp collects word guesses based on text fragments up to and excluding the word to be guessed. For example, when presented with "*As a boy, though, I had often ...*", a user's task is to guess what word would follow "often" in this context. Once a guess is made, Clozapp records the original word, the word guessed by the user, and the time it took to make this guess. The user is then presented with an updated context including the correct word: "*As a boy, though, I had often stayed ...*". Clozapp then carries on until it reaches the end of the text or a pre-defined reviewing point (see definition below). Responses collected by Clozapp are time-stamped. Clozapp collects both the typed input from the user (e.g., demographic information and word guesses) and information about mouse clicks on functional keys: loading, saving, (un)pausing and quitting. Once the user instructs Clozapp to save the current state, all information is saved as an output in in a text file.

Starting the Application

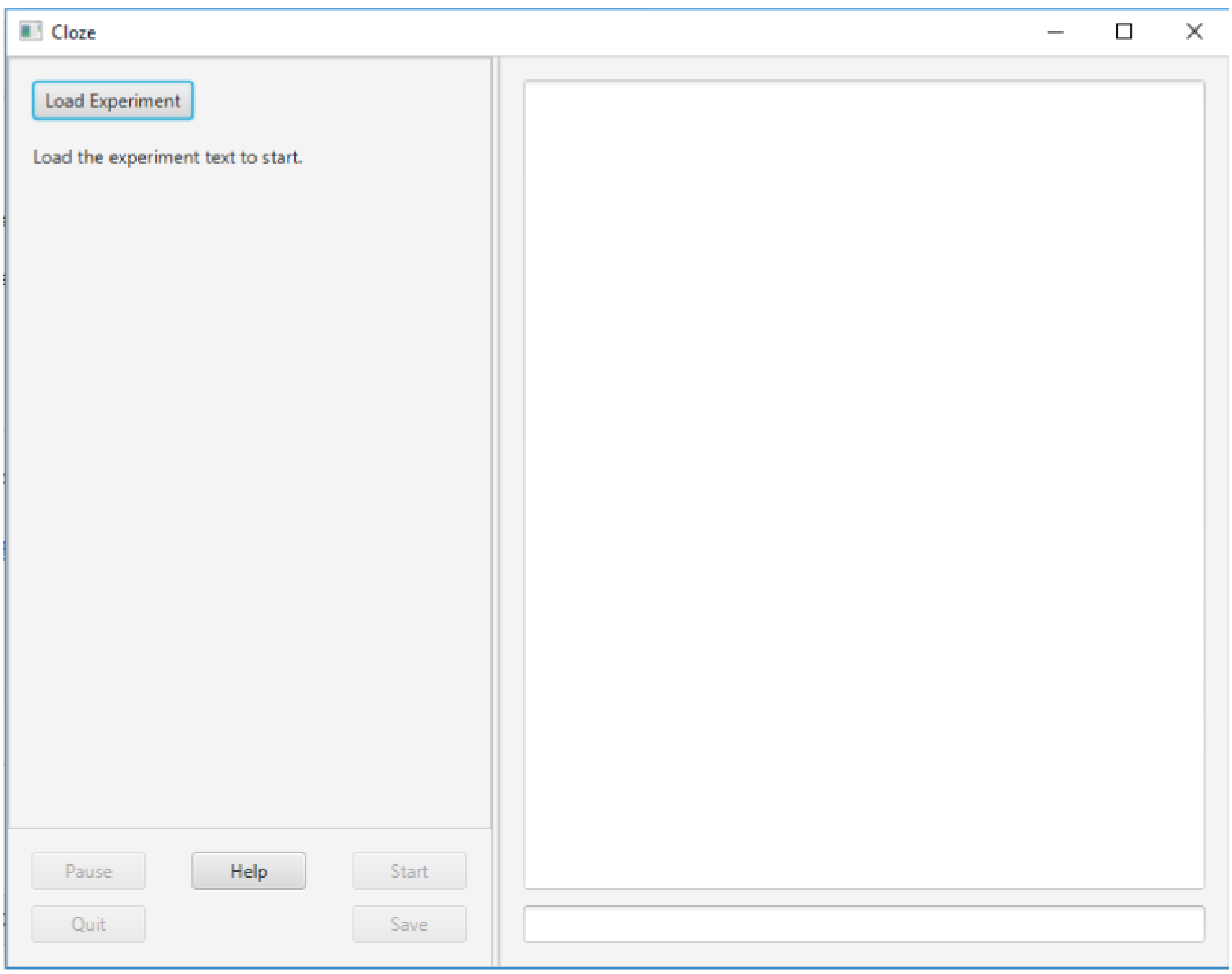
Clozapp can be run by anyone with JRE 8+ installed. As JRE 8 is the default version available for windows, no additional steps generally need to be taken, and the program 'Clozapp.jar' can simply be run. If you are unsure what version of Java you have installed, searching for 'Java' on your operating system and clicking 'About Java' which will display the version you are currently running. For those on JRE 9+, the JavaFX libraries need to be installed separately (See https://openjfx.io). A separate version of the program has been provided for anyone running JRE 13 ('Clozapp_JRE13.jar'), which bundles all the JavaFX libraries, so no additional installation is required.

The application can be started using one of the two following methods:

1. If your operating system supports it, you should be able to start the application by double-clicking on the "Clozapp.jar" file.
2. Otherwise, open a command prompt. Go to your "Clozapp.jar" file directory and type *java -jar Clozapp.jar*.

The initial look of Clozapp is shown in Figure 1, the only available options being the "Help" button and the "Load Experiment" button. The "Help" function can be used at any point to provide a quick snapshot of the instructions as a reminder of the task. This screen can be modified by editing the plain text file containing all the stimuli trials and other modifications (see an example experiment template). As a researcher you can link out to an external web page or resource if greater details need to be provided for the participant. Upon clicking the "Load Experiment" button, you can select and upload an experiment markup file (see detailed description below, or the Experiment Template), and the remainder of the user interface will be populated.

Figure 1. Clozapp upon opening the Java file.



Experiment Markup Files

All information that will be presented to the user is defined in a plain text (.txt) file loaded at the beginning of the experiment. The file is coded in a very simple customized markup language, similar to html, where tags indicate different modes of text presentation. We will refer to this file as the experiment markup file. Note that all information (instructions and stimuli) can be supplied in any language.

At the bare minimum, this file needs to contain the *Stimuli* that will be presented to the user. *Stimuli* define the sentences or other texts that are presented to the participants in the experiment to collect Cloze probability ratings, as well as the target parts of those sentences or texts for which responses are elicited. For example, "*we wish you a merry Christmas*" is a sentence-long stimulus and "*Christmas*" can be defined as a target part in that sentence, for which a participant's guess is prompted. It is possible to program each stimulus such that it prompts the participant to respond to every word in this stimulus, or only to a part of that sentence/text stimulus. Also, responses can be elicited on the word-by-word basis or using word sequences as units to be guessed.

These options are implemented through the use of syntactic tags '<blank>', '</blank>', '<end>', and '<review>', explained below and in the Clozapp User Manual uploaded on OSF.

Stimulus presentation: No tag is required in the beginning of a stimulus. But each stimulus needs to end with either an <end> or <review> tag. The <end> tag will make the application proceed to the next stimulus immediately upon completion of the last response to the target part of the stimulus. Conversely, the <review> tag will allow the user to review the completed stimulus before moving to the next stimulus: this functionality is especially useful when stimuli are texts consisting of multiple sentences. See examples of each tag in the Clozapp User Manual and Experiment Template.

Defining the target part of the stimulus for responding: We chose some typical experimental designs to highlight some of the structures that can be programmed using our tags. In all designs, the <blank> tag before the target part of a stimulus marks the beginning of the part that will be presented to the reader for responding.

In many Cloze probability tasks, researchers choose to have the sentence (or larger passage) as a whole guessed word-by-word by the participant. This can be programmed by inserting a <blank> tag at the start of each stimulus sentence. For example, the researcher would program as follows: "<blank> *The girl walked her dog in the park* <end>". This stimulus does not contain a </blank> tag and thus participants will be prompted to respond to every word between the <blank> tag and the end of the stimulus marked by the <end> tag. Using this formatting, the experiment would ask the participant to input a guess for each of the 8 words in the sentence one at a time.

Alternatively, an experimenter may choose to provide most of the context and only ask participants to fill in or guess a particular word or two. For example, the researcher would program a stimulus as follows: "*The girl walked* <blank> *her dog* </blank> *in the park* <end>". This would result in the first three words 'the girl walked' followed by the experiment asking for input from the participant to guess the words '*her*' and '*dog*' one at a time, and then would provide the remainder of the sentence '*in the park*' to conclude the trial.

Furthermore, an experimenter may choose to ask for multiple word guesses at the same time by using the syntax '<blank multiple>'. For example, if the researcher programmed as follows: "*The girl walked* <blank> *her dog* </blank> <blank multiple> *in the park* <end>. This would result in a similar beginning as the last example where the first three words '*the girl* *walked*' were presented at the beginning, followed by the participant guessing one by one the words '*her*' and '*dog*'. However, for the ending of this sentence, three blanks will be presented to the participant all at once for them to fill in '*in the park*' after which the trial would conclude. There are many possible options to choose from leading to a wide variety of experimental paradigms and questions that can be answered.

Instructions and Demographics

Instructions are defined in the experiment markup file by including the tag <instruction> at the beginning of the instruction text and </instruction> at its end. Furthermore, researchers are able to schedule breaks for participants using the same syntax. For example, if you have a particularly long session, at the half way mark, you can use the instruction tags to stop the experiment and simply insert text that says 'break' instead of additional instructions to give the participant some time before starting the second half.

There is a default set of demographic questions that will be presented to participants, which includes the participant's age, gender, and highest level of education. However, tags are available to override those defaults and customize the demographic questions to be shown to the user, if needed. Custom demographic questions are contained in a single section at the beginning of the experiment markup file that ends with the tag </demographics>. Each question has its own tag depending on the type of question desired. There are options for different input types including text, integers, floating point (decimal) numbers, radio buttons, or combo boxes (drop-down menus). Within any <question> tag, you must also include the text that will be presented to the user: for example, 'please state your current age' or 'select from the drop-down menu your highest achieved level of education'. In the case of radio buttons or combo boxes, each option must fall within <choice> tags. The name of the field is what appears in the output files, as well as the user prompts to fill uncompleted fields. Example code for the questions described above are included below:

```
<question integer ExampleNumber>Example Integer input:</question>

<question radio ExampleRadio>Example Radio Buttons:
        <choice>One</choice>
        <choice>Two</choice>
        <choice>Three</choice>
```

User functionality

Once the experiment markup file is loaded, the user interface will be populated with the desired questionnaire information, see Figure 2. This will display all of the demographic questions contained in your experiment markup file: either the default questions or the customized questions. The participants enter their demographic information, and upon completion they click the 'Start' button to begin. After clicking Start, the instructions are provided as set up in the experiment markup file. If no instructions are defined, the experiment will begin directly. Figure 3 provides an example snapshot of the application with instructions showing up before the stimuli are presented.

Figure 2. After loading the experiment markup file, participants will be prompted to fill in the demographic information requested by the experimenter. Pressing 'Start' begins the experiment.

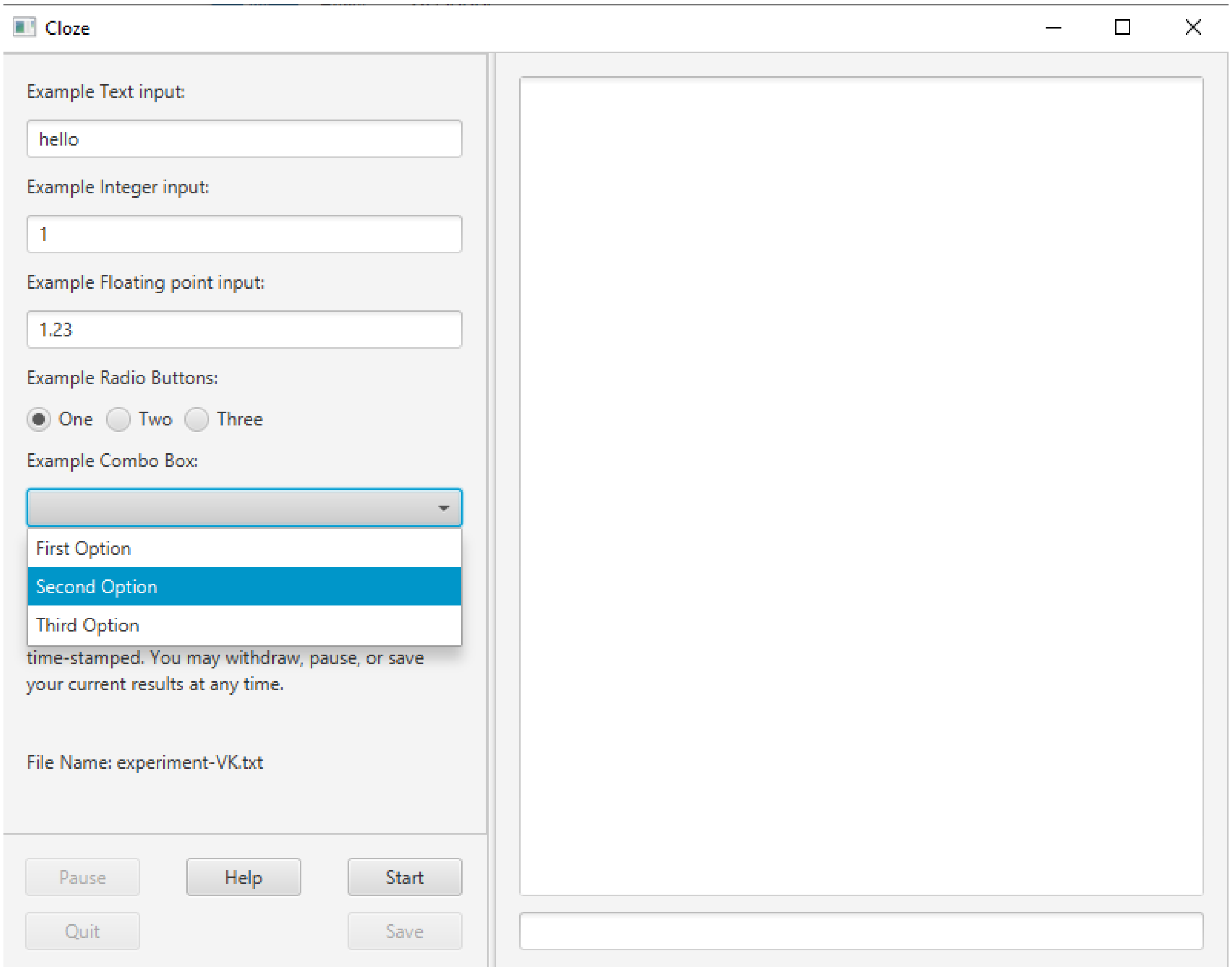
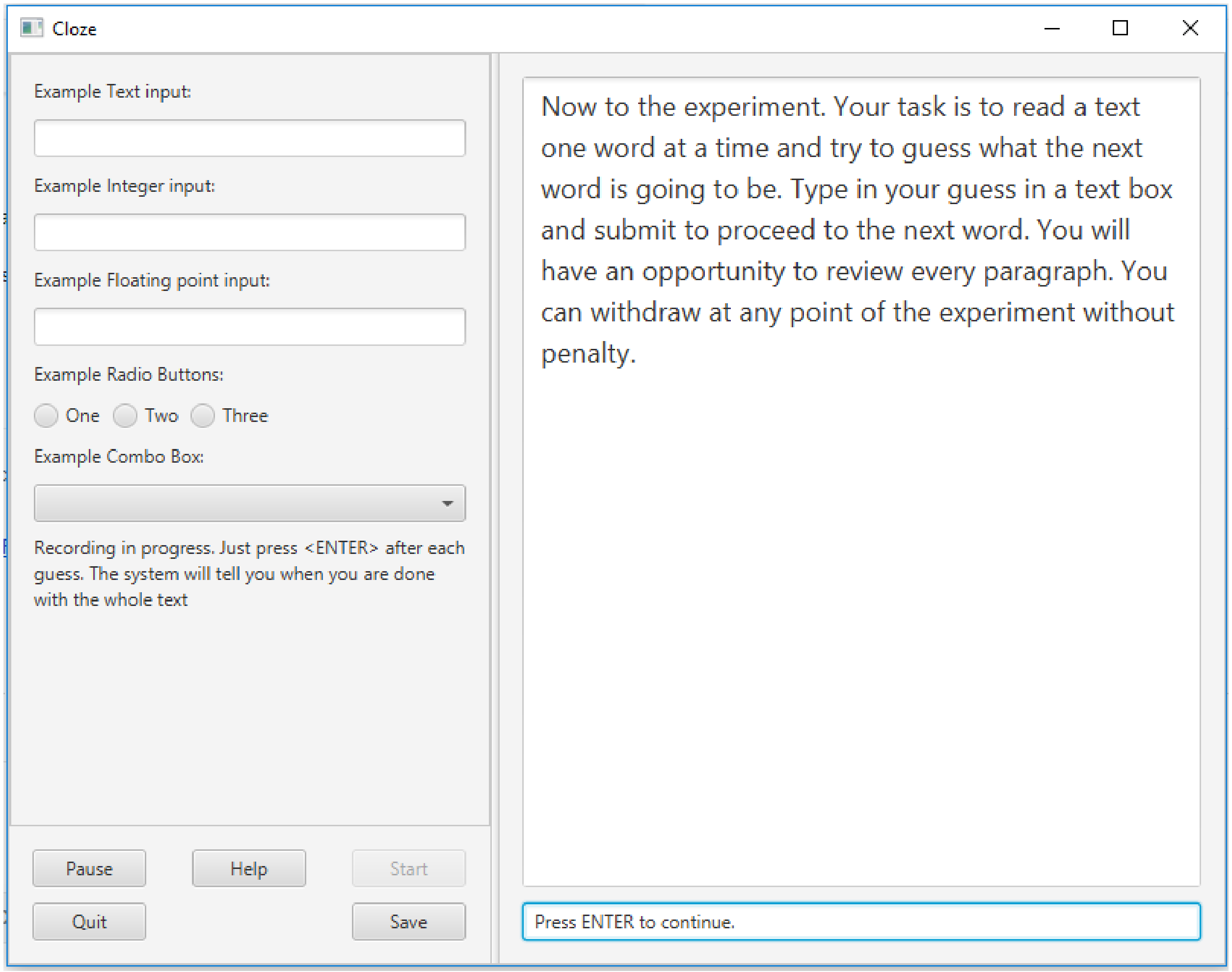


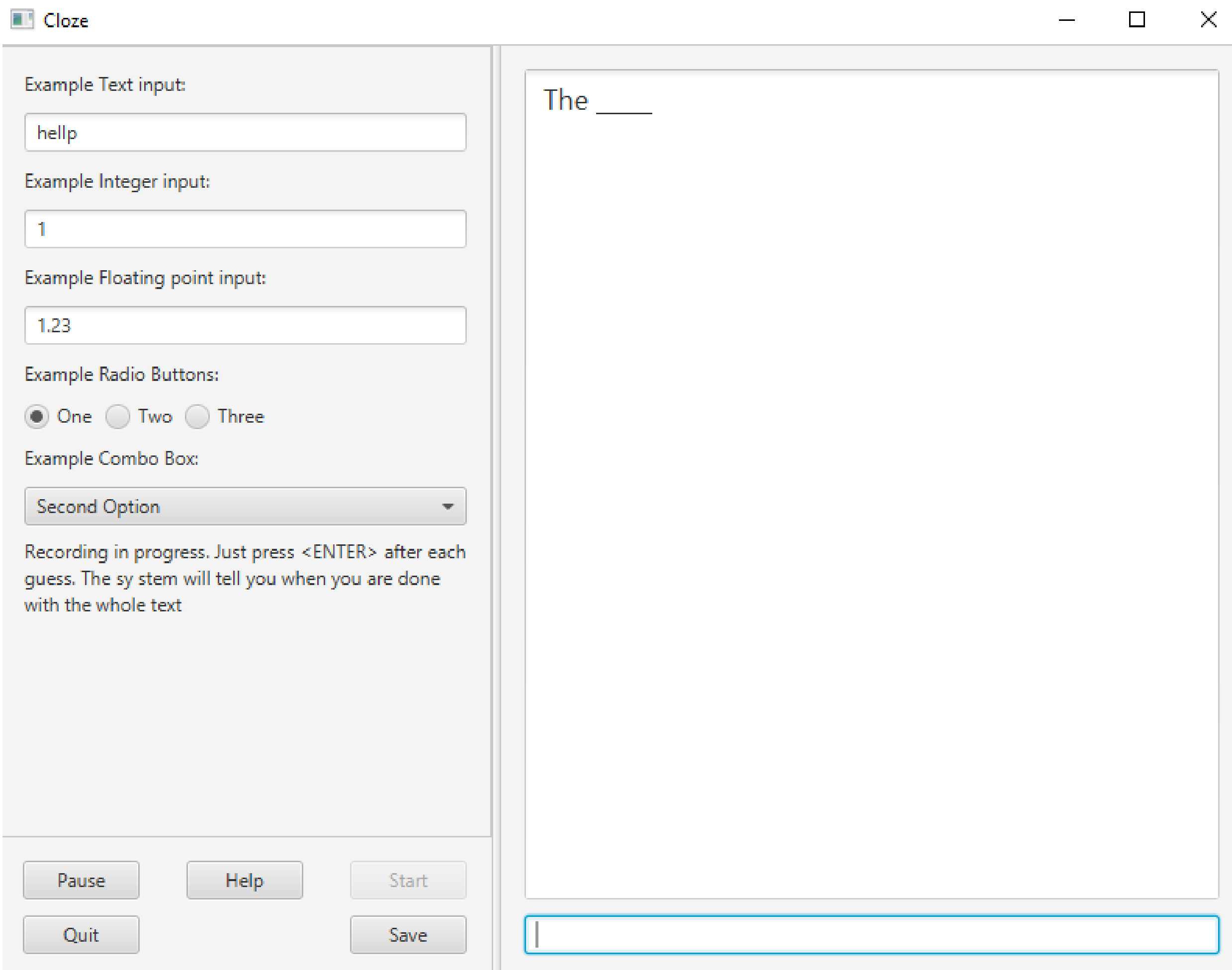
Figure 3. After pressing 'Start', the participants are given the instructions for the experiment as well as the 'in progress' text reminding them of the task. Buttons to 'Pause', ask for 'Help', or to 'Quit' or 'Save' the experiment are also available.



Once the participants are ready to begin the experiment, the screen will look like Figure 4. Blanks (i.e., words to be filled in) will be denoted by underscores, and the user will be prompted to enter their guesses into the box below. Answers are submitted by pressing the 'ENTER' key. At any point while answering questions, the participant may pause the experiment, save their results, or quit. Progress text is provided for each stage of the experiment and is always included to the left margin in the application. These texts outline each stage of the experiment and are always visible depending on the progress of the experiment. For example, during the experiment it reads "Recording in progress. Just press <ENTER> after each guess. The system will tell you when you are done with the whole text.".

<experiment> This text is always visible before the user clicks start </experiment> (see Figure 2)
<progress> This text is visible while the user is answering questions </progress> (see Figure 3)
<complete> This text is visible after the user answers the last question </complete>

Figure 4. A blank '____' indicates the word(s) to guess. Users type in their response and press <ENTER>.



Saving / Output file

A user's session is recorded and saved in the output file using the *save* button. The created file can be saved in one of two formats. Figure 5 shows an example of the default method for saving a recorded session.

Figure 5. The default format of recording a session.

```
ExampleNumber    1
ExampleDecimal   1.23
ExampleRadio     One
ExampleText      hello
ExampleDropdown  Second Option
Current File     experiment_KN_Nov20.txt
START Fri Nov 22 14:01:55 EST 2019
==================================
  1-1                          [A]   11.27   Fri Nov 22 14:02:12 EST 2019
  1-2         The     [researcher]   15.34   Fri Nov 22 14:02:24 EST 2019
  1-3        girl           [went]    2.42   Fri Nov 22 14:02:39 EST 2019
  1-4      walked            [her]    2.37   Fri Nov 22 14:02:41 EST 2019
  1-5         her            [dog]    1.32   Fri Nov 22 14:02:44 EST 2019
  1-6         dog             [at]    2.09   Fri Nov 22 14:02:45 EST 2019
  1-7          in            [the]    1.36   Fri Nov 22 14:02:47 EST 2019
  1-8         the           [park]    3.26   Fri Nov 22 14:02:49 EST 2019
  2-1      walked            [her]    3.46   Fri Nov 22 14:02:53 EST 2019
  2-2         her            [dog]    1.26   Fri Nov 22 14:02:56 EST 2019
  3-1         The            [man]    5.39   Fri Nov 22 14:03:03 EST 2019
  3-2        girl            [had]    6.46   Fri Nov 22 14:03:08 EST 2019
  3-3      walked            [her]    2.72   Fri Nov 22 14:03:15 EST 2019
  3-4         her            [dog]    5.27   Fri Nov 22 14:03:17 EST 2019
  3-5         dog          [after]    8.19   Fri Nov 22 14:03:23 EST 2019
  3-6          in            [the]    1.48   Fri Nov 22 14:03:31 EST 2019
  3-7         the           [park]    9.11   Fri Nov 22 14:03:32 EST 2019
  3-8       park.            [The]    3.60   Fri Nov 22 14:03:41 EST 2019
  3-9         She            [had]    2.33   Fri Nov 22 14:03:45 EST 2019
 3-10       walks            [her]    1.83   Fri Nov 22 14:03:47 EST 2019
 3-11         her            [dog]    1.10   Fri Nov 22 14:03:49 EST 2019
 3-12         dog             [at]    4.45   Fri Nov 22 14:03:50 EST 2019
 3-13       every            [day]    1.97   Fri Nov 22 14:03:55 EST 2019
 3-14     morning            [and]    3.64   Fri Nov 22 14:03:57 EST 2019
 3-15          at            [the]    1.45   Fri Nov 22 14:04:00 EST 2019
 3-16         the           [park]    3.23   Fri Nov 22 14:04:02 EST 2019
 3-17        same           [park]    2.49   Fri Nov 22 14:04:05 EST 2019
  4-1         dog    [at the park]    3.81   Fri Nov 22 14:04:11 EST 2019
  5-1         dog             [in]    2.95   Fri Nov 22 14:04:15 EST 2019
  5-2          in            [the]    2.85   Fri Nov 22 14:04:18 EST 2019
  5-3         the           [park]    1.25   Fri Nov 22 14:04:21 EST 2019
  6-1         #13            [ran]    5.33   Fri Nov 22 14:04:32 EST 2019
  6-2         won            [the]    1.35   Fri Nov 22 14:04:37 EST 2019
  6-3       first          [place]    2.54   Fri Nov 22 14:04:38 EST 2019
==================================
SAVE  Fri Nov 22 14:04:48 EST 2019
```

The first lines include all the demographic information as answered by the participant. The next line specified the name of the experiment markup text loaded, followed by the time and date the experiment was started at.

Following the break, all of the participant responses are recorded. The columns represent (in order):

1) The sentence/question number, and the word number within the question (as calculated from the markup file)
2) The last word of context available to the participant
3) The response by the user
4) The time elapsed (in seconds) between when the context was presented, and when a response was submitted. The elapsed time does not include time spent paused, though this can be deduced from the time stamps
5) The time at which the context was presented to the user.

Alternatively, data can be saved in JSON format, which may be easier to load into third party programs. The JSON export format can be seen in the Clozapp User Manual.

Experiment

In order to test the effectiveness of Clozapp, we designed an experiment to collect Cloze probability norms using the application. We compare norms collected via Clozapp to probability norms to the same stimuli collected using different methods in two previous studies: the Schilling corpus of Cloze probability norms (Reichle et al., 1998), and the norms from Schnoebelen and Kuperman's (2010) study.

Methods

Participants

The experiment collected data from 62 participants (5 male; age: 17 to 31 y.o.), and after removing participants who failed to complete the test, the final data set contained 59 participants. Participants were recruited from a convenience pool of undergraduate students. All participants were native speakers of English and were compensated by course credit.

Materials and Design

This experiment consists of a set of 48 sentences (531 words total) such as "*Margie moved into her new apartment over the summer*" from the Schilling corpus (Reichle et al., 1998), which were presented to participants word-by-word for completion. The same sentences were used in the Cloze task by Reichle et al. (1998) in a laboratory study and by Schnoebelen and Kuperman (2010) in a web-based study using the online crowdsourcing platform Amazon's Mechanical Turk. We wanted to see how norms collected for all words in all sentences compared across the three data sets.

Procedure

Our experiment follows a similar procedure to Reichle et al.'s (1998) study in which participants are presented with a full sentence a word at a time and are asked to guess which word comes next in the sentence. For example, sentence 7 of the experiment is as follows: "*The policeman demanded to see Jim's license and registration*". The participants would first see "*The* ____" and have to guess what comes next. Once they have made a guess, a button is pressed to save their response, and the actual word is presented to them. In this example, they would see "*The policeman* ____". They now have the original presentation plus the new correct word and are asked to guess the following word again. This continues until the full sentence is presented to the participant and a new sentence begins.

Variables

The percentage of correct guesses from all participants were calculated for each word in every sentence. This percentage was an estimate of the word's Cloze probability. Using the same example "*The policeman demanded to see Jim's licence and registration*", for the word 'see' 25 participants out of 59 correctly guessed the word, so the Cloze rating would be 0.42. By the end of the sentence, for the word 'registration', 37 out of 59 chose the correct word, giving a probability rating of 0.63.

Results and Discussion

Table 1 provides descriptive statistics of Cloze probability in the Schilling corpus for the three data sets under comparison. Included are the mean score for cloze probability, standard deviation, and the minimum, median, and maximum scores. In addition, we have included the number of "full misses", which are words in which no participant was able to correctly guess the actual word.

Table 1
Descriptive statistics for Clozapp, Schnoebelen and Kuperman (2010), Reichle et al. (1998), and Clozapp.

	Reichle	Schnoebelen	Clozapp
Mean Probability	0.329	0.182	0.359
S.D.	0.364	0.241	0.334
Min Probability	0.000	0.000	0.000
Median	0.150	0.075	0.119
Max Probability	1.000	1.000	1.000
Number of "full misses"	163	152	123

Results collected through Clozapp were on par with those from the Reiche and Schnoebelen datasets (e.g., the median predictability rate from the Clozapp data, 12%, was in between the medians from the other data sets, 15% and 8% respectively). We also recalculated descriptive statistics based on the data from which words were removed if they did not elicit a single correct guess across the three subsets. Table 2 reports the results: again, Clozapp shows a median predictability rate (24%) which is in between the two earlier datasets (30% and 12% respectively).

Table 2
Descriptive statistics for Clozapp, Schnoebelen and Kuperman (2010), and Reichle et al. (1998), when words are removed if they did not elicit a single correct guess in all three corpora (in other words, "full misses" were excluded).

	Reichle	Schnoebelen	Clozapp
Mean Probability	0.403	0.232	0.293
S.D.	0.364	0.249	0.336
Min Probability	0.000	0.000	0.000
Median	0.300	0.121	0.237
Max Probability	1.000	1.000	1.000

Another comparison was conduction using Spearman's correlations between Cloze probability estimates: the non-parametric method was used since our data do not follow a normal distribution. The correlation between Reichle et al.'s (1998) data set and Schnoebelen and Kuperman's (2010) data set is strong at 0.853 ($p < 0.0001$). As for the Clozapp dataset, the correlation value with the original Schilling dataset is 0.821 ($p < 0.0001$) and with Schnoebelen and Kuperman's dataset is 0.743 ($p < 0.0001$). Based on these values, we can say that the use of Clozapp in a sentence completion task results in similar probability norms as prior datasets.

Limitations and Future Directions

Clozapp provides a simple, easy to use application for gathering Cloze probability ratings. This important measure is commonly used in many different kinds of psycholinguistic experiments and having a standard way of collecting this information is critical for research moving forwards. One potential limitation is that currently, the user interface is only available in English. While non-English languages could be input into the experiment markup files, this does not allow complete translation of the user interface (e.g., the buttons such as 'Help' and 'Save' are hard coded in English). However, anyone with experience in Java could replace these buttons with other language specific translations.

While much of the Java code is cross platform, the user interface (JavaFX) and file API calls are only supported on desktop (Windows, MacOS, Linux), so Clozapp cannot be easily ported to a mobile phone and/or tablet. To address these issues, it would be possible to re-implement this functionality into a web-based application, which would also allow data collection from remote participants.

Availability

The Clozapp User Manual, the Experiment Template, and the application itself 'Clozapp.jar' ('Clozapp_JRE13.jar' for those using JRE 13), are all available on Open Science Framework, at https://osf.io/zgxay/?view_only=26c8bb33ce624728b3c1dfb6689b3113.

References

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & cognition*, *24*(4), 477-492.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, *17*(3), 364-390.

Bickley, A. C., Ellington, B. J., & Bickley, R. T. (1970). The cloze procedure: A conspectus. *Journal of Reading Behavior*, *2*(3), 232-249.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, *20*(6), 641-655.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior*, *23*(5), 612-624.

Jonz, J. (1976). Improving on the basic egg: The M-C Cloze. *Language Learning*, *26*(2), 255-265.

Jones, C. & McRoy, S. (2003). The Cloze Test Project User's Manual. Retrieved from: http://dreuarchive.cra.org/2003/Jones/UserManual.html.

Klare, G. R., Sinaiko, H. W., & Stolurow, L. M. (1972). The cloze procedure: A convenient readability test for training materials and translations. *Applied Psychology*, *21*(2), 77-105.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504-509.

Rayner, S. (1998). Educating Pupils with Emotional And Behaviour Difficulties: Pedagogy Is the Key!. *Emotional and Behavioural Difficulties*, *3*(2), 39-47.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, *105*(1), 125.

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon mechanical turk for linguistic research. *Psihologija*, *43*(4), 441-464.

Schustack, M. W., Ehrlich, S. F., & Rayner, K. (1987). Local and global sources of contextual facilitation in reading. *Journal of Memory and Language*, *26*(3), 322-340.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311-327.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1-17.

Taylor, W. L. (1956). Recent developments in the use of "Cloze Procedure". *Journalism Quarterly*, *33*(1), 42-99.

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics*, *36*(3), 277-284.

CHAPTER 3.

Quantifying the Difference in Reading Fluency Between L1 and L2 Readers of English

Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2021). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Manuscript accepted to Studies in Second Language Acquisition*, 1-28. doi: 10.1017/S0272263121000279

Abstract

This study is a comparative examination of reading behavior of first-language (L1) Canadian and second-language (L2) Finnish and German readers of English. We measured eye-movement patterns during reading the same set of English sentences and administered tests of English vocabulary, spelling, and exposure to print. The core of our study is a novel method of statistical prediction used to generate hypothetical Finnish and German participants with maximum observed L1 scores in all component skills. We found that with L1-like component skills, hypothetical German readers can show the same reading speed as the L1 group. We hypothesize this advantage comes from the short linguistic distance to English. Conversely, hypothetical Finnish readers remain disadvantaged even with maximum component skills, likely due to a larger linguistic distance. We discuss theoretical and applied implications of our method for second language acquisition research.

Background

English is rapidly becoming one of the most widely spoken languages world-wide with 339 million native speakers, and a staggering 1.66 billion learners of English as a non-native language (Pecorari, 2018). This is largely due to the fact that many countries require a mandatory second language in its education system and English is a prevalent choice. In Europe, for instance, roughly 73% of students learn English in primary school, and over 90% learn English in upper secondary education (Eurostat, 2018).

English proficiency of non-native (labeled here as L2) speakers is a strong predictor of their academic success, employability, and -- among immigrants -- social and economic assimilation (Arkoudis, Hawthorne, Baik, Hawthorne, O'Loughlin, Leach, & Bexley 2009; Arkoudis, Baik, Bexley, & Doughney 2014; Bleakley & Chin, 2010; Graham, 1987; McManus, Gould, & Welch, 1983; Woodrow, 2006). Given the benefits and pressures of mastering English, many L2 learners have the goal of becoming as close to native (labeled here as L1) fluency as possible (Jenkins, 2009; Lambert, 2008; Olsen, 2000). A major topic of interest both in research and in educational practice is to what degree L2 speakers of English can attain proficiency in major faculties of the English language (listening, speaking, reading, and writing), what obstacles can hamper this progress, and how this proficiency compares to that of L1 speakers (e.g., Segalowitz, 2010). This topic begs two questions: how do we measure proficiency in L1 and L2 speakers, and how do we identify and quantify specific obstacles to proficiency. This paper offers a methodological proof of concept that has a potential of answering both questions for the case of fluency observed when reading for comprehension. We present the method in an eye-tracking experiment of reading behavior of English texts observed in three distinct cohorts: English L1 Canadian readers and highly proficient English L2 Finnish and German readers. We precede this presentation by a brief review of the literature on component skills of reading comprehension in English and other languages; typical discrepancies in the reading performance between L1 and L2 readers; and proposed sources of those discrepancies.

Development of reading proficiency in a foreign language is an extremely well-studied topic in educational and psychological literature as shown from a number of books (e.g., Bernhardt, 2011; Doughty & Long, 2008; Nuttall, 1996; Urquhart & Weir, 2014), and many reviews (e.g., Bernhardt, 2005; Grabe, 1991; Jeon & Yamashita, 2014; Mason & Krashen, 1997). Reading proficiency is argued to have at least two inter-related aspects: fluency (typically gauged as speed of processing) and quality of comprehension (gauged via comprehension questions or paraphrasing). As we specify below, this paper is primarily concerned with reading fluency. A wealth of psycholinguistic literature shows that L2 speakers demonstrate a slower reading rate for comprehension than L1 speakers, even when they are highly proficient in L2 (see a recent review by Brysbaert, 2019 and summaries in Beglar & Hunt, 2014 and Fraser, 2007). While proposed accounts of L2 reading comprehension vary considerably, they largely point to the same pool of causal factors. The quality of L2 reading comprehension in an individual is co-determined by her "L1 literacy", "L2 language knowledge", and "unexplained variance": see Bernhardt (2011) and Brevik, Olsen, and Hellekjær (2016), among others. L1 literacy refers to the individual proficiency in reading comprehension and component skills in one's first language which serve as the foundation for additional abilities to develop including L2 acquisition (Geva & Wang, 2001; Kuperman et al., 2020). L2 language knowledge refers to the proficiency in L2 component skills (discussed in detail below). Finally, the unexplained variance refers to the motivation or attitudes towards the L2 as well as domain knowledge or cognitive abilities.

Recent studies have quantified relative contributions of these factors to L2 reading comprehension in different L1-L2 language pairs (see reviews by Brevik et al., 2016 and Yamashita & Shiotsu, 2017). Invariably, one's proficiency in L2 has been found to be a stronger predictor than one's L1 proficiency. For example, the L1 and L2 proficiency explained 10%–16% and 30%–38% respectively in L2 reading comprehension among English learners of Spanish (Bernhardt & Kamil, 1995); and 3% for L1 and 57% for L2 proficiency respectively among Korean readers of English (Lee & Schallert, 1997). This imbalance motivates our focus on what defines "L2 knowledge" in current theoretical proposals.

A consensus exists in the current literature that reading proficiency in L2 – and both its fluency and comprehension facets -- is a multi-componential skill. It engages both general L2 verbal skills (e.g., familiarity with L2 phonology, morphology, grammar, and vocabulary, and listening comprehension), and reading-specific L2 skills and habits (e.g., decoding, spelling, and exposure to print in L2), and non-verbal variables like general knowledge, cultural capital, motivation, attitude towards L2, working memory and other meta-cognitive skills (e.g., Bernhardt, 2011; Grabe, 2009; Koda, 1996; 2005). The research effort of establishing the strongest and most robust predictors of L2 reading proficiency is massive and ongoing. A recent meta-analysis by Jeon and Yamashita (2014) examined relative contributions of 10 widely used components to L2 reading comprehension and reported L2 grammar knowledge, vocabulary knowledge, decoding and listening comprehension as the strongest and most robust predictors (see also Yamashita & Shiotsu, 2017).

Importantly, the influence of all speaker-specific factors on L2 reading proficiency is modulated by several group variables, including the distance between L1 and L2 languages and L1 and L2 scripts. As an illustration of these L1-L2 distances, consider language pairs like Serbian and Croatian, and Finnish and German. The first pair is a case of virtually identical languages but different scripts (Cyrillic versus Roman alphabetic): these languages show a small L1-L2 language distance but a large L1-L2 script distance. The second pair uses virtually identical Roman-based alphabetic scripts but the languages are fundamentally different in their genealogy (Finno-Ugric versus Indo-European), lexicon, morphology, phonology and syntax. The L1-L2 language distance for this language pair is large but the script distance is small. Generally, L1-L2 pairs with a shorter language or script distance facilitate transfer of L1 skills to L2 proficiency and are argued to give advantage in developing L2 reading comprehension and component skills (Jeon & Yamashita, 2014; Melby-Lervåg & Lervåg, 2011). Shorter L1-L2 distances tend to come with a larger advantage in L2 reading comprehension: for recent big-data studies see Schepens, van der Slik and Van Hout (2013; 2016) and van der Slik (2010). Perhaps the most studied dimension of the L1-L2 language distance has been the presence or absence of cognates, that is, L2 words orthographically or phonologically similar to their L1 equivalents (de Groot, 2006). For example, English *house* and German *Haus* and English *university* and Hebrew אוניברסיטה /univeʁsita/. Language pairs with shorter distance tend to exhibit a higher number of cognates. As a rule, cognate words in L2 are recognized faster than their non-cognate counterparts, even if the amount of facilitation varies between language pairs and tasks (Blumenfeld & Marian, 2005; Costa, Caramazza, & Sebastián-Gallés, 2000; Lemhöfer & Dijkstra, 2004; Schwartz, Kroll, & Diaz, 2007; Van Hell & Dijkstra, 2002). However, this effect has been found to vary depending on variables such as the task, the characteristics of the words and the context in which they are found. The eye-tracking study by Bultena, Dijkstra, and van Hell (2014) reports that the cognate effect in the context of a sentence depends on word class. Additional eye-tracking studies suggest that the cognate effect is especially elusive in later eye-movement measures such as the total reading time for the word (Libben & Titone, 2009; Cop, Dirix, Van Assche, Drieghe, & Duyck, 2017).

In sum, reading proficiency in L2 is a complex interaction of an individual's skills in L1 and especially L2, an individual's meta-cognitive abilities, and the script and linguistic distance between L1 and L2. Any and all of these factors may become an obstacle for developing fluency in L2 reading for comprehension. Yet so far, no analytical method has been offered to disentangle relative contributions of these factors to L2 fluency of an individual or group. Suppose a group of Russian learners of English show a lower performance both on reading fluency in English and component skills contributing to this fluency, compared to L1 English readers. It is difficult to determine how much of this difference is due to the contrast in their mastery of component skills and how much can be attributed to the L1-L2 distance. The current study presents a novel statistical method that enables researchers to disentangle these factors and estimate the expected theoretical maximum of L2 reading fluency possible for a given L1 and quantify any advantage incurred by a given L1-L2 distance. This novel statistical prediction method allows us to project different component skill levels onto hypothetical readers of a given language background. This allows us to examine the influence of L1-like skills on reading fluency in L2 speakers - data we might not otherwise have access to.

Current Study

We examine reading fluency in English by L1 and highly proficient L2 readers in English (i.e., undergraduate students in Canadian, German and Finnish universities, respectively). While reading proficiency has an additional aspect of quality of comprehension, the materials that we use provide more insight into fluency, see below. Specifically, in this study we define fluency as the speed of reading simple sentences in English. The speed is gauged via the readers' eye-movements during reading to determine both a sentence and word level measure of reading fluency.

We have four goals in our study. First, we set out to establish how L2 reading fluency compares against the L1 baseline when reading the same sentences in English. Importantly, we consider the L1 reading fluency nothing but a convenient benchmark for estimating L2 reading performance; we do not suggest that an actual or potential performance below the level of an average L1 reader is a failure on the side of a learner of English. Second, we aim to determine how the L1 and L2 cohorts vary in terms of select few component skills of reading in English. A third goal is to identify how much the fluency of L2 readers is hampered by suboptimal component skills, and how much it can be improved if those skills were as developed as L1 readers of English. Finally, we aim to point to component skills of L2 reading comprehension that contribute the most to L2 reading fluency in Finnish versus German L2 readers.

To pursue the first goal, we conducted an eye-tracking study in all three cohorts (Canadian, Finnish, and German) that read an identical set of sentences in English while their eye-movements were recorded. The empirical base for the second goal was a small battery of tests tapping into critical component skills of reading comprehension in English (vocabulary size and spelling), and a more general index of reading proficiency (exposure to print): see details below. We expect L2 readers to show lower reading fluency and a lower level of L2 component skills than their L1 counterparts, in line with multiple reports in the literature (see above).

The remaining two goals constitute the novel methodological part of this study. The core of the method is a statistical prediction of the reading speed of L2 readers with different levels of proficiency. As discussed above, this method offers researchers a different perspective of L2 reading fluency by telling apart the contributions of individual component skills of L2 reading proficiency (e.g., English orthographic knowledge, or vocabulary) and a systemic impact on L2 readers that derives from the linguistic distance between their L1 and L2. More specifically, we used the predictive function of multiple regression models to estimate what their reading speed *would be* if the L2 readers had a maximum L1-like proficiency in component skills. One possible outcome of the statistical prediction analysis is that the predicted reading speed of a hypothetical German or a Finnish reader endowed with maximum performance of an L1 speaker would become comparable or faster than a typical L1 reader of English. In this case, the main reason for observed lower reading speed in the actual L2 readers is their imperfect mastery of English component skills, and L1-L2 distance carries no inherent processing cost. The difference between the observed reading fluency in L1 and L2 readers can be eliminated in this scenario if L2 readers improve one or more of the skills. An alternative possibility is that a hypothetical L2 reader would still lag behind a typical observed L1 reader in their predicted reading fluency, even if they were assigned the top L1 scores in all component skills. In this case, the difference in reading speed between the 'ideal' hypothetical L2 readers and observed L1 readers will signify the cost for the L2, which cannot be compensated for by the maximum proficiency in component skills. We compare the Canadian vs German and Canadian vs Finnish readers of English to provide quantitative evidence and theoretical interpretation for one of the scenarios.

Several linguistic considerations are of relevance for this goal. In the set of languages considered here, both English, Finnish, and German show little script distance: all three utilize the Roman-based alphabetic script. However, German and English share much more with one another in terms of the lexicon, phonology, morphology, and syntax than Finnish and English do. German and English are from the same branch of Germanic languages with substantial overlap in their lexicon and similar morphosyntactic principles. In contrast, Finnish is a Finno-Ugric language with a largely different lexicon and agglutinative morphosyntax, leading to much more inflectional complexity than in English or German. We predict that an overall language advantage will be apparent when comparing German and English than when comparing Finnish and English, because the L1-L2 distance is shorter in the former case. The two language pairs may thus exhibit different patterns of reading behaviour as described in the hypotheses above.

Our use of regression models also enables identification of the component skills that give L2 readers the most leverage (i.e., have the strongest effect on increasing their reading speed). This step has an immediate educational relevance, as it points to skills that are the most beneficial to develop for speakers of a given L2.

In sum, the methodological approach that we propose offers a tool to answer the question of whether native-like fluency in L2 is strictly a matter of acquiring an ideal performance in component skills, and what specific skills are mostly responsible for the differences between L1 and L2 reading fluency. We emphasize that both the novelty and the focus of this paper is developing and demonstrating effectiveness of a statistical procedure. To this end, we made use of reading materials, component tests, and cohorts from earlier published (Rahmanian & Kuperman, 2019) or unpublished studies that pursued a separate purpose, such as examining effects of spelling variability on L1 and L2 reading of correctly spelled words embedded in English sentences (see Rahmanian & Kuperman, 2019). For this reason, we do not claim that the present texts, tests and samples of participants are optimal for drawing a comprehensive cross-linguistic picture of reading in English (see 'Limitations and future directions'). We do argue, however, that the empirical base available for this study is sufficient for developing an accurate statistical procedure to determine whether there is an irreducible processing cost that accompanies a specific L1-L2 pair (see above for motivation).

Methods

This study consists of three cohorts of university students completing the same set of tasks in English: one cohort is comprised of L1 speakers of English recruited in Canada, while the remaining two cohorts are L2 speakers of English recruited in Finland and Germany, respectively. The L1 data have been reported in Rahmanian and Kuperman (2019) and serve as a baseline for the two additional datasets from L2 readers of English.

Participants

The L1 cohort was recruited from an undergraduate departmental subject pool at McMaster University (Hamilton, ON, Canada). Thirty-five participants were tested and two were dropped due to below average reading comprehension scores leaving 33 total participants (24 female, mean age 20.9). The L1 cohort were all native speakers of English. A total of 27 native Finnish speakers (25 female, mean age 21.5) from the subject pool at the University of Turku, Finland were tested, as well as 33 native German speakers from the subject pool at the University of Würzburg, Germany (28 female, mean age 21.9). All participants reported normal or corrected-to-normal vision and no visual or learning impairments.

Both L2 cohorts and the L1 cohort were undergraduate students. None of the L2 participants have lived more than 3 months in an English-speaking country. All of them had acquired English proficiency by virtue of going through an educational system that mandates multiple years of English instruction at the primary, secondary and post-secondary level, and imposes rigorous exams of English proficiency as an entrance condition for university education. While English is not the medium of instruction in either Finland or Germany, most reading materials for university students are only available in English, so their L2 reading skills are constantly practiced.

Materials

This study used stimuli from Rahmanian and Kuperman (2019). These were seventy syntactically simple sentences. The sentences were made so that they fit onto one line, such as '*He was bothered by the asymmetric painting in the office'.* All cohorts saw the same stimuli during the reading task, while their eye-movements were recorded. Following half of the trials, simple yes-no questions were administered. These simple questions were included to ensure participants were indeed engaged in reading the sentences presented and paying attention to the task at hand. These were not intended to represent the depth of comprehension of the participants. The comprehension scores were expected to be at ceiling for participants who were paying attention and were used as a removal criterion for participants who responded with low accuracy (i.e., less than 80%) on these questions. For these reasons (a small percentage of questions and a high response accuracy), the present materials are not optimal for a statistical prediction of reading comprehension. While this paper focuses on reading fluency determined through eye-movements, the same mathematical apparatus can be applied to estimating how accurate the L2 readers comprehension would be if they possessed the L1-level of English component skills.

The experimental manipulation in the original paper was the selection of words that are often misspelled in English: however, all words presented to participants were spelled in a conventional way and did not pose a problem for the present study. Additionally, in Rahmanian and Kuperman (2019), two target words contained typos which were corrected in the stimulus list presented to the Finnish and German participants. Thus, the analyses below are based on 68 sentences for the L1 cohort and 70 sentences for both L2 cohorts. Stimuli are available as supplementary materials to Rahmanian and Kuperman (2019).

In addition to the sentence reading task, three offline tests were chosen to evaluate the English skill level of each individual participant: that is, tests of spelling skill, exposure to print, and vocabulary size. The spelling test was chosen as a measure of orthographic processing ability which has been shown to impact reading speed and comprehension (Andrews & Hersch, 2010; Andrews & Lo, 2013). Participants are presented with a list of 88 words and are asked to circle those that are spelled correctly with points awarded for those correctly identified for a maximum possible score of 88. The spelling recognition test is known to be correlated with many linguistic abilities and shows high test-retest reliability at $r = 0.93$ (Andrews, Veldre, & Clarke, 2020). Published reliability estimates are not available for L2 readers of English, as far as we know.

The task tapping into one's exposure to print was the Author Recognition Task (ART) first developed by Stanovich and West (1989), and later refined by Acheson, Wells, and MacDonald (2008). The ART presents participants with a list of 130 names, only half of which are real authors. Participants are instructed to indicate all the authors they are familiar with and are scored based on how many they indicate correctly minus those they incorrectly identified. A maximum score of 65 is possible, negative scores are possible as well. The authors in the stimulus list predominantly represent the Western and especially the English-language literary tradition, yet – as will be come important in comparisons below – knowledge of these authors can be gained from reading in other languages than English. Thus, the ART is a test of exposure to print but is not specific to English reading, thus the L1 advantage over L2 readers is less expected in this test than in the language specific ones (e.g., English spelling and vocabulary), see McCarron and Kuperman (2020).

The ART is known to be a strong predictor of reading performance as well as eye-movements characteristic of reading skill (Mol & Bus, 2011, Moore & Gordon, 2015). Reliability of the ART has been shown through a split-half test at $r = .82$ (Stanovich & West, 1993). Moreover, in the meta-analysis by Mol & Bus (2011), Cronbach's alpha values for reliability for print exposure tests fell between $r = .75$ and .89. Published reliability estimates are not available for non-native readers, to our knowledge.

To measure vocabulary, the test LexTALE (Lemhöfer & Broersma, 2012) was chosen, in which participants are presented with 60 items (40 words and 20 nonwords) on a computer screen one at a time and are asked to judge whether or not the presented item is a word in English. An accuracy score is given in percentage out of 100. Vocabulary measures are highly correlated with a number of linguistic characteristics and the reliability of the LexTALE test has been shown at $r = .81$ for Dutch L2 speakers, and $r = .68$ for Korean L2 speakers (Lemhöfer & Broersma, 2012).

The spelling test and ART were administered to all three cohorts, while the vocabulary test was only administered to the two L2 cohorts.

Design

Reading fluency of L1 and L2 English readers is of critical interest in our study. Temporal data was obtained using eye-tracking, an effective and well-developed methodology for studying both the basic parameters of reading fluency and the individual variability that it gives rise to (Ashby & Clifton, 2005; Chace, Rayner, & Well, 2005; Kuperman & Van Dyke, 2011; Rayner, 1998). Eye-tracking enables researchers to focus on both cumulative measures of reading fluency at the level of a sentence or a longer text, and more specific ones reflective of the fine-grained oculomotor and lexical processes in reading behavior at the word level. To ensure completeness of examination, we measured fluency as reading speed both at the sentence- and the word level.

The sentence level analysis defined reading fluency as reading rate, in words per minute (wpm): sentence length in words was divided by the total time spent reading the sentence (and recalculated per minute). This measure is considered most related to the received ways of assessing individual performance in reading comprehension tests and other educational skills (see reviews by Brysbaert, 2019, Fraser, 2007 and references therein). We complemented our report of reading rate by analyses on the word level. While there are many measures that can be used to investigate the time-course of individual word processing (e.g., first fixation duration, gaze duration etc.), we focused on total fixation duration (the summed duration of all fixations on the word) as it is a cumulative measure for word level processing and is more indicative of the overall processing effort associated with the word and a good reflection of processing speed (Rayner, 1998). Below we demonstrate our method by presenting analyses of the observed reading fluency (reading rate and total fixation duration) and the reading fluency predicted for a given level of English proficiency. In the analyses of word-level fluency we took into account the fact that words varied in their word length and frequency of occurrence. Word length and word frequency are known as benchmark predictors of eye movements during reading and their effects are found both in L1 and L2 reading (Rayner, 1998). Word length was measured in characters, and frequency information was collected from the 51 million-token SUBTLEX lexical resource based on subtitles to the US films and media (Brysbaert & New, 2009).

Experimental Procedure

Participants started with the eye-tracking task where they were asked to read sentences silently for comprehension while their eye-movements were recorded. Each participant read all the sentences individually, one at a time, in a randomized order. Participants were instructed to press the space bar after they had finished reading each sentence. Following half of the trials, yes-no comprehension questions were presented to which they responded either ‘true’ or ‘false’ using specified buttons on the keyboard.

Participants were comfortably seated approximately 60 cm away from the screen with their head stabilized on a chin rest to avoid movement. A three-point horizontal calibration was used, and the experiment would only begin if good validation (maximum average error below 0.3 degree of visual angle) was achieved. Eye movements were collected from a single eye using the Eyelink 1000 (SR Research, Kanata, ON, Canada). Each trial started with a drift correction to ensure calibration was stable throughout the experiment and to increase the accuracy of eye-movements. There were four practice trials to familiarize the participants and allow them the chance to ask any questions, after which the experiment started.

Following the eye-tracking portion, the participants completed the spelling test (Andrews & Hersch, 2010), and the ART (Acheson et al., 2008; Stanovich & West, 1989). The L2 cohorts also completed the vocabulary test (Lemhöfer & Broersma, 2012). An unrelated experiment was run during the same session. In total, the participants spent approximately 50 minutes completing the study.

Statistical Prediction Procedure

The core of the present study is a method for obtaining statistical predictions of reading behavior for participants at different levels of proficiency in component skills of English as L2. We first describe the method for reading rate as a dependent variable. We fitted linear multiple-regression mixed-effects models to German and, separately, Finnish data, with reading rate as a dependent variable and scores in spelling, vocabulary size and ART as participant-level predictors. The random effects structure included by-subject and by-item intercepts. Function lmer from library lme4 v 1.1-19 in the statistical software environment R v 3.4.4 was used to fit the regression models (R Core Team, 2013; Bates, Maechler, Bolker, & Walker, 2015). The models for both the German and Finnish data were fitted, trimmed of outliers (data points with standardized residuals greater than 2.5 standard deviations), and fitted again. Since our main goal for the model use was statistical prediction and we did not rely on the inferential estimates of the slopes, the potential harmful impact of collinearity was not an issue.

As a next step, we used individual models for each L2 cohort to generate predictions for hypothetical readers in those cohorts. Specifically, we were interested in the performance of an *ideal* L2 reader (i.e., a reader with maximum scores in all component skills obtained by the L1 cohort), as well as the performance of a *median* L2 reader (i.e. a hypothetical reader at the median level of L1 proficiency). To study hypothetical readers, we extracted the median and the maximum scores in the spelling test (77 and 87, respectively) and ART (7.5 and 41 respectively) from observed L1 participants. Since L1 participants did not complete the vocabulary task, we estimated the range of this score as the median score observed in a combined pool of L2 participants (72.5) and the maximum of the scale (100). Then we generated a set of hypothetical readers that would represent all combinations of extreme score values for the three tests: this pool included a hypothetical reader with the median scores in all component skills (spelling 77, ART 7.5 and vocabulary 72.5), an ideal reader with all highest scores (87, 41 and 100, respectively) and readers with all other combinations of the median and maximum scores. (Any other percentile in the skill distribution can be chosen instead of the presently used median and maximum).

Finally, we used the combinations of test scores that represented hypothetical readers to predict their reading rate based on the regression models fitted to the data observed in German and, separately, Finnish readers. To this end, both the models and the test scores were submitted to function predict() available in the stats package of the statistical software environment R. Function predict() generates values of the dependent variable for unseen values of predictors based on the estimates for the intercept and the slopes in the original regression model. As a result, reading rate estimates were generated for each hypothetical German and Finnish reader and each unique combination of test scores. These estimates represented our best guess at the behavior of German and Finnish readers representing the median-to-maximum range of native proficiency in those skills, see below.

Our method allows for fine-grained predictions not only for participants with a specific proficiency level, but also for words with specific lexical parameters. This is demonstrated in our analyses of total fixation times. Linear multiple-regression mixed-effects models were fitted to German and Finnish data, with total fixation times to words as a dependent variable, word length and frequency as text-level predictors, and scores in the spelling task, vocabulary size test and the ART as participant-level predictors. Furthermore, we allowed both word length and word frequency to interact with each of the participant-level predictors: the interactions had the following form (word length + word frequency) x (spelling + vocabulary + ART). The random effects structure included by-subject and by-item intercepts, by-subject slopes of length and frequency, as well as correlations between all by-subject estimates. No three-way interactions were included. Total fixation times and word frequency values were log-transformed to reduce the influence of outliers on the model estimates.

All unique combinations of word length, word frequency, and median and maximum scores in the three tests of individual differences were supplied as input to predict() function, along with the regression models fitted to total fixation times in Finnish and German. The outcome was the predicted total fixation time that a hypothetical Finnish or German L2 reader of English (including the ideal reader with maximum scores) would demonstrate when reading a word of a given length and frequency.

Predictions for both reading rate and total fixation time are pointwise estimates of how an average L2 reader would perform given certain skill levels. To estimate variability of these estimates, we bootstrapped (sampled without replacement) eye-movement data in 1000 iterations and fitted a linear mixed-effects model to each iteration using function bootMer in the statistical platform R. We used each model to derive predictions for the hypothetical readers as described above. The resulting distribution of estimates produces a prediction interval, that is, a range within which a future observation (e.g., reading rate) is expected to occur given prior observations. We use the range between 2.5 and 97.5 quantiles of the range as an estimated 95% prediction interval: below we report those intervals along with the pointwise estimates. We expect the prediction intervals to be narrower for the median than for the maximum hypothetical scores because the median skill of an L1 speaker is more likely to be represented within the L2 sample that the prediction-generating model is based on. The maximum L1 skill is more likely to be out of the range that L2 skills occupy and so this prediction is expected to be more variable. All datasets and R code used for the statistical prediction methodology outlined here are available from Open Science Framework (OSF) at this link: https://osf.io/ex9fj/.

Results

Component skills

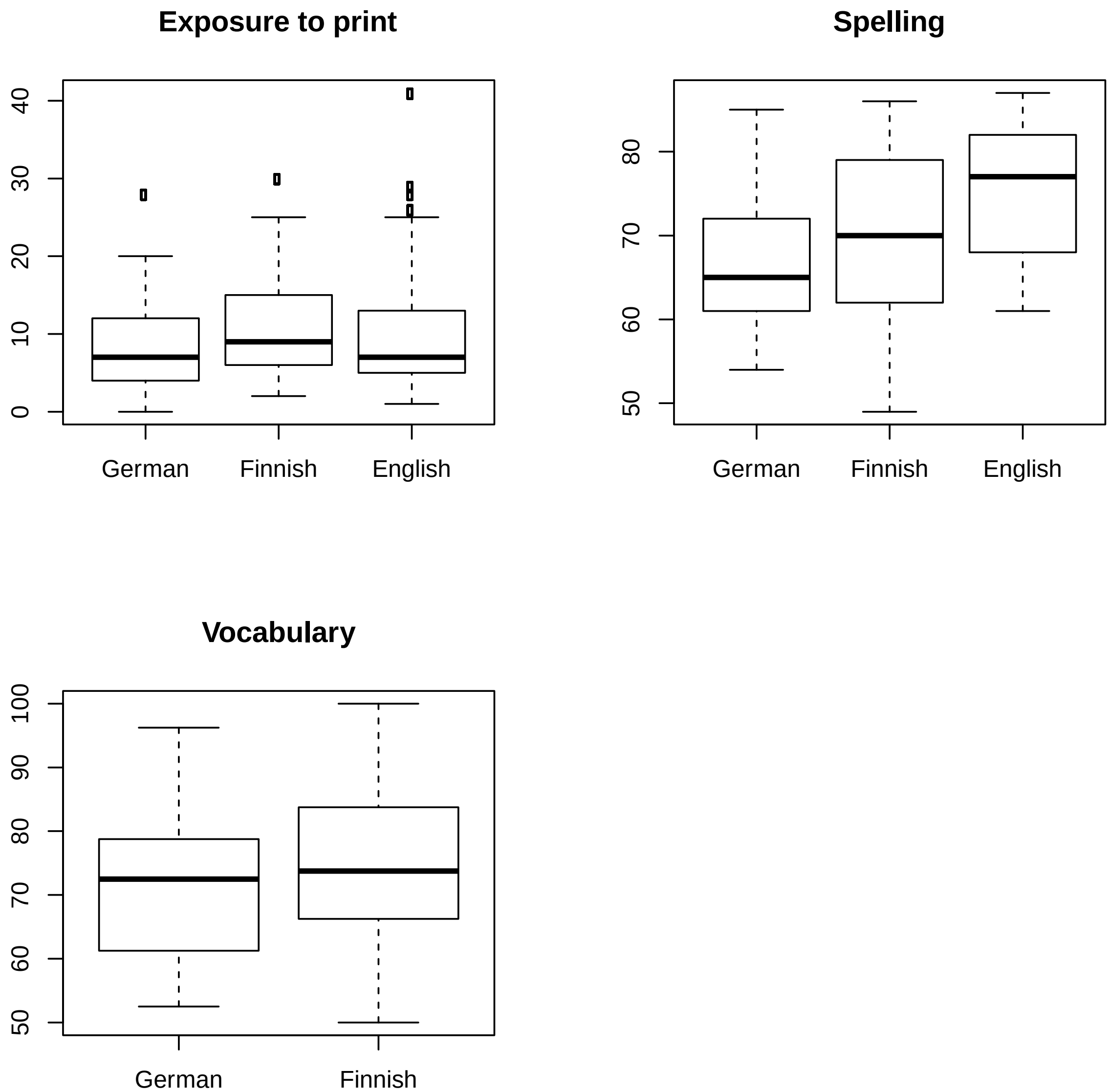
Our initial analysis examined whether different component skills varied by language background. The descriptive statistics for the ART, spelling and vocabulary tests for each language is found in Table 1, and Figure 1 illustrates the distributions of scores for L1 and L2 cohorts.

Table 1. Descriptive statistics for the three component skills tests by L1.

	ART				Spelling				Vocabulary			
	Mean	Med	SD	Range	Mean	Med	SD	Range	Mean	Med	SD	Range
English	10.74	7.5	9.07	1:41	74.79	77.0	7.75	61:87	------	-----	------	--------
Finnish	11.14	8.5	6.08	2:30	70.14	69.50	10.28	49:86	74.0	74.0	13.0	50:100
German	8.55	7.0	5.99	0:28	66.52	65.00	7.54	54:85	72.46	72.5	11.73	52:96

Multiple-regression mixed-effects models fitted separately to ART, spelling, and vocabulary scores with language as a predictor revealed significant differences between the cohorts as shown in Supplementary Materials (Tables S1, S2, and S3). On average, the Finnish cohort had reliably higher ART scores than the Canadian cohort, and both performed reliably better than the German cohort. The advantage of the Finnish cohort over the Canadian cohort may appear surprising, as the former represents L2 speakers of English. It is worth a reminder however that the ART test consists of names of fiction authors belonging to the Western literary tradition. The observed advantage in the Finnish cohort may simply indicate their better familiarity with this tradition, which they may have gained while reading both in Finnish, English, and other languages. The boxplot additionally shows that, unlike the average trend, more Canadian speakers attained very high scores on the ART test compared to the Finnish cohort: this observation will become important when predicting reading behavior of hypothetical L2 speakers. For the spelling test, there was a clear advantage for the Canadian cohort followed by the Finnish and then German cohorts: all contrasts were reliable. Finally, in the vocabulary test, Finnish scores were significantly higher than German scores.

Figure 1. Boxplots of the three component skill tests by L1.



The Canadian cohort accounted for the highest scores of both the ART and spelling tests. While we did not conduct LexTale in the Canadian sample, previous studies using L1 English speakers within the same age group have cited an average of about 98% on the LexTale vocabulary test (Diependaele, Lemhofer, & Brysbaert 2013). The significantly higher ranges in all tests for the L1 cohort suggests that assignment of maximum scores to L2 readers of English may give them a boost in reading fluency. The contrasts between the German and Finnish performance also suggest that the magnitude of the boost may differ by language cohort. We examine these possibilities below.

Eye-movements

The original data pools included 63,555 total observations from all three datasets, 22,746 for English, 22,077 for German, and 18,732 for Finnish. Reading rate was calculated as the number of words in a sentence divided by the time it took for the participant to read that sentence, measured in words per minute (wpm). In calculating total fixation duration, some observations were removed. Specifically, all words that were not fixated during reading (skips) were removed from analysis, leaving 14,199 observations for English, 17,353 for German, and 15,177 for Finnish. From this we can calculate skipping rate for each language. The Finnish cohort skipped the least number of words (19%), followed by the German cohort which skipped 21% of words, while the Canadian cohort skipped the most (38%). This is in line with prior research that shows that more proficient readers skip more words (Rayner, 1998; 2009).

Common function words are most likely to be skipped and provide little information regarding text processing (Rayner, 1998; Drieghe, 2008). All function words and their respective fixations were removed, leaving us 8,951 English observations, 10,567 German observations, and 9,108 Finnish observations remaining. We further removed all data points in which ToFD was lower than 80 ms and any words whose total number of fixations was greater than or equal to 5. Finally, outliers from the outer 1% from both sides of the distribution for each participant were removed. This left a total of 8,405 observations for English, 9,722 for German, and 8,289 for Finnish.

Descriptive statistics for the words used in the sentences for the eye-tracking task were calculated for word length (range = 3:12, mean = 6.62, SD = 2.17) and word frequency per million (range = 0.039:9773, mean = 254, and SD = 788).

Our main eye-tracking variables of interest were reading rate (RR) and total fixation duration (ToFD). The descriptive statistics for each language cohort can be found in Table 2.

Table 2. Descriptive statistics for the main eye-tracking variables for each cohort. RR = reading rate. ToFD = total fixation duration.

	RR (words per minute)				ToFD (ms)			
	Mean	Median	SD	Range	Mean	Median	SD	Range
English	329.08	312.50	120.71	92-990	323.40	270.0	201.11	88-2747
Finnish	178.51	174.14	70.32	36-505	483.01	408.0	288.5	98-2526
German	211.04	200.24	85.18	49-541	418.65	356.0	247.80	88-2355

Reading Rate

We first looked at the overall differences in how fast each cohort read the sentences (measured in wpm). As Table 2 shows, reading rate of L1 readers of English is much higher (329 wpm) than that reported as typical for text reading in Brysbaert's (2019) meta-analysis, even after correcting for word length. This is likely because sentence reading is simpler than passage reading, which Brysbaert's estimates are based on: there is no inferential structure or discourse variables to keep track of within sentences. Moreover, the reading rate of the Canadian cohort was much higher than that of the German cohort (211 wpm) and Finnish cohort (179 wpm). Median reading rates of the German and Finnish cohorts were equal to the 12th and 6th percentiles of the reading rate in the Canadian cohort, respectively. The lower level of fluency in L2 compared to L1 is expected even in highly proficient L2 readers, as suggested in the literature (Bernhardt, 2011; Nuttall, 1996; Urquhart & Weir, 2014).

Mixed-effects multiple-regression models confirm all contrasts between languages as highly statistically significant (all $p < 0.001$). see Table 3.

Table 3. Linear regression model of reading rate as a function of language with German as baseline.

	Estimate	SE	t	p
Intercept	211. 037	2.037	103.58	< 0.001
English	118.039	2.838	41.59	< 0.001
Finnish	-32.531	2.993	-10.80	< 0.001

Statistical prediction

This analysis highlights differences in the baseline reading rate of the three cohorts. These may be partly due to the differences in English L2 skills of individual German and Finnish readers, and partly due to a systematic cost that L2 speakers experience regardless of skill level. Our next step was to examine whether the observed differences in reading rate would be reduced if we project native-like proficiency in the chosen component skills on L2 reading performance.

We first created mixed-effects linear models for German and Finnish reading rates independently, taking into consideration all available skill components and between-item and between-subject variability, see Methods (Table 4 for German; Table 5 for Finnish). These models served as the basis of the statistical prediction of hypothetical readers.

Table 4. Reading rate as a function of component skills: German. SD of the by-items random intercepts is 27.99, SD of the by-subject random intercepts is 52.10; the model residual SD is 49.20; N = 1618.

	Estimate	SE	Df	t	P
Intercept	48.7274	87.6731	29.0605	0.556	0.5826
Spelling	-0.9072	1.6895	28.9714	-0.537	0.5954
Exposure	3.0504	1.8034	29.0023	1.692	0.1015
Vocabulary	2.7184	1.0700	28.9803	2.541	0.0167

Table 5. Reading rate as a function of component skills: Finnish. SD of the by-items random intercepts is 24.48; SD of the by-subject random intercepts is 51.32; the model residual SD is 40.69; N = 1920.

	Estimate	SE	Df	t	P
Intercept	-13.2601	73.1063	24.0214	-0.181	0.8576
Spelling	2.7172	1.4553	23.9470	1.867	0.0742
Exposure	-0.6177	1.7464	23.9502	-0.354	0.7266
Vocabulary	0.1101	1.3014	23.9433	0.085	0.9333

Based on the models shown in Table 4 and 5, we used statistical prediction (the predict() function) to estimate the reading rate of hypothetical readers with the median and maximum scores in all component skill tests: See Methods above for details. The results of the prediction are shown in Figure 2. It shows observed median reading rates for the three cohorts as well as the 10^{th} and the 90^{th} percentiles of the reading rate in the Canadian cohort. Critically, it also shows the pointwise estimates and the prediction intervals (error bars) of the reading rate expected from a hypothetical German and a Finnish reader with the maximum L1 scores on all component skill tests and the median L1 scores.

Figure 2. Predicted reading rates for hypothetical L2 readers with median and maximum performance in component skills (in grey) with prediction intervals (error bars), and observed median (and for the Canadian cohort, the 10th and 90th percentile) reading rates from the three cohorts (in black).

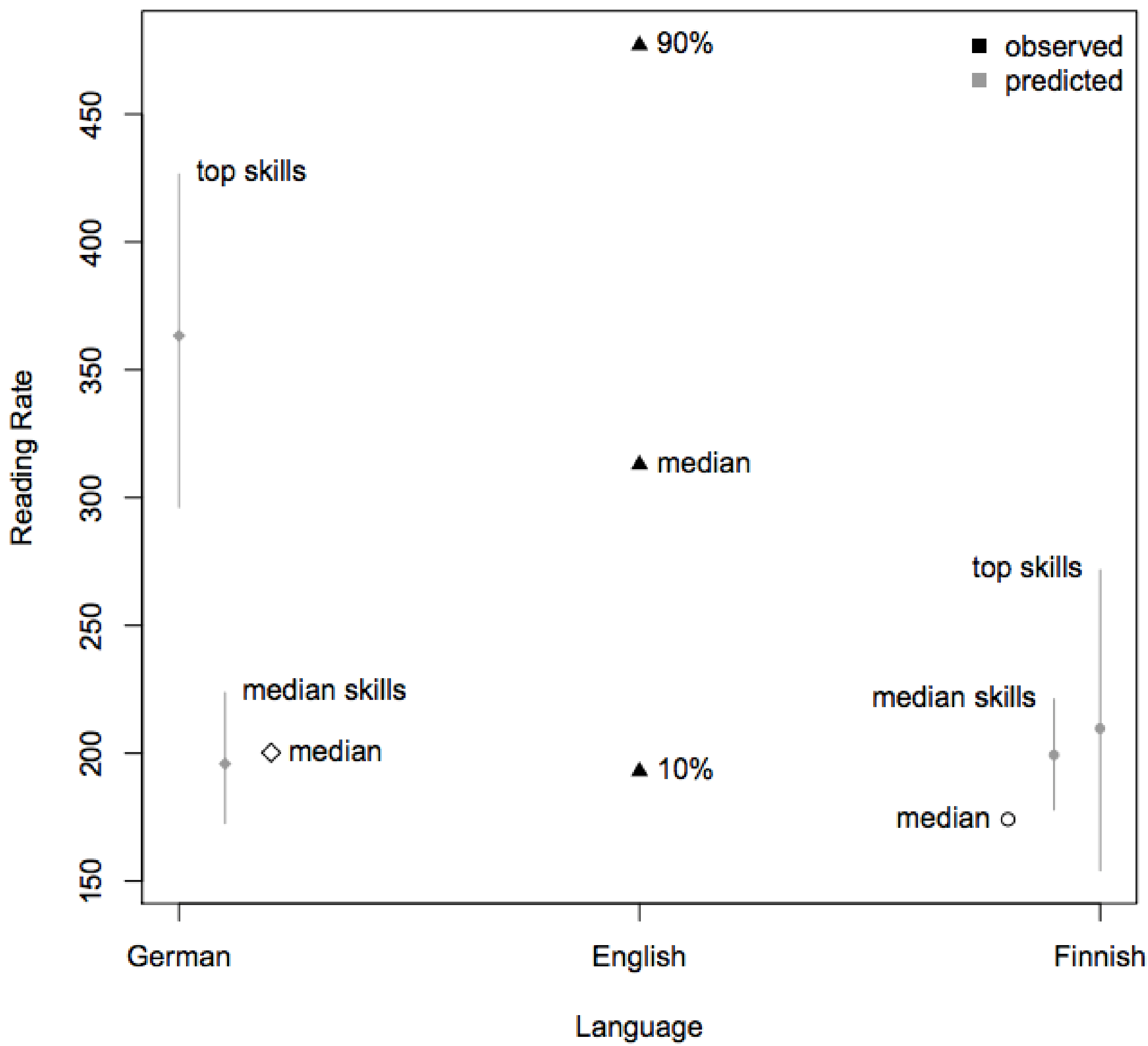


Figure 2 shows clear differences between L2 cohorts. For Finnish hypothetical readers with imposed median L1 skills, the predicted reading rate is only slightly higher than the observed median reading rate of our Finnish sample. This performance is equivalent to the 11th percentile of reading rate distribution in the Canadian cohort. Even with the top-level L1 skills projected onto them, the predicted value of reading rate in Finnish hypothetical speakers is not much higher than the observed Finnish median reading rate (the 6th percentile) and is equivalent to the 13th percentile in the L1 cohort.

For the German hypothetical readers, if given median skills of a L1 reader, the predicted reading rate (198 wpm) is very similar to the observed German median reading rate (200 wpm) and is equivalent to the 11^{th} percentile in the L1 cohort. However, we see that if German hypothetical speakers are given top skills, they exceed the observed median reading rate of an L1 reader: they are at the 67^{th} percentile. In other words, an ideal German reader outperforms an average Canadian reader.

As expected, we see greater variance in the prediction intervals for the top skills than the median skills. For Finnish readers, since the advantage of receiving the top skills is quite small (2% higher) compared to the median skills, the lower boundary of the prediction interval is pushed below the actual median of the observed reading rate. On the other hand, for German readers, we do not find any overlap of the prediction interval with the actual median scores since the advantage of receiving the top skills is quite large (56% higher).

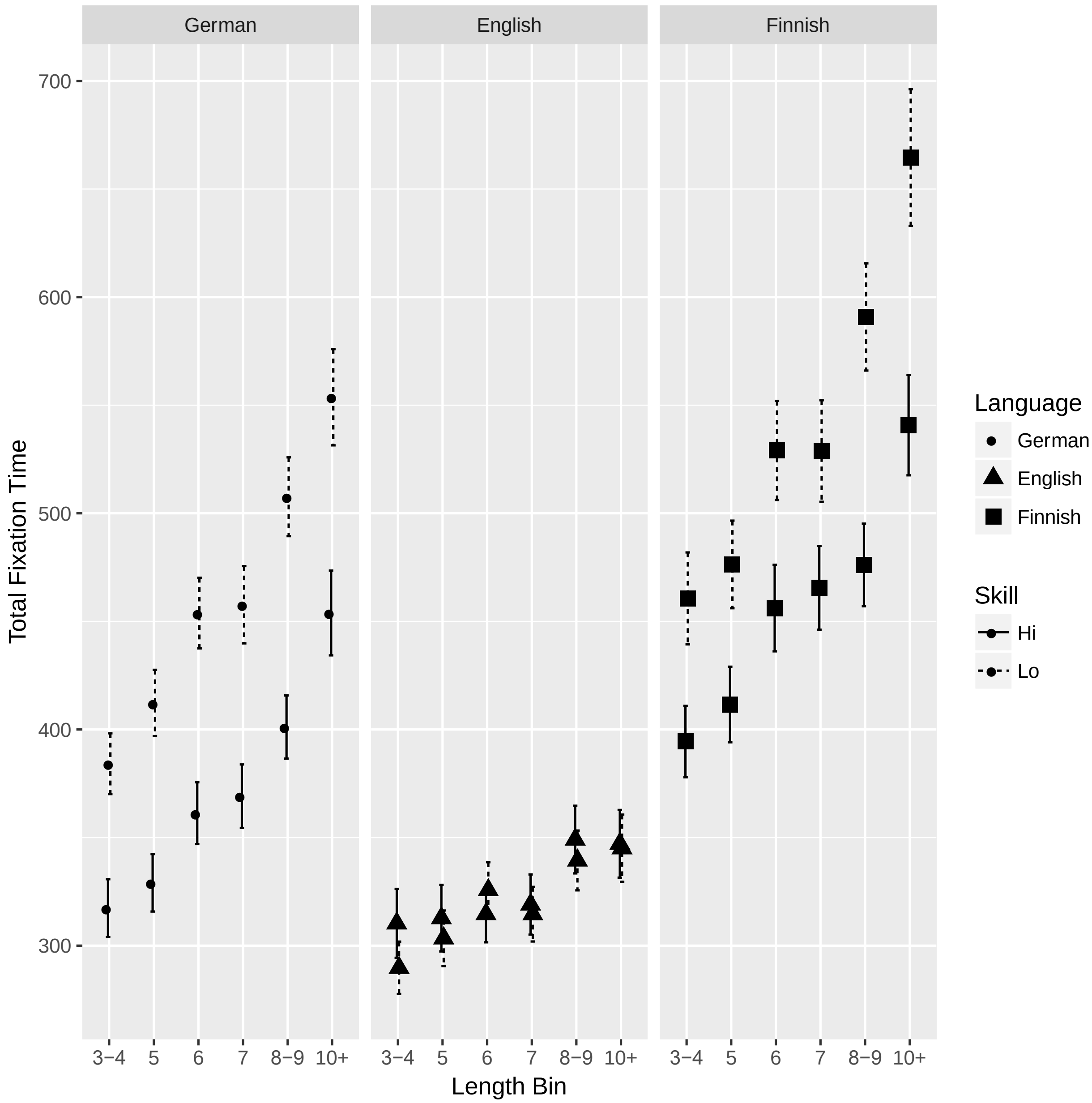
These results are consistent with the idea that there may be a potential processing advantage that comes with a specific L1. For German participants, we see an advantage of language similarity because with maximum component skills, they reach above-median L1 fluency. Conversely, we see no advantage for Finnish participants. Given the same affordances, they are only ever predicted to reach a level of fluency that is below the median of L1 fluency. We discuss below what linguistic features might account for the advantage that German readers have over Finnish in reading English as L2.

Total Fixation Duration (ToFD)

While reading rate characterizes overall differences between L1 and L2 readers, additional information can be gleaned from more fine-grained measures at the word level. This analysis looks at ToFD and examines whether reading fluency is modulated not only by language distance but also by length and frequency of individual words.

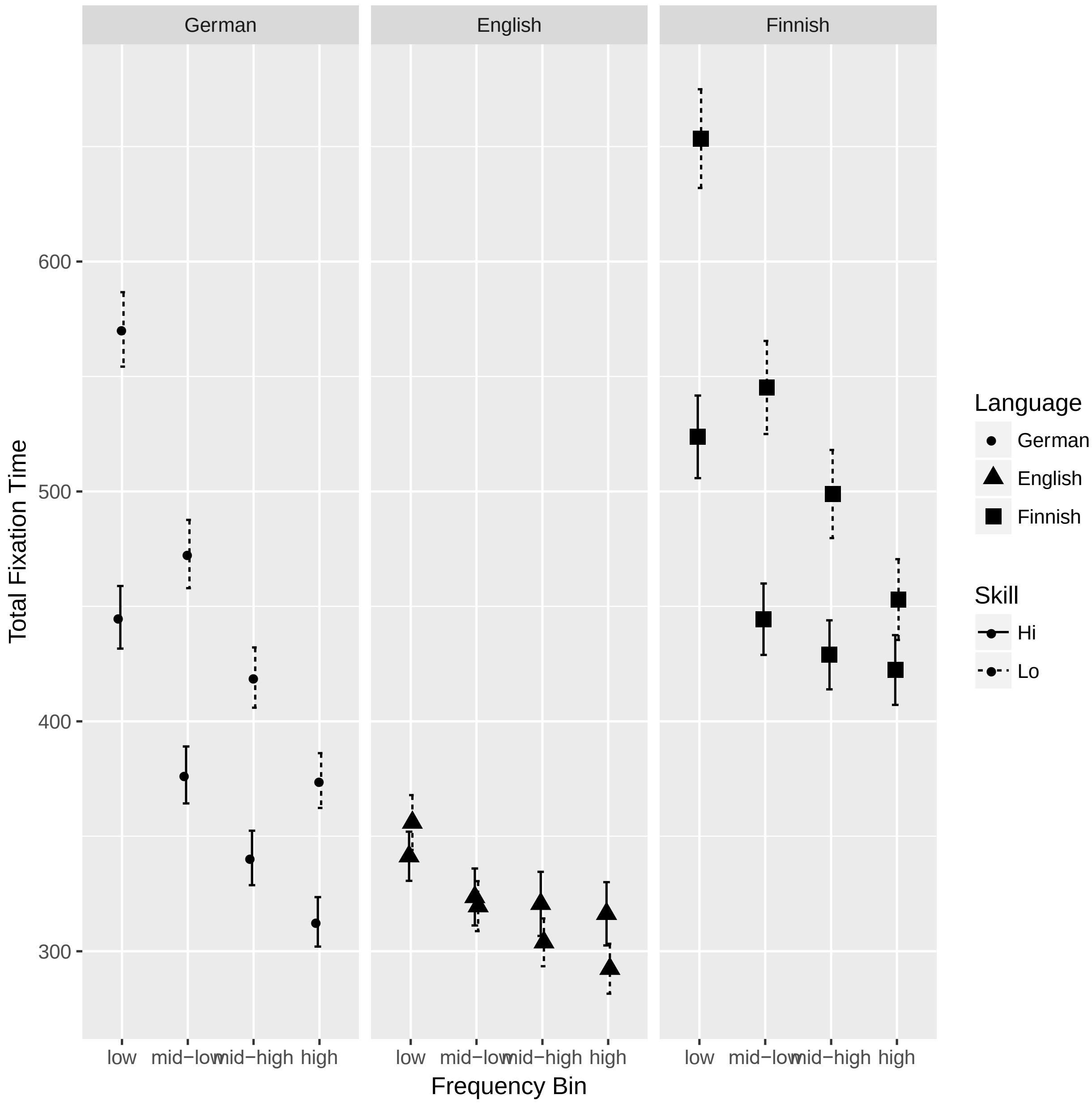
Figure 3 shows how word length impacts total fixation time. For illustrative purposes, we half-split participants by skill to create a ‘high skill’ and ‘low skill’ group in each language. We wanted to use the same criteria for all three languages while ensuring roughly equal groups. This was done by pooling the available data from each language (spelling and ART from all three groups, and vocabulary data for the L2 groups) to find an appropriate cut off point. The criteria chosen to split into the high skill group was a spelling score higher than 79 and an ART score higher than 8. This split ended up assigning 14 German participants to the high group (19 to the low group), 21 Canadian participants to high (13 to low), and 17 Finnish participants to the high group (11 low). We also binned word length into 6 distinct categories (‘3-4’, ‘5’, ‘6’, ‘7’, ‘8-9’, ‘10+’) to ensure a roughly equal number of words per bin. Analyses below were conducted on the entire data set, without splitting or binning. Overall, Figure 3 showed a large main effect of language, with Canadian participants the fastest followed by German and then Finnish. It also revealed an expected effect of word length: longer words were read slower. These observations were supported by the regression model (see Supplementary Materials, Table S4). Skill level was a strong factor in L2 cohorts but showed little influence for L1 readers. Importantly, Finnish readers showed longer total fixation times than the other two cohorts across the word length range. Conversely, the high-skill German readers of English showed L1-like speed of word processing of short English words.

Figure 3. Total fixation time as a function of length, for each language and skill level.



The influence of word frequency on readers broken down by their L1 and skill level is displayed in a similar manner in Figure 4. The word frequency data was split into 4 distinct categories based on quartiles. Skill level cut-offs for high and low are the same. Again, we observed that skill came with reduced total fixation times in both L2 cohorts but not the L1 cohort. Finnish readers were slower than the other cohorts, while high-skilled German readers approached L1-like total fixation times in high-frequency words.

Figure 4. Total fixation time as a function of frequency for each language and skill level.



Statistical prediction

Similar to reading rate, we used the statistical prediction procedure to look at word-level differences in hypothetical German and Finnish readers. These predictions were based on baseline models for German and Finnish that included length and frequency as well as all three component skills (See Supplementary materials Table S5 and S6, respectively). We took all combinations of median and maximum scores from the distribution of L1 speakers (see above) and used these to create hypothetical readers for all possible combinations of word length and frequency. For example, there would be a datapoint for a hypothetical German reader with the highest ART score, highest spelling score and the median vocabulary score for a word length of 6 in the high frequency bin. A hypothetical predicted datapoint exists for all combinations of these variables based on both the German and Finnish observed data. The prediction procedure can be easily adjusted to produce predictions for any other percentile in the distribution of individual skills.

Figure 5 compares English to Finnish speakers in which the median observed data for each combination of length and word frequency is shown as black triangles for English and black open circles for Finnish. The predicted data for Finnish speakers is shown as grey circles in the background. The fastest-reading hypothetical Finnish participants were typically on par with the 10^{th}-30^{th} percentile of ToFD in L1 readers, and only occasionally approximated their median reading speed.

Figure 5. Observed median total fixation times for Finnish (black open circles), and English (filled black triangles), and predicted total fixation times for hypothetical Finnish (small grey circles) for all combinations of length and frequency using median and top skills. For legibility purposes, prediction intervals are not shown.

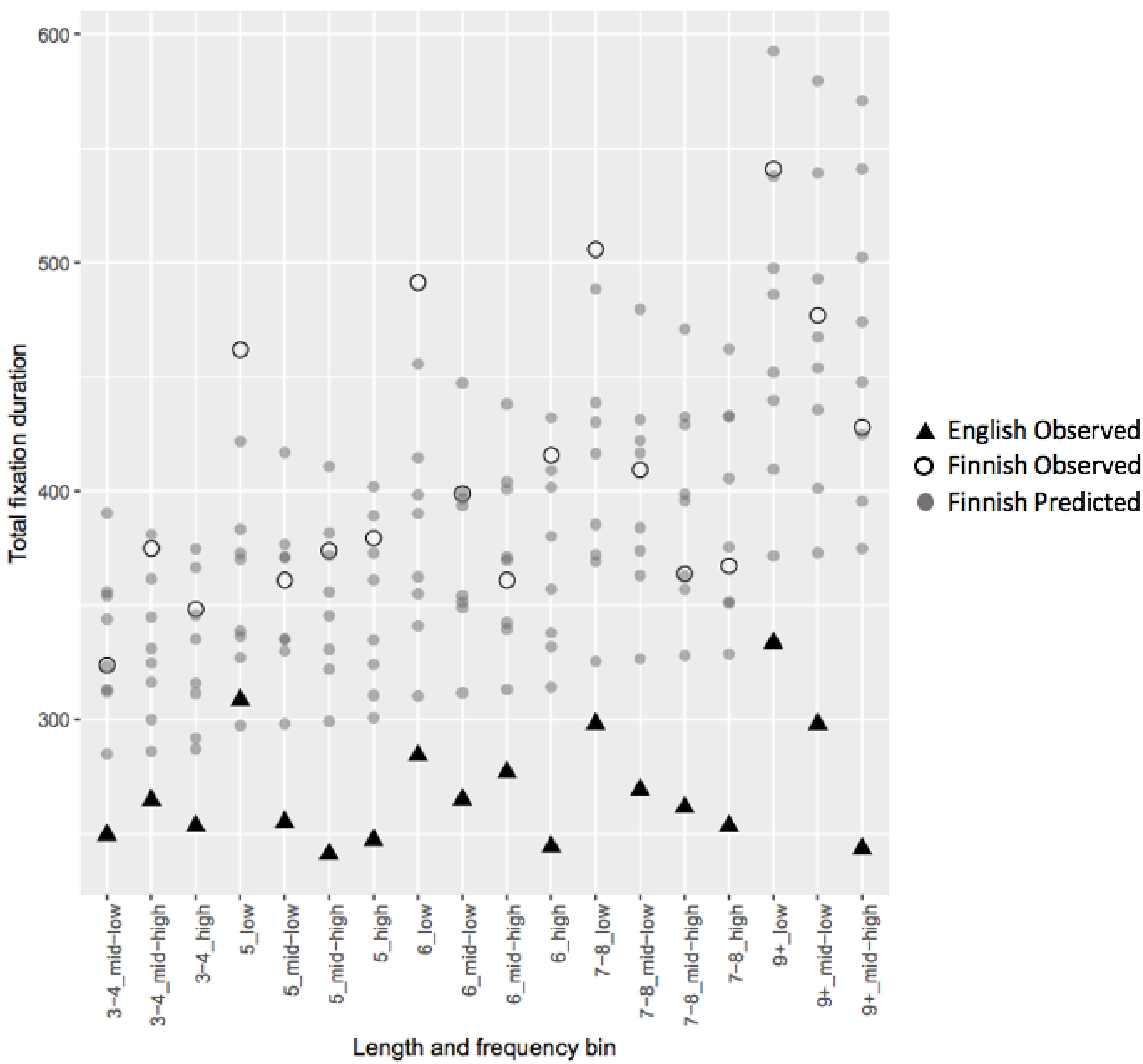
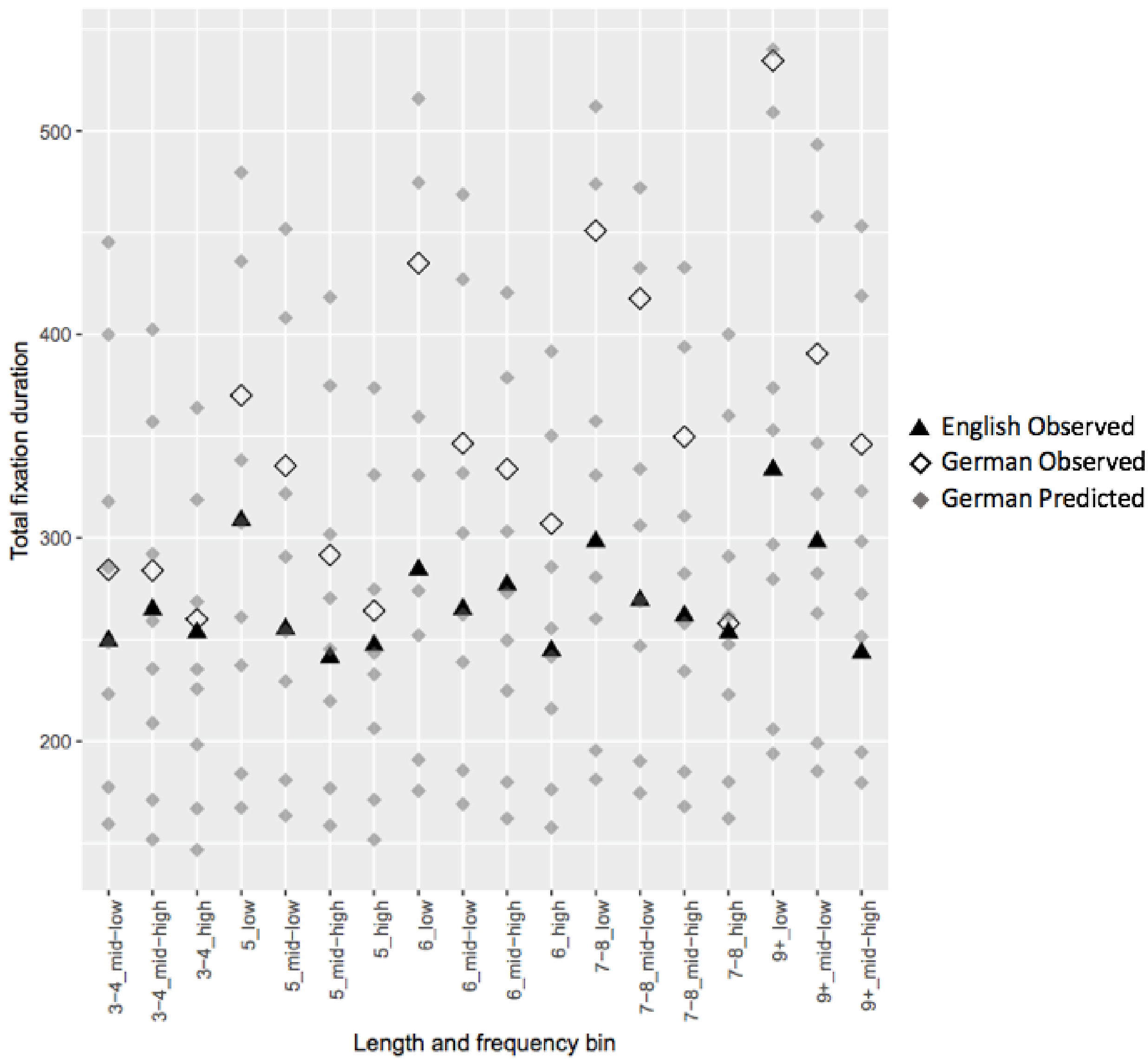


Figure 6 compares English to German speakers in which the general outline and English data are identical to the previous figure. German observed data is shown as open diamonds and predicted data as grey diamonds in the background. We see in this comparison that in most word-frequency bins, hypothetical German readers were either on par or exceeded the median reading speed of observed L1 readers.

Figure 6. Observed median total fixation times for German (black open diamonds), and English (filled black triangles), and predicted total fixation times for hypothetical German (small grey diamonds) for all combinations of length and frequency using median and top skills. For legibility purposes, prediction intervals are not shown.



Figures 5 and 6 show that both Finnish and German readers have longer median total fixation times than their Canadian counterparts in all combinations of frequency and length.

However, hypothetical German readers with maximum L1 skills show total fixation times much shorter than those observed in a median L1 speaker and approach the fastest L1 speaker. That is, in this subset of words, full mastery of component skills would lead German readers to faster recognition times than those observed in L1 readers. In all bins, the hypothetical German readers showed a wider range of performance as compared to the Finnish readers. The advantage of maximum component skills is more moderate in the hypothetical Finnish readers who only occasionally approach the median L1 performance when given top skills. We address potential reasons for the discrepancy between hypothetical German and Finnish readers in the General Discussion.

Cognates

It is possible that the advantage the German cohort had over the Finnish cohort is partly explained by a higher prevalence of cognates in the German-English language pair, than the Finnish-English one. Cognates are expected to be more common in genetically and typologically related languages. For example, a cognate among all three languages in this data set is the English word ‘vanilla’, in German ‘vanille’ and Finnish ‘vanilla’. Additional examples of German and English cognates exist like the English word ‘knee’ which in German is ‘Knie’ (but ‘polvi’ in Finnish). L1 German speakers examined our stimuli (316 unique word types, 669 word tokens) and identified 156 cognate word types (268 word tokens) between English and German (49% of total word types, 40% of word tokens). L1 speakers of Finnish examined the stimuli and found 44 cognate word types and tokens between English and Finnish (14% of word types, 7% of word tokens). To examine whether the cognate status affected total fixation time on a word in L2 readers, we looked at pairwise comparisons of the English dataset separately with the Finnish and German cohorts. In both cases, log-transformed ToFD was a dependent variable predicted by whether a given English word is a cognate in the readers’ L1 or not. Scores in the skill tests, as well as word frequency and length were included as controls. The results suggested that cognates did not have an independent effect on reading English for either German ($b = 0.003$, $SE = 0.019$, $t = 0.193$, $p = 0.846$), or Finnish ($b = -0.021$, $SE = 0.030$, $t = -0.690$, $p = 0.491$) participants. We replicated this null effect in analyses of gaze duration and first fixation duration also (not shown). We conclude that a higher prevalence of cognates cannot explain away the speed advantage that German readers have over Finnish readers in the sentence reading task. Our null effect is in line with the mixed evidence regarding the magnitude of the cognate effect in late eye-movement measures reported in eye-tracking studies of sentence and passage reading (see the Introduction).

Which component skills are important?

Regression models with either reading rate or ToFD as dependent variables (Tables 4-5, Supplementary Materials Tables S5, S6) estimate contributions of specific component skills to reading fluency in the L2 readers of English. Judging by inferential estimates, the critical skill for Finnish readers was English spelling while vocabulary size was a critical skill for German readers. These predictors were significant or marginally significant in respective models. While ART and vocabulary had significant or marginally significant main effects on total fixation times in the German L2 cohort (ART $p = 0.0476$, vocabulary $p = 0.0526$), spelling is the only component skill that shows a reliable impact ($p = 0.0407$) on total fixation times in the Finnish cohort (See Supplementary Materials Tables S5 and S6 respectively).

Because inferential estimates may be skewed due to collinearity between scores in component skills, we took an additional approach to highlight the relative importance of skills in L2 cohorts. From our statistical prediction method of total fixation duration for each language (shown visually in Figures 5 & 6), we took the combinations of component skills scores that led to the fastest total fixation times and analyzed them to look for patterns. Specifically, for the German cohort we extracted combinations of test scores that came with predicted total fixation times which were shorter than total fixation time for the fastest observed L1 reader (there were 12 combinations that yielded these results shown in Table 6). For the Finnish cohort, we extracted combinations of test scores that led to total fixation times shorter than the 40th percentile in the distribution of total fixation times for observed English readers (we chose this arbitrary point due to the fact that there were no Finnish speakers that surpassed an observed L1 reader, as you can see from Figure 6). There were 11 combinations that led to a predicted Finnish data point above the 40th percentile (shown in Table 7).

Table 6. Top predicted combinations for German hypothetical readers. Possible values for ART (median 7.5, high 41), Spelling (median 77, high 87), and vocabulary (median 72.5, high 100).

	Length Bin	Frequency Bin	log(ToFD)	Exposure	Spelling	Vocabulary
Prediction 1	3-4	High	4.9879	41	77	100
Prediction 2	3-4	High	5.1176	41	87	100
Prediction 3	5	High	5.0218	41	77	100
Prediction 4	5	Mid-high	5.0661	41	77	100
Prediction 5	5	Mid-low	5.0965	41	77	100
Prediction 6	7-8	High	5.0882	41	77	100
Prediction 7	7-8	Mid-high	5.1239	41	77	100
Prediction 8	7-8	Mid-high	5.2203	41	87	100
Prediction 9	9+	Low	5.2674	41	77	100
Prediction 10	9+	Low	5.3273	41	87	100
Prediction 11	9+	Mid-low	5.2222	41	77	100
Prediction 12	9+	Mid-low	5.2940	41	87	100

Table 7. Top predicted combinations for Finnish hypothetical readers. Possible values for ART (median 7.5, high 41), Spelling (median 77, high 87), and Vocabulary (median 72.5, high 100).

	Length Bin	Frequency Bin	log(ToFD)	Exposure	Spelling	Vocabulary
Prediction 1	3-4	High	5.6601	7.5	87	100
Prediction 2	3-4	High	5.6762	7.5	87	72.5
Prediction 3	3-4	Mid-low	5.6562	7.5	87	100
Prediction 4	5	High	5.7065	7.5	87	100
Prediction 5	5	High	5.7388	7.5	87	72.5
Prediction 6	5	Low	5.6951	7.5	87	100
Prediction 7	5	Low	5.7916	41	87	100
Prediction 8	5	Low	5.8185	7.5	87	72.5
Prediction 9	5	Low	5.8258	7.5	77	100
Prediction 10	5	Mid-low	5.6978	7.5	87	100
Prediction 11	9+	Low	5.9179	7.5	87	100

These analyses demonstrated that the optimal performance in hypothetical German readers of English is associated with maximum scores in vocabulary size (100) and ART (41). However, the spelling score was not a reliable predictor of fast reading speeds and fluctuated between the median and maximum. The situation was reverse in the Finnish cohort, where the fastest performance almost always came with the maximum spelling score (87; only one model used the lower score), but vocabulary and ART scores varied. The stronger contributions of vocabulary size and ART in German readers and of spelling ability in Finnish readers align with the results of the regression models (See Supplementary Materials Tables S5 and S6). These findings point to the different origins of deficits in reading speed for the German and Finnish cohorts. Equally, they suggest that highly-proficient German learners of English would benefit most from acquiring more vocabulary and increasing the volume of their reading in English. These measures, however, are not expected to benefit the Finnish learners of English that much: for this cohort, the strongest boost will come from enhancing their spelling ability. These conclusions are of course specific to the selection of tests implemented in this study and the cohorts of participants.

General Discussion

The main goals of this study were methodological. We proposed the use of a statistical technique that estimates a theoretical maximum of L2 reading performance based on observed L1 and L2 reading data. An eye-tracking study compared highly fluent L2 speakers of English from Germany and Finland reading the same sentences and completing the same tests of component skills as L1 speakers of English from Canada.

Comparative analyses showed that indeed highly proficient L2 speakers do differ from native L1 speakers of English: they were slower in reading (as determined by their reading rate and total fixation times) and they obtained lower scores in English spelling and vocabulary tests. We also found differences between the L2 cohorts. German speakers were faster on average than the Finnish participants when reading English on both a sentence and word level, even though Finnish participants performed better in the individual component tests. Descriptive statistics revealed that German L2 readers of English were able to reach native-like proficiency when reading short words or highly frequent words, whereas Finnish speakers did not approach L1 proficiency in any scenario. As argued in the Introduction, the different contrasts between the Canadian cohort and the two L2 cohorts may be due to differential knowledge of English as L2 in the German and Finnish cohort, the different language distances that German and Finnish have relative to English, or a combination of these and additional factors.

Through the use of statistical prediction, we were able to resolve the question above (for the given set of reading materials, participant samples, and component skill tests). Specifically, we estimated the predicted performance of hypothetical German and Finnish readers that were assigned the native-like proficiency in L2 component skills. An ‘ideal’ German reader (i.e., a hypothetical reader with the maximum scores observed in the Canadian L1 cohort on all tests) showed reading rate and total fixation times on individual words that were on par or exceeded those shown by a median observed L1 reader of English (reaching the 67^{th} percentile in reading rate). In some bins formed by word length and frequency, an ideal German reader would even exceed the observed maximum of the distribution of total fixation times in L1 readers. This suggests that the entire contrast in reading fluency between German and Canadian readers of

English can be cancelled out in principle if the former cohort improves its knowledge of L2 component skills. This signifies an advantage for L2 reading speed associated with the distance between that cohort's L1 (German) and L2 (English). A further analysis revealed that – out of the tests included in the present battery – the strongest boost to the L2 reading fluency of German readers of English is expected to come from an increased vocabulary size and greater exposure to print.

Given the same maximum affordances as the ideal German reader, an ideal Finnish reader is predicted to develop fluency at the 13th percentile of the L1 reading rate and, in select bins formed by word length and frequency, the 40th-50th percentile of the L1 total fixation time. The difference between this performance and that shown by the median L1 reader is likely to be related to the linguistic distance between L1 and L2 English, and in the present data cannot be further reduced by higher scores in the component skills of English reading comprehension.

Analyses also showed that English spelling skill is most beneficial for increased reading fluency in Finnish readers. In general, the difference in L2 English reading fluency of German and Finnish readers of English aligns well with the notion that the L1-L2 language distance is a causal factor of success in second language acquisition (see the Introduction). In our triad of languages, German and English are unquestionably closer to one another than Finnish and English. The former pair is related both historically and typologically, and shares a large number of phonological, orthographic, morphological, and syntactic features (Haspelmath, Dryer, Gil, & Comrie, 2005), and has a large overlap in respective word-stocks (Schepens, Dijkstra, & Grootjen, 2012). Conversely, Finnish and English belong to different language families (Finno-Ugric vs Indo-European) and represent radically different organizational principles in their orthography, phonology, morphology and syntax; beyond several loanwords, there is little lexical overlap between the languages either. An apparent important role of English spelling skill for Finnish L2 readers of English may be traced back to the fact that Finnish is an orthographically shallow language with near one-to-one correspondence between sounds and letters of words, while sound-letter correspondences are relatively opaque in written English. Learning to read Finnish does not require development of memorization which typically arises when learning to decode a complex and deep orthographic system like English (Aro, 2017). This may lead to increased word recognition and reading effort when transitioning to a language like English. While German is more transparent orthographically than English (see Seymour, Aro, & Erskine, 2003), the similarities in lexical, phonological and orthographic systems of these two languages allow German readers to overcome the same problem and further benefit from accumulation of other skills.

Limitations and future directions

This study is a proof of concept and has a number of limitations that need to be rectified in future research. First, these results need to be validated against better reading materials (e.g., continuous texts rather than short sentences), larger samples of participants representing a larger number of languages, and more comprehensive and balanced batteries of component skills (including, among others, listening comprehension and decoding tasks). For example, an accurate test of exposure to print that is proven to work reliably for both L1 and L2 speakers of English would be an asset. Our research shows that the ART does not reach this goal (McCarron & Kuperman, 2020). Therefore, some consideration is necessary when interpreting these scores across language backgrounds. Future work should look to improve the current biases towards the

Western and English-language literature and focus on finding the best tests for both L1 and L2 speakers.

Also, we concentrated our examination on L2 knowledge, which is the strongest but not the only causal factor in second language acquisition. Future batteries need to include measurements of L1 proficiency and motivational factors.

Second, our method assumes that reading speed is *linearly* related to component skills across the entire range of those skills. To illustrate this point in a simplified way, our method assumes that the difference between average reading rates observed for spelling scores of 5 and 10 is the same as the difference in reading rate associated with unobserved spelling scores of 55 and 60. At present, we are not aware of evidence against this assumption, but it may need to be tested separately. Third, we have only considered the impact of a specific L1-L2 distance to be expressed in reading fluency, that is, we only considered a direct impact of the distance on these outcome variables, over and above the influence of component skills. Undoubtedly, the L1-L2 distance also affects reading accuracy as well as the development of component skills per se, and thus has an additional mediated impact on reading fluency (Naghdipour, 2015; Chung, Chen, & Geva, 2019). That is, when indicating that an improvement in a given skill would benefit Finnish readers more than it would German readers of English we gloss over the possibility that specific differences between these L1s and L2 could make such an improvement more difficult to attain for Finnish vs German readers. A mediation analysis would address this issue.

Fourth, in the present study we cannot separate influences of the L1 linguistic background from the educational, cultural, and social aspects of living in a country with this dominant L1. Decoupling language from country of residence (e.g., studying German of Germany, Austria and Switzerland) would remedy this conflation. Finally, we call for consideration of a much more diverse selection of languages and scripts that would represent a broad range of L1-L2 distances. This will make it possible to quantify a correlation between processing penalties established with our method and independent estimates of their distances from English derived from research in linguistic typology.

In sum, the method that we presented here has the ability to disentangle direct contributions of L2 knowledge (measured as component skills) and the L1-L2 distance to reading fluency in L2. It can also point to cross-linguistic differences in L2 reading performance stemming from specific L1 backgrounds and help attribute it to specific parameters of those linguistic systems. Importantly, the method highlights the skills that are particularly beneficial for developing optimal L2 reading fluency in readers of a given L1 background. These results may have applied value for teaching foreign languages because they enable structuring educational curricula in the way that focuses on skills critical for a specific L1 and thus optimizes learning gains.

References

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. https://doi.org/10.3758/BRM.40.1.278

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*(2), 299-318. https://doi.org/10.1037/a0018366

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*(3), 279-296. https://doi.org/10.1016/j.jml.2012.12.001

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior Research Methods,* 52:2257-2282 . https://doi.org/10.3758/s13428-020-01387-3

Arkoudis, S., Hawthorne, L., Baik, C., Hawthorne, G., O'Loughlin, K., Leach, D., & Bexley, E. (2009). The Impact of English Language Proficiency and Workplace Readiness on the Employment Outcomes of Tertiary International Students (Full Report., 157 pp.). Melbourne: Centre for the Study of Higher Education, University of Melbourne.

Arkoudis, S., Baik, C., Bexley, E., & Doughney, L. (2014). English Language Proficiency and Employability Framework for Australian Higher Education Institutions. *Melbourne: Centre for the Study of Higher Education*.

Aro, M. (2017). Learning to Read Finnish. In Ludo, V., & Perfetti, C. (Eds.) *Learning to read across languages and writing systems* (pp.416-436) Cambridge University Press.

Ashby, J., & Clifton Jr, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89-B100. https://doi.org/10.1016/j.cognition.2004.12.006

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Beglar, D. & Hunt., A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language* 26: 29–48.

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, *25*, 133-150. DOI:10.1017/S0267190505000073

Bernhardt, E.B. (2011). *Understanding advanced second-language reading*. New York: Routledge.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied linguistics*, *16*(1), 15-34. https://doi.org/10.1093/applin/16.1.15

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics*, *2*(1), 165-92. DOI: 10.1257/app.2.1.165

Blumenfeld, H. K., & Marian, V. (2005). Covert bilingual language activation through cognate word processing: An eye-tracking study. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 27(27), 286-291.

Brevik, L. M., Olsen, R. V., & Hellekjær, G. O. (2016). The complexity of second language reading: Investigating the L1-L2 relationship. *Reading in a Foreign Language, 28*(2), 161-182. DOI: 10125/66899

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–90. https://doi.org/10.3758/BRM.41.4.977

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1214-1241. https://doi.org/10.1080/17470218.2013.853090

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(3), 209. https://doi.org/10.1037/h0087476

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*, *50*, 149-161. https://doi.org/10.1016/j.jneuroling.2018.01.003

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283-1296. https://doi.org/10.1037/0278-7393.26.5.1283

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, *20*(4), 747-769. doi:10.1017/S1366728916000213

de Groot, A. M. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, *56*(3), 463-506. https://doi.org/10.1111/j.1467-9922.2006.00374.x

Doughty, C. J., & Long, M. H. (Eds.). (2008). *The handbook of second language acquisition* (Vol. 27). John Wiley & Sons.

Drieghe, D. (2008). Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review*, *15*(4), 856-860. DOI: 10.3758/PBR.15.4.856

Eurostat Statistics Explained, Education and training. (2008). Foreign language learning statistics. Retrieved from https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Foreign_language_learning_statistics&oldid=454846.

Fraser, C. A. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal*, *91*(3), 372-394. https://doi.org/10.1111/j.1540-4781.2007.00587.x

Geva, E., & Wang, M. (2001). The development of basic reading skills in children: A cross-language perspective. *Annual Review of Applied Linguistics*, *21*(1), 182. DOI:10.1017/S0267190501000113

Grabe, W. (1991). Current developments in second language reading research. *TESOL quarterly*, *25*(3), 375-406. https://doi.org/10.2307/3586977

Grabe,W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL quarterly*, *21*(3), 505-521. https://doi.org/10.2307/3586500

Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2005). *The world atlas of language structures*. Oxford: Oxford University Press.
Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. Mit Press.
Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160-212. https://doi.org/10.1111/lang.12034
Koda, K. (1996). L2 word recognition research: A critical review. *Modern Language Journal*, *80*, 450–460. https://doi.org/10.1111/j.1540-4781.1996.tb05465.x
Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. New York: Cambridge University Press.
Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of memory and language*, *65*(1), 42-73. https://doi.org/10.1016/j.jml.2011.03.002
Kuperman, V., Siegelman, N., Schroeder, S., Alexeeva, A., Acartürk, C., Amenta, S., Bertram, S., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Marelli, M., Nisbet, K., Papadopoulos, T. C., Protopapas, A., Savo, S., Shalom, D., Slioussar, N., Stein, R., Sui, L., Taboh, A., Tønnesen, V., and Usual, K. A. (2020). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus (MECO). *Manuscript submitted to' Studies in Second Language Acquisition'*.
Lambert, O. (2008). Who are our students? Measuring learner characteristics in adult immigrants studying English. *Adult Basic Education & Literacy Journal*, *2*(3), 162-173.
Lee, J. W., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *Tesol Quarterly*, *31*(4), 713-739. https://doi.org/10.2307/3587757
Leinenger, M., & Rayner, K. (2017). *What we know about skilled, beginning, and older readers from monitoring their eye movements.* In Leon, J. A., & Escudero, I. (Eds.) Reading Comprehension in Educational Settings, 1-28. John Benjamins Publishing Company.
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*(2), 325-343. https://doi.org/10.3758/s13428-011-0146-0
Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533-550. https://doi.org/10.3758/BF03195845
Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381. https://doi.org/10.1037/a0014875
Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System, 25*(1), 91-102. https://doi.org/10.1016/S0346-251X(96)00063-2
McCarron, S., & Kuperman, V. (2020). Is the Author Recognition Test a Useful Metric for Native and Non-Native English Speakers? An Item Response Theory Analysis. *Behavioural Research Methods.* (Manuscript under revision)
McManus, W., Gould, W., & Welch, F. (1983). Earnings of Hispanic men: The role of English language proficiency. *Journal of Labor Economics, 1*(2), 101-130.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114-135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Naghdipour, B. (2015). The impact of L1 reading directionality mode on L2 reading fluency. *Journal of Asia TEFL*, *12*(1) 53-77.

Nuttall, C. (1996). *Teaching reading skills in a foreign language*. Heinemann, 361 Hanover Street, Portsmouth, NH 03801-3912.

Olsen, L. (2000). Learning English and learning America: Immigrants in the center of a storm. *Theory into practice*, *39*(4), 196-202. https://doi.org/10.1207/s15430421tip3904_2

Pecorari, D. (2018). Teaching English to Speakers of Other Languaes (TESOL). In P. Seargeant, A. Hewings, & S. Pihlaja (Eds.), *The Routledge Handbook of English Language Studies* (pp. 199-211). London: Routledge.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

Rahmanian, S., & Kuperman, V. (2019). Spelling errors impede recognition of correctly spelled word forms. *Scientific Studies of Reading*, *23*(1), 24-36. https://doi.org/10.1080/10888438.2017.1359274

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements?. *Vision research*, *16*(8), 829-837. https://doi.org/10.1016/0042-6989(76)90143-7

Rayner, S. (1998). Educating Pupils with Emotional And Behaviour Difficulties: Pedagogy Is the Key!. *Emotional and Behavioural Difficulties*, *3*(2), 39-47. https://doi.org/10.1080/1363275980030206

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*, *62*(8), 1457-1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514. https://doi.org/10.1037/a0020990

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157-166. DOI: https://doi.org/10.1017/S1366728910000623

Schepens, J., Van der Slik, F., & Van Hout, R. (2013). The effect of linguistic distance across IndoEuropean mother tongues on learning Dutch as a second. In Borin, L., & Sazena, A. (Eds.) *Approaches to measuring linguistic differences*, *265*, (pp 199-245).

Schepens, J. J., van der Slik, F., & Van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Language Learning*, *66*(1), 224-256. https://doi.org/10.1111/lang.12150

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive processes*, *22*(1), 106-129. https://doi.org/10.1080/01690960500463920

Segalowitz, N. (2010). *The cognitive bases of second language fluency*. New York: Routledge.

Seymour, P. H., Aro, M., Erskine, J. M., & Collaboration with COST Action A8 Network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143-174. https://doi.org/10.1348/000712603321661859

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 402-433. https://doi.org/10.2307/747605

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.

Van der Slik, F. W. (2010). Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West European first languages. *Studies in Second Language Acquisition*, *32*(3), 401-432. Doi:10.1017/S0272263110000021

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic bulletin & review*, *9*(4), 780-789. https://doi.org/10.3758/BF03196335

Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney papers in TESOL*, *1*(1), 51-70.

Yamashita, J., & Shiotsu, T. (2017). Comprehension and knowledge components that predict L2 reading: A latent-trait approach. *Applied linguistics*, *38*(1), 43-67. https://doi.org/10.1093/applin/amu079

Supplementary Materials

Table S1. Fixed effects of the model fitted to exposure to print (ART) scores with German as the baseline.

	Estimate	SE	t	P
Intercept	8.545	1.298	6.585	< 0.001
English	2.910	1.822	1.202	0.232
Finnish	2.597	1.915	1.356	0.178

Table S2. Fixed effects of the model fitted to the spelling test scores with German as the baseline.

	Estimate	SE	t	P
Intercept	66.515	1.480	44.932	< 0.001
English	8.279	2.078	3.984	< 0.001
Finnish	3.628	2.185	1.660	0.100

Table S3. Fixed effects of the model fitted to vocabulary scores with German as the baseline.

	Estimate	SE	t	P
Intercept	72.462	2.109	34.361	< 0.001
Finnish	1.936	3.144	.0616	0.54

Table S4. Fixed effects of language on total fixation time with German as the reference language. By-subject random intercepts SD = 0.25, by-item random intercepts SD = 0.17, residual SD = 0.46, N = 26416.

	Estimate	SE	Df	t	P
Intercept	5.887	0.044	100.373	132.732	< 0.001
English	-0.265	0.061	92.012	-4.356	< 0.001
Finnish	0.149	0.064	91.771	2.332	0.22

Table S5. Regression model for ToFD: German cohort. SD of the by-items random intercepts is 0.0690; SD of the by-subject random intercepts is 0.2811; the model residual SD is 0.4388; N = 9259.

	Estimate	SE	Df	T	P
Intercept	6.852	0.5809	29.53	11.795	< 0.001
Length	0.0712	0.0335	29.98	2.120	0.0426
Frequency	-0.1389	0.0340	29.46	-4.087	0.0003
Spelling	0.0091	0.0111	28.91	0.821	0.4174
Exposure	-0.0245	0.0119	28.79	-2.064	0.0476
Vocabulary	-0.0143	0.0071	29.34	-2.016	0.0526
Length: Spelling	-0.0006	0.0006	28.65	-0.940	0.3553
Length: Exposure	0.0002	0.0007	29.05	0.306	0.7622
Length: Vocabulary	-0.0000	0.0004	29.26	-0.201	0.8422
Frequency: Spelling	0.0006	0.0006	28.27	0.902	0.3745
Frequency: Exposure	0.0009	0.0007	27.75	1.328	0.1951
Frequency: Vocabulary	0.0003	0.0004	29.08	0.825	0.4160

Table S6. Regression model for ToFD: Finnish cohort. SD of the by-items random intercepts is 0.0164; SD of the by-subjects random intercepts is 0.1274; the model residual SD is 0.4521; N = 7881.

	Estimate	SE	Df	t	P
Intercept	8.3436	0.5804	23.1230	14.374	< 0.001
Length	-0.0006	0.0360	24.7265	-0.017	0.9866
Frequency	-0.1861	0.0355	24.4473	-5.241	< 0.001
Spelling	-0.0249	0.0115	22.6292	-2.171	0.0407
Exposure	-0.0035	0.0138	22.8492	-0.251	0.8041
Vocabulary	-0.0069	0.0102	22.4665	-0.673	0.5077
Length: Spelling	0.0009	0.0007	23.9914	1.289	0.2098
Length: Exposure	0.0006	0.0009	23.9905	0.697	0.5034
Length: Vocabulary	-0.0004	0.0006	23.7514	-0.604	0.5515
Frequency: Spelling	0.0013	0.0007	23.6149	1.796	0.0854
Frequency: Exposure	0.0006	0.0008	23.9950	0.696	0.4932
Frequency: Vocabulary	0.0007	0.0006	23.2413	1.197	0.2435

CHAPTER 4.

The Validity of the Link Between Statistical Learning and Reading Comprehension: Testing for Mediating Variables

Nisbet, K., Siegelman, N., & Kuperman, V. (2021). The Validity of the Link Between Statistical Learning and Reading Comprehension: Testing for Mediating Variables. *Submitted March 2021 to Cognitive Science*.

Abstract

Reading is a complicated skill and many variables have been found to determine an individual's proficiency in this area. One such variable that has gained traction recently is statistical learning (SL). Recent SL views of reading tie successful text processing to sensitivity to statistical regularities embedded in writing systems, and indeed multiple studies have found positive correlations between SL tasks performance and reading outcomes. Still, there are two outstanding questions in regards to the validity of the SL-reading connection. First, it is possible that SL is only related to specific sub-components of reading. Second, it is possible that the link between SL and reading is mediated by a theoretically-irrelevant third variable. To address these questions, this study examines whether specific components of reading and the more general cognitive factor, motivation, mediate the relation between SL and reading comprehension. We conducted an online study (n=188) containing tests of visual SL, reading comprehension, motivation, spelling ability, and exposure to print. We found that a significant portion of the shared variance between SL and reading is mediated by other factors, including in particular participants' motivation. However, there exists a direct link between SL and reading comprehension, above and beyond all other tested factors. These results demonstrate the complex links between SL and other important factors known to influence an individual's reading proficiency. They also suggest that visual SL is not only a reliable measure of performance but is also a valid predictor of reading skill in and of itself.

Introduction

Reading is a complex, multi-componential process that requires simultaneous proficiency in numerous skills such as oral comprehension, decoding, vocabulary, and orthographic, phonological, and morphological awareness to name just a few (Oakhill, Cain, & Bryant, 2003; Perfetti, Landi, & Oakhill, 2005; Share, 1995; and Vellutino, Tunmer, Jaccard, & Chen, 2007). In addition, more domain general cognitive abilities, including memory and attention, have been found to impact reading ability (Cain, Oakhill, & Bryant, 2004; Cutting & Scarborough, 2006; Fedorenko, 2018). This list of variables that demonstrably impact reading is continuously growing. One such variable that has recently gained traction in reading research is statistical learning (SL).

SL refers to the brain's ability to extract regularities from the environment and use re-occurring patterns to process and predict upcoming information. Our strong capacity for SL as humans has been illustrated across different modalities (visual, auditory and tactile), types of patterns (distributional and transitional), throughout the lifespan (from neonates to adults) and across multiple measures (Abla & Okanoya, 2009; Bulf et al., 2011; Campbell et al., 2012; Conway & Christiansen, 2005; Kirkham et al., 2002; Koelsch et al., 2016; Maye et al., 2002; Neger et al., 2014; Saffran, Newport, & Aslin, 1996; Teinonen et al., 2009). Therefore, SL is argued to play a role in many cognitive skills and abilities including segmenting auditory input, contextual cuing, visual search, and conditioning, among others (Baker et al., 2004; Courville et al., 2006; Goujon & Fagot, 2013; Saffran, Aslin, & Newport, 1996; also see review by Frost et al., 2015).

In the current work we focus on the link between SL and reading comprehension. Since SL consists of detecting patterns or regularities in sensory input, it stands to reason that this ability will be of particular utility in various aspects of language acquisition and use. This is because statistical regularities of varying complexity exist at all levels of language. Co-occurrences of letters, phonemes, morphemes, or syntactic elements in language productions are neither deterministic nor random but are best described as probabilistic or quasi-regular (Brand et al., 2009; Chetail, 2017; Thompson & Newport, 2007; Treiman & Kessler, 2006). Detecting and learning these probabilistic patterns is an essential part of proficient language production and comprehension (Erickson & Thiessen, 2015; Romberg & Saffran, 2010; Siegelman, 2020).

Individual Differences in SL and Language

Support for the theorized link between SL and language has been demonstrated in studies tying individual differences in SL ability and various linguistic (dis)abilities. Thus, positive correlations with SL have been shown in multiple aspects of linguistic performance including syntactic processing, lexical knowledge, and vocabulary size (Conway et al., 2007; Lany et al., 2018; Misyak, & Christiansen, 2012; Misyak et al., 2010; Shafto et al., 2012; Singh et al., 2012; Spencer et al., 2014). In the same vein, there have been several reports of SL deficits in populations with different forms of language impairments (for reviews and meta-analysis, see, e.g., Bogaerts et al., 2020; Lammertink et al., 2017; Saffran, 2018). The range of linguistic domains showing an association with SL demonstrates the depth of the connection between SL and language acquisition and use.

This study targets a specific area of language: reading comprehension. Reading is an interesting test-case for SL theories, since being proficient in reading requires mastery of multiple levels of language knowledge, many of which require learning of complex statistical patterns (for reviews, see Arciuli, 2018; Sawi & Rueckl, 2019). Indeed, a few individual-differences studies have found correlations between SL and measures of reading skill (Arciuli & Simpson, 2012; Conway et al., 2010; Frost et al., 2013; Hung et al., 2018). For instance, Arciuli and Simpson (2012) found that performance on a visual SL task was significantly correlated with reading ability in both children and adults. It has also been shown that SL ability can predict reading ability for children in their next year of education (Van de Kleij et al., 2019). Frost, Siegelman, Narkiss, and Afek (2013) found that English L1 speakers learning Hebrew as L2 with better visual SL task performance were better at assimilating to the structure of the new writing system as measured through multiple reading measures. Another large study looked at the relationship between SL and a battery of literacy-related skills (oral language, vocabulary knowledge, and phonological processing) in children, and found that a unique portion of the variance in these critical skills was accounted for by SL (Spencer et al., 2014). Finally, an additional important aspect of reading fluency is being able to predict what word will come next, and Conway et al. (2012) found a significant correlation between adults' SL ability and their performance on a word predictability task. At the same time, and despite these multiple positive findings, not all studies have found a significant correlation (see Schmalz et al., 2019). Additional research is needed to fully understand whether and how SL impacts reading ability. This study looks further into the nature of this relationship.

Potential Underlying Factors: Getting the Whole Picture

Evidence suggests a positive relationship of SL with language abilities. One lingering question that has yet to be answered is how *valid* this hypothesized relationship is. Specifically, this paper addresses validity of the relationship between visual SL and reading comprehension. There are at least two reasons why such an investigation is necessary, one theoretical and one methodological. Theoretically, both SL and reading comprehension are multi-componential skills. Thus, it is possible that the impact of SL on reading comprehension is indirect, i.e., mediated by one or more component skills of reading. Prior reports of correlations between individual variability in SL and reading-related skills (e.g., Lany et al., 2018; Misyak et al., 2010; Singh et al., 2012; Spencer et al., 2014) support this possibility. A methodological concern is that correlations between visual SL and reading comprehension may be inflated or spurious because of a factor which is not immediately related to either cognitive domain but stems from individual differences in the level of engagement with experimental tasks, i.e., individual motivation to perform well in an experiment. More motivated individuals will likely show higher scores in both SL and reading comprehension than the less engaged ones, leading to correlations that are not primarily indicative of underlying cognitive processes.

Both potential threats to validity of the hypothesized SL-reading connection imply that there is a "third" variable (or multiple variables) that mediates the relationship and might be leading us to misinterpret the nature or magnitude of the connection. This study examines the nature of the relationship between individual differences in SL and reading comprehension while accounting for a potential mediating role of select variables of influence. We provide details on potential "third variables" below.

Our central interest is in the role of *motivation*. This term can refer to several different related concepts, but for this project, we are simply referring to an individual's investment in the current study, i.e., their motivation to do well in the current experiment. This contrasts with a person's motivation to read or do well in tasks or situations specifically related to reading (Guthrie & Coddington, 2009; Wigfield et al., 2016). The motivation to engage in the experiment is important when studying any psychological phenomena, as there is always the concern that the higher performance of participants who want to do well is more heavily reliant on their motivation rather than their abilities themselves (McFall et al., 2009; Seitchik et al., 2017; Van Iddekinge et al., 2018). This is particularly difficult to disentangle with tasks that are challenging and require sustained attention for long periods of time. Typical SL tasks are not particularly entertaining (e.g., seeing sequences of symbols) and require participants to stay engaged throughout the familiarization phase (a duration of 5-20 minutes). When we see a positive correlation between an SL task and a reading task outcome, it may reflect a direct relation between abilities, or the participants' motivation. This study asks whether a participant's motivation towards the experiment may be a "third variable", a mediating factor in the relationship between SL and reading comprehension.

Given the nature of the relationship between SL and reading as it is known thus far, it is also reasonable to suggest that SL may affect either only some components of reading ability, or all but to differing degrees. Reading is contingent on a visual sensory ability, therefore we chose a SL task that also relies on the visual domain. While there are many different component skills related to reading that would warrant investigation, given the online nature of our experiment and the type of SL task (visual) we chose to implement, we consider two component skills of reading as potential mediators of SL and reading comprehension: individual spelling ability and exposure to text. These tests were chosen given their strong connections to overall reading ability, their presumed relevance to SL, and their ability to be implemented online. Thus, in terms of their link to reading, it is thought that individuals who demonstrate higher spelling ability also show high reading ability (Andrews, 2012; Andrews, 2015; Vellutino et al., 2007; Caravolas et al., 2001), and that more frequent exposure to print is also highly correlated with an individuals' reading proficiency (Mol & Bus, 2011; Moore & Gordon, 2015; Stanovich & West, 1989). In terms of their link to SL, spelling abilities are considered to involve sensitivity to multiple types of regularities, including orthotactic patterns and correspondences between sounds and printed units (Treiman & Kessler, 2006; Treiman et al. 2018), and more exposure to print is expected to provide more opportunities for SL to occur. Therefore, both spelling and exposure to print are important factors to consider as potential mediators of the link between SL and reading comprehension.

The present study looks to further investigate the complex relationship between SL ability and reading ability. The unquestioned assumption of most prior work in this area is that this relationship is direct: people with a greater SL skill are better readers. Yet SL may show its influence on reading comprehension because it masks the effect of a "third variable" or perhaps because it affects some component skills of reading which in turn influence reading comprehension. In a series of mediation analyses, we determine the magnitude and directionality of relationships between SL, reading comprehension, motivation, and component skills of reading in English.

Methods

Participants

Participants were 188 English L1 university students from McMaster University in Canada (Male = 18, Other = 1), recruited from a convenience pool. The sample of participants spanned from first year to fifth year undergraduate students with an age range from 17 – 34 (mean age = 20). Participants were compensated for their time with partial course credit. This study received appropriate ethics clearance from the McMaster University research ethics board (REB 2018-033).

Materials

Statistical Learning: This study used a self-paced Visual Statistical Learning (VSL) task (Siegelman et al., 2018) that includes both "online" (i.e., collected during learning) and "offline" measures of SL ability. This task employs the typical SL procedure with a familiarization phase followed by a test phase. This familiarization phase consisted of eight triplets (consisting of 24 complex visual shapes) with a transitional probability (TP) of 1 between shapes within triplets (e.g., within triplets, shapes are shown in the same order). These shapes were seen one after the other in the middle of the screen over 24 blocks. Each block contained all 8 triplets in a random order. During familiarization, participants determined the speed at which they saw each shape (i.e., self-paced presentation). In this task, the amount of time the participant takes to advance to the next shape is expected to be impacted by the structure of the familiarization stream (Siegelman et al., 2018). Thus, predictable shapes (in the 2nd and 3rd position within triplets) are responded to faster than unpredictable shapes (in the 1st position). Following this familiarization phase, participants took a 2-alternative-forced-choice offline test consisting of 32 trials. Each trial presented participants with 2 choices of 3-item sequences of shapes. One was a triplet seen in the familiarization phase (i.e., within-triplet TP = 1), and one was a foil consisting of familiar shapes combined in a sequence that never appeared in the familiarization phase (i.e., TP=0). Note that shapes in foils did not violate the internal order within original triplets (e.g., from the triplets ABC, DEF, and GHI, a possible foil was AEI but not AFG). During the test phase, trials were presented shape by shape at a fixed rate of 800 ms with an inter-stimulus interval of 200 ms between shapes within triplets and a blank of 1000 ms between the two options. Each of the triplets appeared 4 times throughout the test with 4 different foils (which were also seen a total of 4 times each). VSL performance was calculated both for online learning and offline learning. The online learning score was calculated by taking the difference in log-transformed reaction time (RT) between the 1st location and the mean of the 2nd and 3rd location in blocks 7 to 24 (where evidence for stable online learning is expected, see Siegelman et al., 2018). The offline test scores ranged from 0 to 32 according to the number of correct identification of targets over foils in the 2-alternative-forced-choice test. This number was then converted to a percentage of correct responses.

Reading Comprehension: To assess reading comprehension, we included a portion of the Grey Oral Reading Test (GORT-4) developed by Wiederholt & Bryant (2001). We specifically selected passages 5 – 12 (8 in total) that are geared towards more advanced readers. Each passage increased in difficulty and was on average 5-11 sentences long followed by five multiple choice questions to assess comprehension of the content. All participants read every passage and answered the questions that followed. The reading comprehension score was calculated by summing up the number of correct responses across all the passages, with a highest attainable score of 40.

Motivation: To determine the motivation our participants had for completing our series of online tests, we used the Student Opinion Scale (SOS) (Thelk et al., 2009). This questionnaire had 10 statements about their motivation where participants responded from ‘1=Strongly Disagree’ to ‘5=Strongly Agree’ with how they feel about each of them in relation to completing the current study (Appendix A). Four items were reverse coded, and therefore flipped prior to scoring. All 10 items were then averaged to get a final motivation score with a maximum of 5.

Exposure to Print: The Author Recognition Test (ART) by Acheson, Wells, and MacDonald (2008) was used to determine the participants' exposure to print. This test contained 130 items (half real authors, half fake names) presented as a checklist of names, and participants were asked to indicate which of the names were published authors while leaving those names that were not. The underlying assumption is that individuals with higher scores know a larger number of authors' names from reading more broadly, even if they did not read those authors. Individual ART scores were calculated by giving 1 point for every correct indication of an author and deducting 1 point for every false indication of a fake name. No points were deducted for missing a real author name. Chance performance on this test would be a score of '0' and negative scores are possible. For predictive validity and reliability of the ART see Moore and Gordon (2015).

Spelling Ability: To measure spelling ability, we used a spelling recognition test adopted from Andrews and Hersch (2010); for reliability estimates see also Andrews, Veldre, & Clark, 2020). The original test consists of 88 English words, of which half are spelled correctly and half are not. In this shortened version of the test, 44 words were selected from the original test (half correct, half incorrect) and participants were asked to make this 'correct' or 'incorrect' judgement. The score given for spelling was calculated by taking the number of correct responses (1 point for saying 'correct' to a properly spelled word and 1 point for saying 'incorrect' for an incorrectly spelled word) from each of the 44 items.

Procedure

The data collection for this study took place entirely online. Participants began by reading important information regarding the online study and providing informed consent. A brief demographic survey was conducted about their year of study, age, sex, faculty, and language history. After the survey, participants completed the reading comprehension portion of the experiment. Participants next filled out a brief survey on reading habits as part of another study which will not be reported here. The ART was administered next followed by the spelling test. After each of the reading, ART, and spelling tests participants were presented with their score as well as the typical average score to see how they compare to their sample. Following this, the motivation survey was administered and finally, the VSL task was completed last. All participants were debriefed after the experiment and were provided contact information if they had any questions about their participation or the study. The entire online experiment took no longer than 1 hour to complete for all participants.

Statistical Considerations

We chose to use mediation analyses to answer some of our questions of interest. For variable 'Z' to mediate a relationship between variables 'X' and 'Y', it is necessary to examine all possible independent relationships among the variables of interest. Two models are required to conduct the mediation analysis (referred to as base models). The first is a statistically significant relationship between the independent variable X and the potential mediator Z, as there can be no mediation if the independent variable does not have an effect on the mediator. This relationship is tested in a regression model where X predicts Z. The second required model is a multiple regression analysis in which both X and Z predict Y. In addition to these requirements, a third model depicting a significant relationship between Y and Z is typically considered as well (although not required, see MacKinnon et al., 2007). After establishing statistically significant

relationships exist among our variables of interest, the mediation analysis can proceed. These analyses were conducted using the ‘mediation’ package in R and bootstrap procedures utilizing 10,000 simulations, as recommended by Hayes and Scharkow (2013) with the lower and upper boundaries of the 95% confidence interval computed by determining the effects at the 2.5th and 97.5th percentiles (Tingley et al., 2014). The mediation analysis creates a model that demonstrates how much of the effect on Y is explained by the direct pathway (from X to Y), and how much is explained by the indirect pathway (from X, through Z, to Y). It is possible for the outcome of the mediation analysis to reveal a relationship that is fully mediated (all of the variance is explained by the indirect pathway), partially mediated (both the direct and indirect pathways significantly account for a portion of the variance), or no mediation (if all of the variance is explained by the direct pathway). The nature of the relationship can be gleaned from the output of the mediation analysis. It has four components 1) the Average Causal Mediation Effect (ACME) – the indirect effect of X on Y that goes through Z and its significance, 2) the Average Direct Effect (ADE) – the direct effect of X on Y and its significance, 3) the Total Effect – the combined effect of both the direct and indirect pathways and its significance, and 4) the Proportion Mediated – the percentage of the total effect that goes through Z. The following analyses were performed using the statistical software R version 4.0.0 (2020). Primary packages used include ‘lme4’ v 1.1.23, ‘mediation’ v 4.5.0, and ‘laavan’ (Bates et al., 2015; Rosseel, 2012; Tingley et al., 2019). All datasets and R code used for our analyses are available from Open Science Framework (OSF) at this link:
https://osf.io/afxsp/?view_only=b936bafc7d2b40bc9383ae6860e01be3

Results

Descriptive Statistics

Table 1 reports descriptive statistics for reading comprehension scores, VSL performance (both offline and online measures), motivation, ART, and spelling.

Table 1
Descriptive statistics for variables of interest (n=188).

	Mean	Median	Standard Deviation	Range
Reading Comprehension	23.80	23.00	6.06	6.00 – 38.00
VSL Online (log RT diff.)	0.013	0.00	0.07	-0.11 – 0.76
VSL Offline (%)	0.66	0.66	0.16	0.31 – 1.00
Motivation	3.48	3.50	0.41	2.10 – 4.40
Author Recognition Test	10.70	9.50	7.45	-7.00 – 32.00
Spelling Test	38.20	39.00	3.99	24.00 – 44.00

Pearson’s correlations were computed among these 6 main variables of interest based on our sample of 188 participants (Table 2). We highlight the most informative correlational patterns: Unless indicated otherwise, all correlations reported in this paragraph were significant ($p < 0.001$). As expected, we found a positive correlation between reading comprehension and VSL offline scores. VSL offline was positively correlated with motivation and ART, but not with spelling ($p = .184$): this raises a possibility that motivation, ART, or both factors mediate the

VSL – reading comprehension relationship. This possibility is further supported by the positive correlations between reading comprehension, motivation, and ART.

The VSL online measure did not have any significant relationships with reading-related variables (all *p*'s > 0.1) and was positively and significantly correlated with the VSL offline measure, albeit this correlation was weaker ($r = .28$) than that reported in the original (in-lab) study with this paradigm (Siegelman et al., 2018). These results suggest that the online VSL measure is not performing in a similar manner in our web-based study as it does in a lab-based setting (Siegelman et al., 2018). Due to these initial findings, potentially owing to the lack of control over the environmental settings and equipment used by the participants (given the online nature of the study), we did not include the online measure in subsequent analyses and the offline measure was used as the sole metric of VSL performance.

Table 2
Correlations between the variables of interest (N = 188). Pearson Correlation Coefficients are reported below the diagonal and p-values above the diagonal. The strength of the correlation is shown in greyscale from weak (in white) to strong (in dark grey).

	Reading Comp.	VSL Offline	VSL Online	Motivation	ART	Spelling
Reading Comp.	--	**p< .001**	p= .685	**p< .001**	**p< .001**	**p< .001**
VSL Offline	0.27	--	**p< .001**	**p< .001**	**p< .001**	p= .184
VSL Online	0.03	0.28	--	p= .746	p= .217	p= .669
Motivation	0.29	0.25	0.02	--	**p< .001**	**p< .001**
ART	0.24	0.26	0.09	0.26	--	**p< .001**
Spelling	0.34	0.10	0.03	0.25	0.53	--

Overall, correlations reported in Table 2 point to two variables that satisfy statistical criteria for possible mediators: motivation scores as well as ART scores correlate significantly both with the VSL offline scores and reading comprehension. The next section presents mediation analyses that quantify the role of motivation and exposure to print – both separately and jointly – in the target relationship between VSL and reading comprehension.

Mediation Analysis: Motivation as Mediator

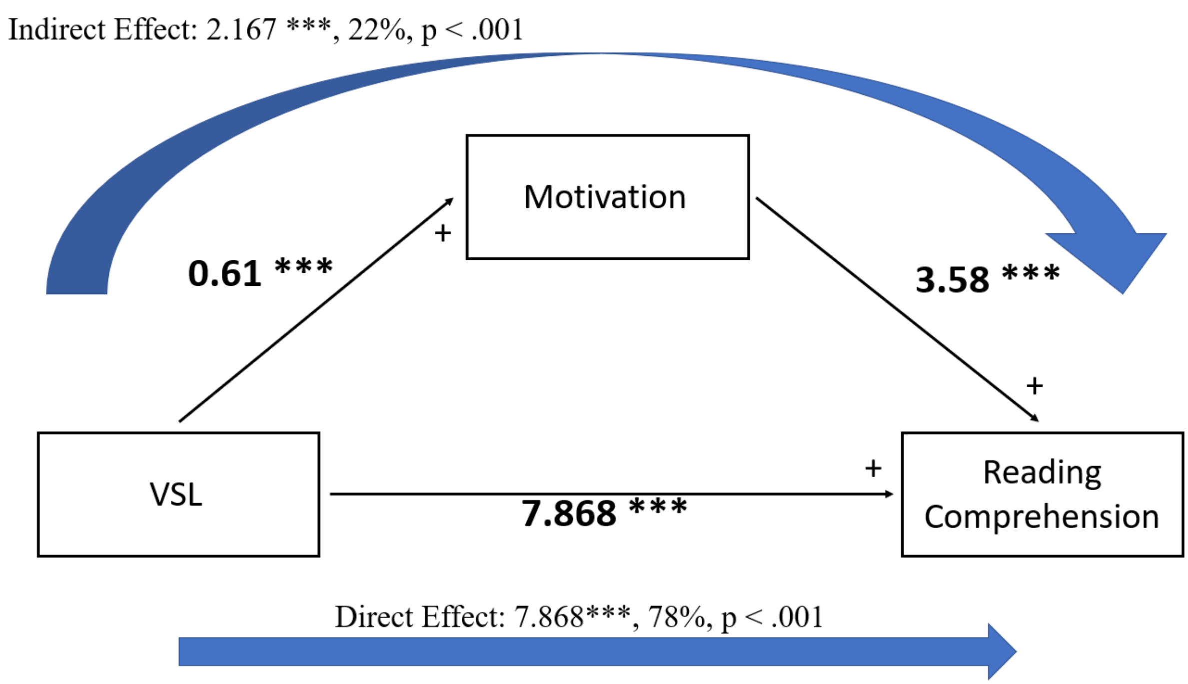
In this mediation analysis, reading comprehension was specified as the dependent variable with VSL score serving as the independent variable, and motivation score as the mediator. The base models required to satisfy the conditions of a mediation analysis were run (Appendix B1 and B2) and were then used to estimate direct and indirect pathways in the mediation analysis. The output is shown in Table 3 and Figure 1.

Table 3
Mediation analysis of the relationship between VSL and reading comprehension mediated by motivation (N = 188). The 'Average Causal Mediation Effect' (ACME), 'Average Direct Effect' (ADE), 'Total Effect', and the 'Proportion Mediated' are displayed. Nonparametric bootstrap confidence intervals with the percentile method simulated in 10000 iterations.

	Estimate	95% CI Lower	95% CI Upper	p value
ACME	2.167	0.671	4.390	< .001
ADE	7.868	2.790	12.720	< .001
Total Effect	10.036	5.007	14.900	< .001
Prop. Mediated	0.216	0.065	0.520	< .001

In Table 3 we see the indirect effect of VSL on reading comprehension that goes through the mediator motivation (ACME) to be statistically significant and account for an estimated 22% of the variance explained (95% CI: [6%, 51%]). The direct effect of VSL on reading comprehension (ADE) was found to be statistically significant and account for the majority of the total variance explained: 78% (95% CI: [49%, 94%]). In summary, while the effect of VSL on reading comprehension was partially mediated by motivation, the direct effect accounted for the majority of the total variance (see Figure 1).

Figure 1
Mediation analysis of the effect of VSL on reading comprehension with motivation as the mediator (N = 188, bootstrap simulations = 10000). The significance values for the individual connections in the triangle come from the estimates of each connection in the base models found in the Appendix (B1 and B2).



Mediation Analysis: ART as Mediator

In the next mediation analysis conducted, reading comprehension was specified as the dependent variable with VSL score as the independent variable like before, but this time ART served as the mediator. Linear regression models demonstrated a significant relationship between VSL and ART, as well as the effects of both VSL and ART on reading comprehension (Appendix B3 and B4). These models were used to estimate direct and indirect pathways in the mediation analysis. The output is shown in Table 4 and Figure 2.

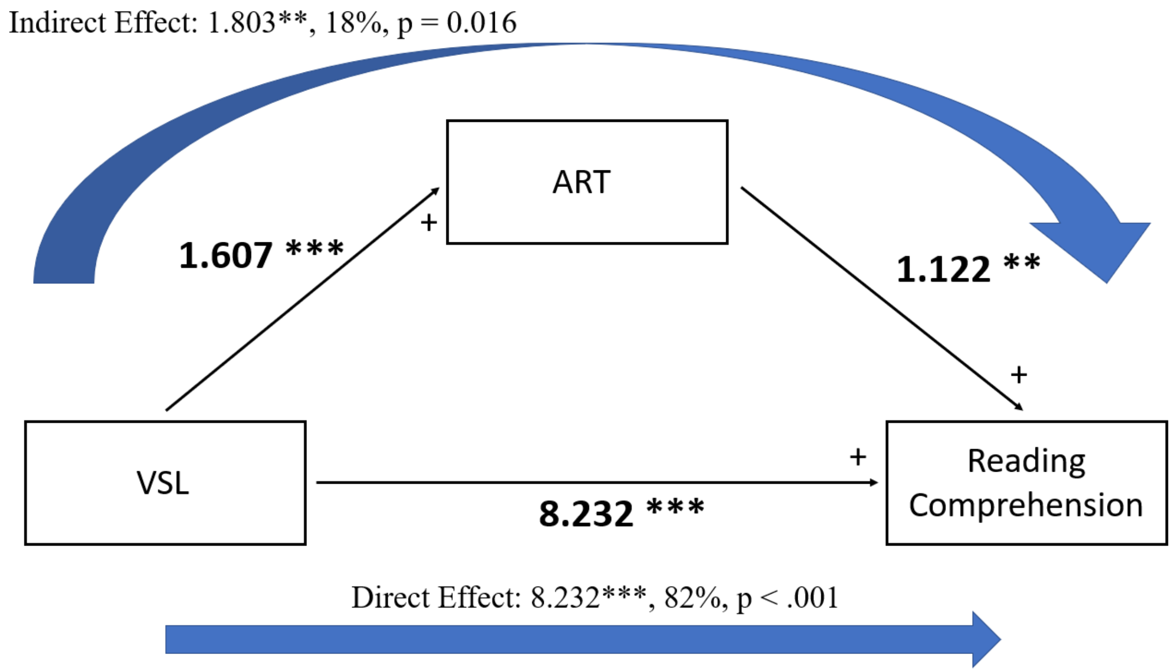
Table 4

Mediation analysis of relationship between VSL and reading comprehension mediated by ART (N = 188). The 'Average Causal Mediation Effect' (ACME), 'Average Direct Effect' (ADE), 'Total Effect', and the 'Proportion Mediated' are displayed. Nonparametric bootstrap confidence intervals with the percentile method simulated in 10000 iterations.

	Estimate	95% CI Lower	95% CI Upper	p value
ACME	1.803	0.297	3.910	.016
ADE	8.232	3.186	13.180	< .001
Total Effect	10.036	5.195	14.890	< .001
Prop. Mediated	0.180	0.031	0.450	.012

The indirect (ACME), direct (ADE), and total effect were all found to be statistically significant at the $p < .05$ level (Table 4). The indirect effect of VSL on reading comprehension that was mediated by ART accounted to 18% of the total explained variance (95% CI: [3%, 45%]). Yet importantly the direct effect of VSL on reading comprehension accounted for the majority of the variance (estimated at 82%, 95% CI: [54%, 96%]). These findings are visualized in Figure 2.

Figure 2
Mediation analysis of the effect of VSL on reading comprehension with ART as the mediator (N = 188, bootstrap simulations = 10000). The significance values for the individual connections in the triangle come from the estimates of each connection in the base models (Appendix B3 and B4).



Mediation Analysis: Motivation and ART as Parallel Mediators

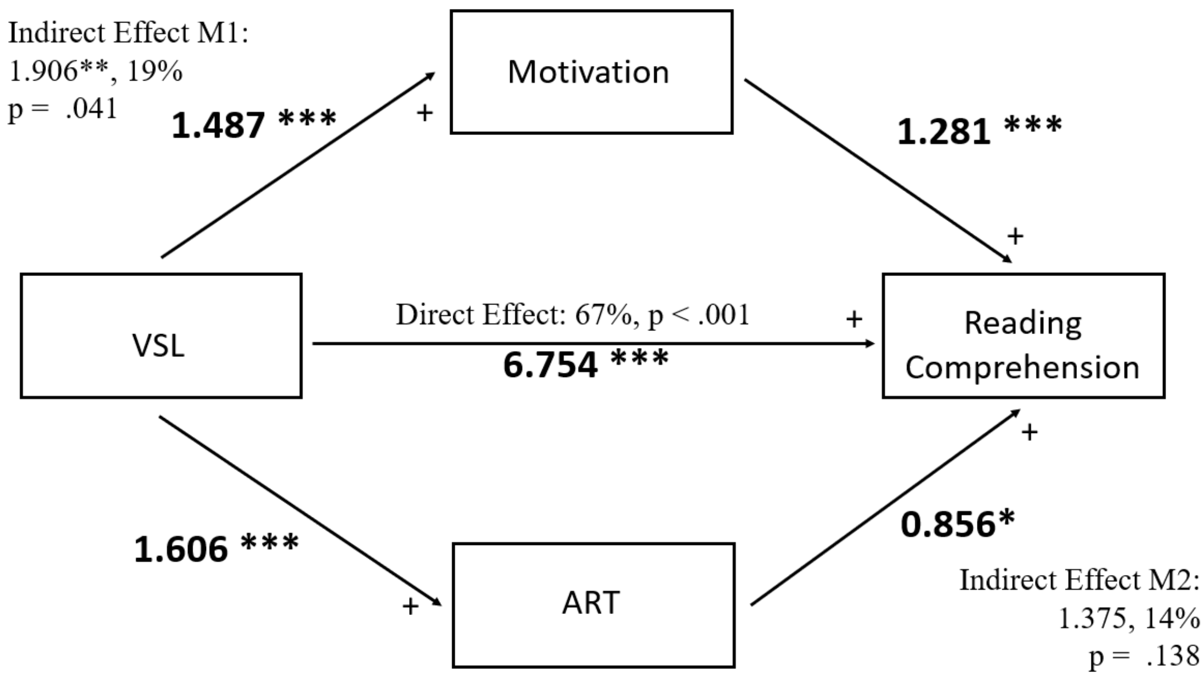
Since both motivation and ART were found to (partially) mediate the effects of VSL on reading comprehension, it was important to see how these variables function as parallel mediators. To look at multiple mediators in the same model, function sem in the laavan package in R was used (Rossell, 2012). In this model, VSL was again the independent variable and reading score was the dependent variable, but this time both motivation (M1) and ART (M2) were included as parallel mediator variables. Table 5 and Figure 3 show the output from the analysis for the total effect, the direct effect of VSL and the indirect effect of VSL through both mediators.

Table 5
Mediation analysis of relationship between VSL and reading comprehension mediated by both motivation and ART (N = 188). The Total Effect, Direct Effect and the Indirect Effects of both the first (M1) and the second mediating variable (M2) are displayed. Nonparametric bootstrap for standard errors simulated in 1000 iterations.

	Estimate	Standard Error	z-value	p value
Total Effect	10.035	2.401	4.018	< .001
Indirect M1: Motivation	1.906	0.932	2.045	.041
Indirect M2: ART	1.375	0.928	1.482	.138
Direct Effect	6.754	2.445	2.762	< .001

We found that the total effect of the model was significant, and the direct effect of VSL on reading comprehension was strong and significant too, with a proportion of variance explained estimated at 67% (95% CI: [33%, 89%]). The estimated proportion of total variance explained that was mediated by motivation was 19% (95% CI: [3.5%, 44%]), and was statistically significant. However, the proportion mediated through ART estimated at 14% (95% CI: [-2%, 37%]) was not statistically significant.

Figure 3
Mediation analysis of the effect of VSL on reading comprehension with motivation and ART as parallel mediators (N = 188, bootstrap simulations = 1000). Significance values for each connection in the figure can be found in Appendix B1-B6.



Nature of the Mediation

Our analysis up until now points to a significant relationship between VSL and reading comprehension that is partially mediated by motivation. This section specifies the nature of this mediation by considering an interaction between VSL offline scores and motivation as predictors of reading comprehension. A multiple linear regression model was fitted to reading comprehension with a critical interaction of VSL and motivation. ART and spelling scores were included as additional main effects (Table 6).

Table 6
Linear regression model fitted to reading comprehension with a critical interaction VSL and motivation. The model residual standard error is 5.43 on 182 degrees of freedom; $R^2 = 0.2183$*.*

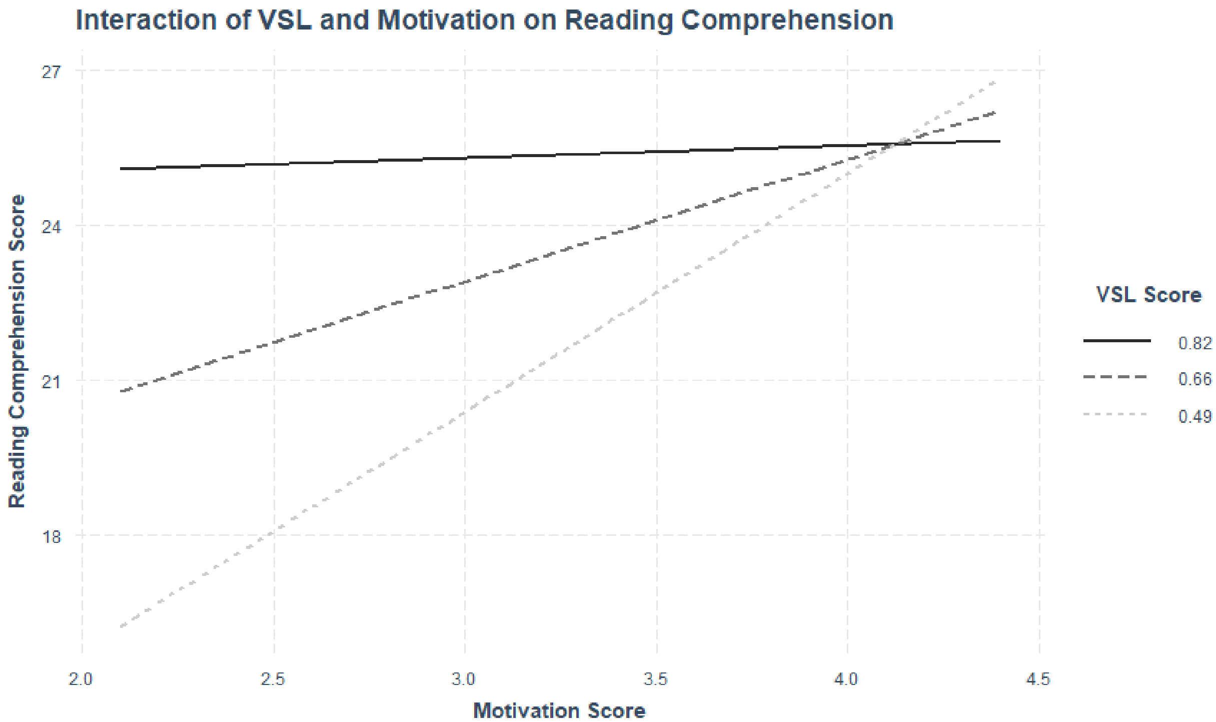
	Estimate	Std. Error	t value	p value
Intercept	18.438	1.740	10.595	< .001
ART	0.054	0.487	0.112	.911
Spelling	1.619	0.476	3.404	< .001
Motivation	4.515	1.796	2.514	.013
VSL	8.484	2.602	3.260	.001
VSL:Motivation	-5.388	2.712	-1.986	.048

This model indicates a statistically significant interaction of VSL offline scores and motivation. Figure 4 visualizes this interaction, and the effects of motivation on reading comprehension at different levels of VSL scores are shown in Table 7. It can be seen that higher motivation boosts comprehension scores but only when VSL accuracy is low. When VSL scores are higher there appears to be less and less of an advantage of having a higher motivation score.

Table 7
Estimates for the VSL scores are the different levels of accuracy of interest: the mean – 1 standard deviation (0.49), *the mean* (0.66), *and the mean + 1 standard deviation* (0.82). *The confidence level used is 95% on 182 degrees of freedom.*

	Estimate	Std. Error	Lower CL	Upper CL
VSL mean – 1SD	1.858	0.588	0.698	3.02
VSL mean	0.970	0.425	0.131	1.18
VSL mean + 1SD	0.082	0.645	-1.190	1.35

Figure 4
Plot of the interaction between VSL and motivation in their influence on Reading Comprehension. VSL score has been broken down into 3 bins representing accuracy at the mean – 1 standard deviation, the mean, and the mean + 1 standard deviation (N = 188).



Additional models (Tables B7 and B8 in Appendix B) revealed that in contrast to motivation, there were no significant interaction between ART and VSL, nor between Spelling and VSL in predicting reading comprehension. In the end, the results suggest that the only variable which has a significant interaction with VSL in predicting comprehension was motivation.

General Discussion

Recent findings show support for the idea that SL ability influences language acquisition and use. This link was also found in a more specific area, between VSL and reading skill (Arciuli & Simpson, 2012; Conway et al., 2012; Frost et al., 2013; Hung et al., 2018). However, the question of how this influence is transpiring is still being investigated (see Siegelman et al., 2018 for a review). An important obstacle to understanding the nature of the VSL-reading comprehension link is that both these cognitive abilities are multi-componential. Thus, it is possible that the impact of SL on reading comprehension is indirect, i.e., mediated by one or more component skills of reading. Prior reports of correlations between individual variability in SL and component skills related to reading (e.g., Lany et al., 2018; Misyak et al., 2010; and Singh et al., 2012) support this possibility.

Another possibility is that the hypothesized relationship between VSL and reading is mediated by a theoretically irrelevant third variable. Given the complex multi-componential nature of SL and reading comprehension, and the nature of experimental tasks that measure these abilities, it might be the case that the critical relationship is driven – partly or fully – by a “third” variable or multiple mediating variables. Such a scenario would cast doubt on the validity of the proposed VSL-reading comprehension connection. One such variable is participants' motivation, as greater individual motivation may cause higher performance of some individuals in all tasks and give rise to spurious, cognitively irrelevant correlations, including between measures of SL and reading abilities.

The present findings based on a sample of 188 proficient L1 readers of English confirmed that VSL is indeed a significant independent predictor of reading comprehension, in line with earlier reports (see the Introduction). We also identified two additional variables that met the formal statistical criteria for mediation: motivation and exposure to print. Further mediation analyses revealed that most of the effect that VSL has on reading comprehension follows a direct pathway. Motivation turned out to be a statistically significant mediator but this indirect pathway – from VSL to motivation to reading comprehension – only accounted for a small proportion of total explained variance.

A further inspection of the role of motivation revealed an interactive pattern. Participants with higher levels of SL ability achieve high reading comprehension scores regardless of how motivated they are to do well in the experiment. Yet among less proficient learners, higher motivation facilitated their reading comprehension. This suggests that motivation does present undesirable source of variability, i.e., it inflates correlations between VSL and reading comprehension scores for reasons unrelated to either cognitive domain. However, the scope of this mediation is confined to the low range of VSL ability and does not explain much variance overall.

The mediation analyses in our study highlighted the importance of accounting for exposure to print and motivation when studying the relationship between reading and SL. In addition to this information, they also served to rule out some potential confounding variables and helped to validate the SL task itself when investigating reading comprehension. Multiple recent studies in the literature have focused on the issue of reliability of SL tasks (Arnon, 2020; Isbilen et al. 2017; Siegelman et al. 2017). We add to this discussion an important aspect for the topic of SL as an individual ability – the issue of validity. We demonstrated that the impact of SL ability on reading comprehension cannot be accounted for by motivation, exposure to print, spelling, or a combination of these additional variables. These results suggest that VSL is not only a reliable measure of performance, but is also a valid predictor of reading skill in and of itself.

Our current findings add to the growing support of a connection between SL and reading ability. However, future research is required in order to fully understand how these complex concepts are related, specifically in light of the multi-componential nature of SL and reading. There are many skills and elements connected to reading ability, and many different aspects of SL. Our study looked at reading comprehension and two relevant component skills - exposure to print and spelling ability. However, there are countless additional components of interest that need to be addressed in relation to SL ability and its validity as an independent predictor of reading comprehension. Furthermore, our measure of SL ability was limited to the visual domain, but there are many other domains and facets of SL that could be measured and tested. There exists some inconsistency between different measures of SL which brings up a potential reliability issue potentially stemming from measurement error. Given the possibly misleading interpretations this can present, particularly in individual differences studies, additional research should focus on including multiple measures of each variable in these types of circumstances (see overview by Rouder & Haaf, 2019). Previous studies have suggested that certain SL domains are more or less related to specific linguistic phenomena; therefore, a larger scale study investigating the many components of reading ability alongside multiple measures of SL ability across domains is needed (see Hung et al., 2018; Qi et al., 2019). Even within reading, which is widely thought to primarily involve the visual domain, we know certain skills rely on other sensory systems. Future studies could focus on highlighting the multi-componential nature of both SL and reading and how they are interconnected.

In summary, this study shows that SL has an influential role in reading comprehension and provides insight into how SL relates to other skills and abilities in its influence on reading ability. This study further highlights the importance of investigating multiple available types of SL against multiple components of reading ability and tailoring these tests to the appropriate components of reading and language. While attention has been focused on the reliability of SL, this study demonstrates that it is a valid measure to use when studying reading ability.

Appendix A

The motivation survey used at the end of testing to gauge the general motivation of the participants in completing the tasks and doing well. Developed in the Student Opinion Scale (SOS) (Thelk, Sundre, Horst, & Finney, 2009).

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.

1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

______ Doing well on this test was important to me.

______ I engaged in good effort throughout this test.

______I am not curious about how I did on this test relative to others.

______I am not concerned about the score I receive on this test.

______This was an important test to me.

______I gave my best effort on this test.

______While taking this test, I could have worked harder on it.

______I would like to know how well I did on this test.

______I did not give this test my full attention while completing it.

______While taking this test, I was able to persist to completion of the task.

Appendix B

B1

Regression model for motivation as explained by VSL. The model residual standard error is 0.972 on 186 degrees of freedom; $R^2 = 0.060$.

	Estimate	Std. Error	t value	p value
Intercept	-0.979	0.293	-3.344	< .001
VSL Offline	1.487	0.432	3.447	< .001

B2

Regression model for reading comprehension as explained by VSL and motivation. The model residual standard error is 5.686 on 185 degrees of freedom; $R^2 = 0.129$.

	Estimate	Std. Error	t value	p value
Intercept	18.627	1.762	10.570	< .001
VSL Offline	7.868	2.603	3.022	< .001
Motivation	1.457	0.429	3.398	< .001

B3

Regression model for ART as explained by VSL. The model residual standard error is 7.205 on 186 degrees of freedom; $R^2 = 0.070$.

	Estimate	Std. Error	t value	p value
Intercept	-1.057	0.291	-3.633	< .001
VSL Offline	1.607	0.429	3.74	< .001

B4

Regression model for reading comprehension as explained by VSL and ART. The model residual standard error is 5.759 on 185 degrees of freedom; $R^2 = 0.106$.

	Estimate	Std. Error	t value	p value
Intercept	18.387	1.793	10.250	< .001
VSL Offline	8.232	2.651	3.106	< .001
ART	1.122	0.437	2.569	.011

B5

Regression model for Motivation as explained by VSL and ART. The model residual standard error is 0.3881 on 185 degrees of freedom; $R^2 = 0.100$.

	Estimate	Std. Error	t value	p value
Intercept	-0.789	0.297	-2.554	.011
VSL Offline	1.153	0.439	2.627	.009
ART	0.208	0.072	2.874	.005

B6
Regression model for reading comprehension as explained by VSL, motivation, and ART. The model residual standard error is 5.643 on 184 degrees of freedom; $R^2 = 0.147$.

	Estimate	Std. Error	t value	p value
Intercept	19.359	1.789	10.824	< .001
VSL Offline	6.754	2.646	2.553	.011
Motivation	1.281	0.435	2.945	.004
ART	0.856	0.437	1.957	.052

B7
Regression model for reading comprehension as explained by VSL, ART, and the interaction between VSL and ART. The model residual standard error is 5.775 on 184 degrees of freedom; $R^2 = 0.106$.

	Estimate	Std. Error	t value	p value
Intercept	16.880	3.097	5.451	< .001
VSL Offline	8.070	4.801	1.618	.094
ART	0.141	0.233	0.606	.545
VSL Offline *ART	0.014	0.338	0.041	.968

B8
Regression model for reading comprehension as explained by VSL, spelling, and the interaction between these VSL and spelling. The model residual standard error is 5.540 on 184 degrees of freedom; $R^2 = 0.177$.

	Estimate	Std. Error	t value	p value
Intercept	-12.894	17.091	-0.754	.452
VSL Offline	28.255	26.054	1.084	.280
Spelling	0.802	0.439	1.826	.070
VSL Offline *Spelling	-0.497	0.666	-0.746	.456

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience research*, *64*(2), 185-190.

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.

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. S. Adelman (Ed.), *Visual word recognition* (Vol. 2, pp. 151–172). London, UK: Psychology Press.

Andrews, S. (2015). Individual differences among skilled readers: The role of lexical quality. In A. Pollatsek & R. Treiman (Eds.), The Oxford handbook of reading (pp. 129 –148). New York, NY: Oxford University Press.

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*, 1-26.

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 634-643.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286-304.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior research methods*, *52*(1), 68-81.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological science*, *15*(7), 460-466.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bogaerts, L., Siegelman, N., & Frost, R. (2020). Statistical Learning and Language Impairments: Toward More Precise Theoretical Accounts. *Perspectives on Psychological Science*, 1745691620953082.

Brand, R. J., McGee, A., Kominsky, J. F., Briggs, K., Gruneisen, A., & Orbach, T. (2009). Repetition in infant-directed action depends on the goal structure of the object: Evidence for statistical regularities. *Gesture*, *9*(3), 337-353.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, *96*(1), 31.

Campbell, K. L., Zimerman, S., Healey, M. K., Lee, M., & Hasher, L. (2012). Age differences in visual statistical learning. *Psychology and Aging*, *27*(3), 650.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of memory and language*, *45*(4), 751-774.

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103-120.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356-371.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.

Ravignani, A., Westphal-Fitch, G., Aust, U., Schlumpp, M. M., & Fitch, W. T. (2015). More than one way to see it: Individual heuristics in avian visual computation. *Cognition*, *143*, 13-24.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences*, *10*(7), 294-300.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific studies of reading*, *10*(3), 277-299.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in psychology*, *5*, 335.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, *19*(3), 117-125.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language?. *Psychological science*, *24*(7), 1243-1252.

Goujon, A., & Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (Papio papio). *Behavioural Brain Research*, *247*, 101-109.

Guthrie, J. T., & Coddington, C. S. (2009). *Reading motivation.* In K. R. Wenzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (p. 503–525). Routledge/Taylor & Francis Group.

Hung, Y.-H., Frost, S. J., & Pugh, K. R. (2018). Domain generality and specificity of statistical learning and its relation with reading ability. In T. Lachmann & T. Weis (Eds.), Reading and dyslexia: From basic functions to higher order cognition (pp. 33–55). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-90805-2_2

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In *the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 564-569). Cognitive Science Society.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific reports*, *6*, 19741.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3474-3486.

Lany, J., Shoaib, A., Thompson, A., & Estes, K. G. (2018). Infant statistical-learning ability is related to real-time language processing. *Journal of child language*, *45*(2), 368-391.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in psychology*, *1*, 31.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. Annual Review of Psychology 2007 58:1, 593-614

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McFall, S. R., Jamieson, J. P., & Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *96*(1), 135.

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.

Neger, T. M., Rietveld, T., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in human neuroscience*, *8*, 628.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes*, *18*(4), 443-468.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). *The Acquisition of Reading Comprehension Skill*. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology*. *The science of reading: A handbook* (p. 227–247). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch13

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, *23*(1), 101-115.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, *48*(2), 1-36.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. Psychonomic bulletin & review, 26(2), 452-467. doi: 10.3758/s13423-018-1558-y

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, *35*(4), 606-621.

Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of neurodevelopmental disorders*, *10*(1), 1-5.

Sawi, O. M., & Rueckl, J. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading*, *23*(1), 8-23.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? Scientific Studies of Reading, 23(1), 64–76.

https://doi.org/10.1080/10888438.2018.1482304

Seitchik, A. E., Brown, A. J., & Harkins, S. G. (2017). *Social facilitation: Using the molecular to inform the molar.* In S. G. Harkins, K. D. Williams, & J. M. Burger (Eds.), *Oxford library of psychology. The Oxford handbook of social influence* (p. 183–203). Oxford University Press.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247-271.

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

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*(3), e12365.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior research methods*, *49*(2), 418-432.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities?. *Cognitive science*, *42*, 692-727.

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental science*, *15*(4), 482-495.

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2014). Statistical learning is related to early literacy related skills. Reading and Writing, 28(4), 467–490. https://doi.org/10.1007/s11145-014-9533-0

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

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC neuroscience*, *10*(1), 1-8.

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*, 129-151.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language learning and development*, *3*(1), 1-42.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K., & Yamamoto, M. T. (2019). Package 'mediation'. *Computer software manual*.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, *98*(3), 642.

Treiman, R., Kessler, B., Boland, K., Clocksin, H., & Chen, Z. (2018). Statistical learning and spelling: Older prephonological spellers produce more wordlike spellings than younger prephonological spellers. *Child development*, *89*(4), e431-e443.

van der Kleij, S. W., Groen, M. A., Segers, E., & Verhoeven, L. (2019). Sequential implicit learning ability predicts growth in reading skills in typical readers and children with dyslexia. Scientific Studies of Reading, 23(1),77–88. https://doi.org/10.1080/10888438.2018.1491582

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*, *44*(1), 249-279.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific studies of reading*, *11*(1), 3-32.

Wiederholt, J. L., & Bryant, B. R. (2001). GORT 4 Gray Oral Reading Tests Examiner's Manual. *Austin, TX: Pro-Ed*.

Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child development perspectives*, *10*(3), 190-195.

CHAPTER 5.

Discussion

Overview

The goal of this thesis was to investigate individual differences in reading-related skills and cognitive abilities that influence reading proficiency across language backgrounds, and to use a variety of different measures and methods of analysis to see how they interact. Overall, the objectives of this thesis were addressed by the combination of three projects included in this thesis. The first objective was to investigate how different measures of reading proficiency contribute to understanding individual differences in reading. The second objective looked at how different methods of data collection and analysis can be used to inform us about individual differences in reading proficiency. Finally, the third objective asked whether individual differences interact when explaining variation in reading proficiency. This objective also aimed to answer whether language background plays a role in the impact of individual differences in reading proficiency. To meet these objectives, this thesis employed a multi-faceted approach to gain further understanding of this complex topic.

To address the first objective, both reading comprehension and reading fluency were included as measures of reading proficiency. The use of eye-tracking in Chapter 3 addressed reading fluency by looking at both overall reading speed of a sentence, as well as the total time it took to process individual words. Reading comprehension was directly assessed using the Grey Oral Reading Test (GORT) in Chapter 4; and Cloze probability (in Chapter 2) is closely associated with both reading fluency and reading comprehension.

To investigate the second objective, all three papers used different combinations and methods of data collection and analysis. Chapter 2 refined a method of collecting cloze probability data, Chapter 3 collected data from three different populations of English speakers, and Chapter 4 was a large online study resulting in the analysis of 188 participants. These studies culminated in a number of different measures that included reading component skills such as spelling, vocabulary, exposure to print, and cloze probability; cognitive factors including statistical learning (SL) and motivation; and language background focussed on both first language (L1) and second language (L2) information. To build upon the typical analyses, Chapter 3 outlines a novel statistical prediction methodology, and Chapter 4 utilized advanced mediation analyses to further investigate the data collected.

Finally, to answer the question posed in the third objective, both Chapter 3 and Chapter 4 primarily focus on interactions between variables of interest. Chapter 3 looked at the interaction between language background and the individual differences measures of spelling, vocabulary, and exposure to print. In Chapter 4, the nature of the relationship between SL and reading comprehension was examined by determining whether there are interactions between SL and motivation, spelling, vocabulary, and exposure to print.

Taken together, these three papers address the objectives set out for the thesis and have provided some important insights into the nature of individual differences in reading proficiency.

Implications & Significance

The results of the three studies presented in this thesis all contribute new and insightful information towards the main goal of this thesis, and to reading research in general. This section will briefly summarize the results found in each of the chapters, state the overall implications that can be drawn from these results, and discuss how this research significantly advances knowledge in the field of reading research.

Chapter 2

The primary goal of this paper was to introduce an easy-to-use application for collecting cloze probability ratings called 'Clozapp'. A major component of this study was comparing the data collected using this application to prior data collected using other measures. Upon analysis of the data, similar descriptive statistics of the predictability rate and number of 'full misses' were found between Clozapp data and the two comparison experiments (Reichle et al., 1998; Schnoebelen & Kuperman, 2010). The strong correlations between our data and the prior experiments suggests that Clozapp provides a reliable measure that can be used confidently moving forwards.

The significance of this project is straightforward, as it fills the need for an easy-to-use, widely available and free application to collect cloze probability ratings, which will be helpful for both current and future research. With the customizable script that accompanies this manuscript, researchers can adapt the demographic information and instructions to fit their unique needs without compromising the precision timing of the experiment. Since the instructions can be customized, researchers are also able to modify them to fit any language requirements. Given this application can be tailored for many different purposes, there is an opportunity for it to become commonplace amongst researchers in the field. If that were the case, an additional benefit from this project will be an increase in consistency when studying this variable, which will facilitate easier comparison across studies. Cloze probability is an important and interesting tool of reading research, and with the availability of an open-source application, user manual, and sample scripts it is hoped this will spark and support further investigation of this variable.

Chapter 3

This study found that L2 speakers with different L1 language backgrounds perform differently in both component skills tests of spelling, vocabulary, and exposure to print, as well as in their eye-movement fluency in sentence reading. The finding that L1 and highly proficient L2 speakers of English have different eye-movement patterns when reading confirms what prior studies have already posited (Beglar & Hunt, 2014; Brysbaert, 2019). This suggests that regardless of how skilled L2 readers of English are in a number of different areas, their fluency in reading may not match L1 counterparts. The impact of this paper is both theoretical and methodological in nature when considering the implications for the field.

From a theoretical perspective, there were several findings with significant implications for research in this field. While many factors are at play in the findings, the matter of L1-L2 distance highlights an important distinction that may be driving these differences (Jeon & Yamashita, 2014; Melby-Lervåg & Lervåg, 2011). Chapter 3 found that German L2 speakers of English were able to reach reading speeds closer to that of the Canadian cohort of L1 English speakers, while Finnish L2 speakers of English were not, which suggests a potential L1 language impact. This study found what appears to be an impact of L1-L2 distance, in that the more similar an individual's L1 is to English (their L2), the closer their eye-movements are to L1 English speakers. This raises the possibility that the L1 an individual learns in life can potentially shape their ability to reach L1-like fluency in another language. This topic has been of interest for several decades in the reading literature, and this study supports previous findings using a different methodological approach. Chapter 3 contributes to this literature and suggests that not all L2 speakers are necessarily alike. Furthermore, while there has been a growing number of studies addressing L2 reading, there are relatively few studies that demonstrate comparable groups of L2 readers on the same texts (see Duyck et al., 2007; Godfroid, 2020). These comparisons are important in order to advance the important question of what factors contribute to becoming a proficient L2 reader. This study contributed important findings of the similarities and differences of reading in English from a number of different language backgrounds.

Another important discovery from this study was that some component reading skills are more important than others in achieving high fluency in English as an L2 speaker. Furthermore, the specific skills that are statistically most important appear to be dependent upon the L1 language background of the individual. This suggests that in order for a given individual to gain the highest possible fluency, they may need to focus their efforts on specific component skills as determined by their L1.

As stated previously, there was also a methodological impact of this paper, which highlighted a useful statistical approach used to investigate the differences between our populations of interest. The availability of this data analysis technique provides opportunity to advance our knowledge in this field in a unique way. Using hypothetical participants that fit a particular performance or outcome on variables of interest, by-passed the need to recruit individuals that fit these requirements. In situations where these participants would likely be outliers or at the far ends of the spectrum of variance, this is a highly beneficial approach that can be used across a number of areas in language research. Specifically, it will be helpful for populations where it is difficult to collect large amounts of data. In addition, even if the values in question are not thought to exist naturally, this method can allow researchers to extrapolate into unknown areas.

This project resulted in another significant implication when considering the societal application of this work, primarily when thinking about teaching English as an L2. English is growing as a requirement both for work and social interactions in many areas around the world and in many industries (Arkoudis et al., 2009; Arkoudis et al., 2014; Bleakly & Chin, 2010). Better understanding the underlying factors for various L1's towards demonstrating proficiency in English as an L2 speaker, may allow for more targeted approaches for language instruction. Therefore, the applied value of the project is in its potential of boosting the efficiency of teaching English as an L2 by focusing on component skills that are most important for a given language background.

Chapter 4

This project sought to investigate the relationship between SL and reading comprehension, and made some advancements toward answering critical questions in the reading field. Primarily, this thesis supported previous findings that SL ability is related to reading ability (Arciuli & Simpson, 2012; Misyak, Christiansen, & Tomblin, 2010). This study found that visual SL is a significant independent predictor of reading comprehension, which supports the link between SL and reading comprehension as valid. In addition, this paper included a number of component skills to model their relationship to reading comprehension. All component skills such as spelling, vocabulary and exposure to print were found to have a statistically significant relationship or role in determining an individual's reading ability, but it was also found that they are not responsible for the effects of SL on reading comprehension. This significant finding rules out the potential for these factors to serve as confounding variables in this relationship, which implies that SL is not simply an expression of component skills. However, additional research is required to determine if and how these component skills interact with various forms of statistical learning, and whether additional reading or general cognitive skills play a role.

One general cognitive factor included in this study, task motivation, was found to significantly mediate the relationship between SL and reading. While the direct effect of SL on reading remained significant and strong, this finding is interesting as it suggests that motivation, too, is an important variable to consider when determining an individual's reading comprehension ability. The nature of this mediating relationship was determined to be an interaction between SL and motivation. Motivation was found to have a significant effect when an individual's SL ability was low, but was not significant when an individual's SL ability was high. Another way to think about this relationship is that motivation can only significantly improve reading comprehension if there is a lack of SL ability, indicating that motivation may be a compensating factor for those with a low SL ability. To my knowledge, this is the first study to demonstrate an interaction between SL and motivation. This is a significant finding with future research implications, as it shows that the relationship between SL ability and reading comprehension can be tied to cognitive factors, and opens the possibility that there may be additional factors that could be playing a role.

This study significantly contributes to the body of research suggesting a link between SL and reading. Given that this field is still in its infancy, further evidence that there is a strong connection between SL and reading proposes that this area of inquiry is worth investigating further. In addition, this study further advances our knowledge regarding the relationship between the cognitive factors of SL and motivation. The evidence suggesting a relationship between SL and motivation is novel, and identifies an interesting new field of investigation into the nature of the role of motivation in reading comprehension. It also highlights the possibility of investigating additional cognitive factors, and creates a new area for SL research in relation to reading and language in general.

Overall Thesis Impact

These chapters address ongoing questions surrounding individual differences in reading proficiency, and work together to contribute new analysis and understanding of reading proficiency as demonstrated through individual and group variation using a variety of measures and methods.

Using different measures of proficiency - fluency and comprehension - allowed for insight into the different aspects that impact reading ability. Both the time it takes for an individual to read text, as well as the ability to understand and apply the information they have processed, contribute to success in reading (Jenkins et al., 2003; Kim et al., 2012; Zadeh et al., 2012). This thesis looked at fluency using eye-tracking methodology which allowed a view into this important aspect of reading by examining how quickly individuals can process text. Importantly, the view of fluency was shown from both L1 and L2 readers of English, giving further insight into questions surrounding language background and reading. Investigating reading comprehension through complex texts was another way to look at reading proficiency, and allowed us to gain insight into what factors are important when reading for understanding.

Examining both fluency and comprehension is essential when determining reading proficiency. This thesis demonstrated which specific critical reading component skills are involved in both fluency and comprehension, and identified how each of these component reading skills differ in their relationship to, and influence on, these different aspects of reading proficiency. The fact that the reading-related skills studied in this thesis contributed differing amounts of explained variance to the different measures of reading proficiency suggests a complex interplay between variables, and supports the notion that reading proficiency is indeed multi-componential in nature.

Assessing both fluency and comprehension using various skills, cognitive abilities, and factors that contribute to reading proficiency was also essential in pushing the field forward. By investigating a variety of variables in tackling the main goal of this thesis, the complex nature of this field of research was demonstrated along with illustrating the connections between various factors. Being a proficient reader requires not only a mastery of the different components of reading, but also relies on certain cognitive abilities like SL and motivation, and can vary depending on an individual's language background. This thesis was a beneficial contribution to the literature as it demonstrated ways in which a wide variety of variables can interact with one another to create an overall picture of reading proficiency.

This thesis also utilized different experimental designs and methodologies including eye-tracking methodology and a large-scale online study, which all contributed uniquely to addressing the goals and objectives. This work also prioritized data analysis tools and techniques through the development of a novel statistical prediction procedure, and the utilization of mediation analysis. Furthermore, all three projects contributed complementary code that detailed the analyses completed on the data in order to contribute to future research in the field.

This thesis used novel approaches and analyses that will have an enduring impact in this field of research. It highlights the interactions, associations, and interplay between important variables of interest related to the topic of individual differences in reading proficiency. This was a novel approach towards addressing important questions in the field for a number of reasons. First, this thesis resulted in a newly available application for collecting data that did not previously exist, which will hopefully benefit the field with increased consistency and ease of access and use for cloze probability data collection. Secondly, it saw the development of a new method of analysis that can allow researchers to further investigate differences between groups with variance on a number of different factors. The approach of creating hypothetical participants for data analysis also has its merits, and can be applied to many different fields. Finally, the investigation into the nature of the relationship between SL and reading comprehension focused on a new variable, motivation. This will contribute to further investigation into this complex variable and open up new avenues of discussion in the field. Additional discussion as to how this thesis fits into the current field and outstanding questions will be addressed in the next section.

Limitations and Future Directions

In addition to the impact of these three projects and the thesis as a whole, it is also important to discuss related limitations and future directions. This section will outline some planned projects that supported my goals and objectives that were begun but not completed due to external circumstances, and were therefore not included in this thesis. It will also include a discussion of the overall limitations of this thesis as a cohesive body of work, with ideas as to how these limitations may be addressed, and future outstanding questions that remain in the field.

COVID Impact

It likely comes as no surprise that I encountered some setbacks as this thesis was completed during the COVID-19 global pandemic. Along with the vast majority of researchers, I was primarily impacted by the lack of access to the lab and the inability to run in- person participants. I am very grateful to the university for prioritizing the health and safety of the students, faculty, and staff, however this closure did have a significant impact on my thesis. The primary impact was the freeze on data collection for a key study that directly deals with the goals and objectives of this thesis, which was designed and programmed with data collection partially underway. Another study was planned and programmed and was about to begin piloting. Below I will discuss the details of these projects, provide any preliminary findings, and discuss how these projects fit into the major goals of this thesis.

One key study that was underway used the application Clozapp, which was outlined in Chapter 2, to address some critical aspects of this thesis by combining the investigation of both reading fluency and comprehension. This project utilized data collected from the Multilingual Eye-tracking Corpus L2 (MECO L2), a project that I closely assisted with throughout my graduate studies. MECO L2 is a resource for eye-tracking data of reading texts in English by L2 speakers from 11 different language backgrounds and includes a large battery of component skills of reading in each groups' L1 and L2 (English) (Kuperman et al., 2021b). The goal of this in-progress project is to investigate how individuals from multiple language backgrounds perform in predicting upcoming words in the same texts as the MECO corpus, taken from the Grey Oral Reading Test (GORT), and the corresponding reading comprehension questions. The data collected using Clozapp was to be matched to data in the MECO corpus in order to compare individuals with similar linguistic backgrounds. This would allow us to see how language background interacted with cloze probability ratings and whether there was a relationship between reading comprehension and fluency. In other words, does language background and the ability to predict upcoming text influence reading comprehension and fluency in English? Previous research has shown that there is an impact of how predictable a word is in text on the speed of processing that word (for a review see Staub, 2015). Looking at this interaction would have been a valuable addition to the objectives of this thesis. This study also sets out to determine how individuals with similar language backgrounds perform on both comprehension and fluency assessments, and whether any connections exist based on L1-L2 distance. Given the complexity of the design of this study, a large sample of participants was needed in order to have enough data (30 participants for each of the GORT texts), and unfortunately there was too little data available prior to the campus closures to conduct a sound analysis. A very preliminary look at the data shows consistent rates of cloze probability to what has previously been seen in the literature, and it is anticipated this experiment will resume and continue the investigation into these important questions.

Another project was designed, programmed and about to begin piloting. This study focused on investigating individual differences in SL ability using electroencephalography (EEG). Few studies have investigated the time-course of SL using EEG; however, it appears to be a promising area of study (Abla et al., 2008; Abla & Okanoya, 2009; Batterink, 2017; Jeste, 2015; Kudo, 201; Teinonen, 2009). There are still quite a few unknowns regarding the process of SL in and of itself, and this study hoped to determine what Event Related Potential (ERP) components are impacted during SL to better understand the general time course of these effects. Specifically, this study hoped to build upon a study by Koelsch et al., 2015 that looked at auditory SL with tones while measuring EEG, focusing on the specific ERP known as statistical mismatch negativity (sMMN). The primary distinction in the planned study is the use of auditory speech sounds rather than tones to look at a linguistic component. Given the complexity of SL and its still unknown mechanisms, this study could build upon our current knowledge and potentially spark a new area of investigation into its connection to reading and literacy.

While these studies were unable to be completed and included in this thesis, they still pose interesting questions that fit well within the current goals and objectives of this thesis. The first study combining the use of MECO and Clozapp would primarily address the first objective by investigating whether the level of predictability influences reading speed (using reading rate and total fixation duration on a word) and comprehension (using performance on the GORT texts) of the participants in the corpus. It would also address the second and third objectives by investigating a number of different reading-related component skills. Another primary goal was to see how predictability changes over the duration of the paragraph, and to determine whether there are any influences of language background on the cloze probability task, and whether this correlates to any differences seen within reading fluency or comprehension. The EEG study primarily relates to the second objective, as it is a deep dive into one variable that appears to have an influential role in reading proficiency, but the mechanisms of this relationship are still relatively unknown. These studies will continue to investigate important questions in the field related to individual differences in reading proficiency.

Limitations and Future Directions

This section will briefly outline high-level limitations in this thesis, discuss how these limitations could be addressed, and identify future directions for this field of research.

Complexity of Variables

Despite reading proficiency being a highly applicable topic, and therefore a thoroughly studied area, there is still much to be learned regarding the variation in an individual's ability to master this crucial skill. In part, this is due to reading proficiency being highly multi-componential in nature, requiring the mastery of a number of different skills and abilities across domains, and given it can be influenced by numerous external factors. This notion has been at the forefront of this thesis, and while the goal was to look at some of these factors and skills, there are still many influences and interactions that were not accounted for. Of critical importance to note is the expansive number of reading component skills that have previously been found to impact reading ability both during early L1 acquisition and L2 acquisition (Oakhill, Cain, & Bryant, 2003; Vellutino et al., 2007). The component skills chosen for this thesis were deliberately picked to answer specific questions, however it is important to note that there are many others that may have a role in addressing the objectives posited in this thesis.

SL is another important variable included in this thesis that is known to be multi-componential in nature. SL can be demonstrated by learning different types of patterns (e.g., distributional, transitional), across different modalities (e.g., auditory, visual and tactile), on both verbal and non-verbal tests, and spanning all age groups (see the section 'Cognitive Factors' in the 'Introduction'). However, how and when in one's lifespan SL contributes to language learning is not fully understood. Visual SL was chosen as the modality in this study since it closely corresponds to a dominant modality used in reading. It also focused on transitional probability, which closely aligns to predictability on a number of different levels (from which letter may follow from another, to what the next word may be in a sentence), which are all instrumental skills in language, and reading in particular. The nature of statistical regularities within language especially within reading were highlighted throughout this thesis and an ongoing investigation into how these areas align would be an interesting topic of future research. The goal of this investigation into SL was to examine its relationship to reading, and questions surrounding the nature of this relationship are still unclear (see Siegelman et al., 2018). Therefore, there may be other aspects of SL that correspond to different aspects of reading proficiency that were not covered in this thesis and could help demonstrate the interconnected nature of these concepts.

Motivation is of critical interest when studying reading proficiency, and in and of itself is quite multifaceted. Specifically, when looking at language acquisition and the factors that apply to increasing the interest in learning a language, and in one's exposure to printed materials. This study looked specifically at task motivation, or simply wanting to do well on a test. This is an interesting factor for psychological research generally, as individuals vary in their desire to do well on tasks, and knowing this information can help to disentangle this factor from other variables of interest. In addition to general task motivation, it can also be beneficial to study another aspect of motivation particularly relevant when studying reading proficiency, which is an individual's motivation to engage in reading. This aspect of reading motivation has been heavily studied and broken down into a number of different components as well (see the section 'Cognitive Factors' in the 'Introduction'). Further research should look at the many different facets of motivation that are involved in an individual's decision to engage and perform to the best of their ability in reading and reading-related tasks.

In order to understand the complex notion of reading proficiency, research needs to focus on the myriad of components that contribute to it. Furthermore, the factors that may impact this complicated process could also be multi-componential in nature. It is important to see what aspects of these variables impact or correspond to which component(s) of reading proficiency in order get the full picture. As mentioned in the previous section, one study in progress takes into account a few additional reading component skills, and measures both reading comprehension as well as reading fluency. However, there are still countless additional combinations of factors that would be beneficial to study. Moving forward, it will be important to further study what specific aspects of a variable of interest are having an impact on which specific components of the reading process.

Language Representation

There is a dominant trend in language research to specifically focus on the English language when investigating cognitive processes in acquisition and use. While learning English has become an important aspect of globalization and fills a strong need for the many individuals that require proficiency in English (largely to navigate in an English-dominant country or field of work), it is important to also consider the plethora of other languages in which people read. Languages vary in a number of critical areas, from the type of script, the depth of orthography, or the morphological complexity, and it is important to understand the role of these linguistic aspects on reading. Gaining an appropriate understanding of the reading process requires investigating the many languages in which reading is being taught and learned, both as an L1 and L2, and currently the literature is primarily focused on English. This is a limitation of many language-focussed lines of research currently being pursued in addition to reading. Ideally the trend to focus on additional languages will continue to increase. With additional languages being focused on as target languages, the field as a whole will benefit from getting a better idea of the process of reading in general, rather than simply what this process looks like in English. In addition to reading literature being heavily English-dominated, virtually no knowledge exists on how the role of SL varies across different languages and writing systems that are widely represented in Canadian society (e.g., Chinese, Hindi, French, Arabic, and many others). This opens another largely unexplored area for future related research.

In addition to broadening the scope of L1 target languages, a limitation in this thesis and for many researchers interested in language acquisition, is the lack of diversity in language background. That is, taking into consideration the wide pool of L1 backgrounds that English learners stem from. As demonstrated in Chapter 3 of this thesis, there may be a significant impact on reading fluency given a specific language background, since some L1 speakers may be able to reach higher levels of fluency in English than others. This thesis represents two possible language backgrounds out of many, and it is therefore critical to increase languages studied to be more representative moving forwards. We found a large difference for the L2 English readers from two primarily transparent alphabetic scripts, so understanding these differences can help to direct future research. Furthermore, languages can vary on a number of different factors, and while Finnish and German vary in their similarities and differences to English, this is just a small subset of the many variations that exist between languages. In order to improve our knowledge of what contributes to L1-L2 distance, other aspects of languages that can differ, such as varying scripts need to be included. The MECO addresses this critically important limitation and offers data from a number of different languages (Kuperman et al., 2021a; Kuperman et al., 2021b). To address this issue and build our knowledge of proficiency in other languages, we plan to use the same statistical methodology developed in Chapter 3 to model the differences in language backgrounds provided by the MECO L2 corpus. Doing so will allow us to explore which aspects contributing to linguistic distance have an impact on reading fluency and comprehension.

Another consideration when investigating individual differences in reading proficiency across language backgrounds is ensuring that comprehensive language information can be collected. Getting an accurate picture of an individual's language skill, or even reading skill in general, requires looking into all aspects of their linguistic environment including their abilities in the L1 in addition to their L2 (see Bernhardt, 2011). One critical obstacle in collecting this information is the general western bias in testing material availability. While there are a whole host of tests and measures that tackle the different reading component skills and abilities in English, finding appropriate or matching tests in other languages can prove to be quite difficult. Furthermore, when tests that were designed specifically for one language are used to assess speakers of another language, accurate results are not always possible (McCarron & Kuperman, 2020). Therefore, it is important that future research begin to diversify the populations being studied in order to make significant contributions towards building the infrastructure for research to continue in operationally sound ways.

Further compounding this issue is the fact that for many researchers, the most easily accessible group of participants tend to be white, female university students, typically studying either psychology or linguistics. If the goal is to learn how to improve literacy in society, a broader, more diverse population is required to fully understand the mechanisms that influence becoming a proficient reader. It has been consistently demonstrated that variables such as socioeconomic status and cultural background, among other environmental factors, have significant impact on language and reading abilities (Abidin & Mohamad, 2019; Aikens & Barbarin, 2008). If these group differences could be influencing the individual differences in reading ability we are hoping to study, it is important that future research branch out to study these communities in order to gain a better perspective of the question at hand.

Conclusion

It is commonly understood that the study of literacy and reading proficiency is an important and complex topic of research with broad implications across the world. In fact, it is estimated that over the past 40 years, approximately one out of every thousand studies published in the scientific community relates to the topic of literacy (Kyröläinen & Kuperman, 2021). This thesis sought to add to our understanding of the multi-componential nature of reading by confirming previous findings, and identifying new results through the study of a variety of factors and their interactions, using novel approaches for data collection and analysis. Together, this thesis forms a coherent and substantial body of work that both adds credibility to previous findings, and identifies new and exciting directions for this research field.

As a whole, this thesis demonstrates the benefits of using an individual differences approach to studying reading proficiency. Additionally, it showcases the advantages of using a variety of measures of reading proficiency, data collection and data analysis methods, and investigating different populations. Examining individual variation within reading component skills, general cognitive abilities, and language backgrounds has demonstrated the complex interplay among a number of factors and their influence in reading proficiency. This work has also contributed tools and methodologies that could bring a greater degree of consistency across this research, and benefit others who are looking at similar important questions.

There are still many unanswered questions in the search for the factors that contribute to achieving reading proficiency. This thesis takes an important step towards this understanding and contributes lasting measures, methods, and important knowledge in this pursuit.

References

Abidin, H. Z. Z., & Mohamad, M. (2019). Reading and Culture; The Connection. *International Journal*, *1*(2), 48-59. doi: 10.35631/ijmoe.12005

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952-964. doi: 10.1162/jocn.2008.20058

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience research*, *64*(2), 185-190. doi: 10.1016/j.neures.2009.02.013

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

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Arkoudis, S., Hawthorne, L., Baik, C., Hawthorne, G., O'Loughlin, K., Leach, D., and Bexley, E. (2009). The impact of English language proficiency and workplace readiness on the employment outcomes of tertiary international students (Full Report., 157 pp.). Melbourne: Centre for the Study of Higher Education, University of Melbourne.

Arkoudis, S., Baik, C., Bexley, E., & Doughney, L. (2014). English language proficiency and employability framework. *Melbourne: CSHE*.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31-45. doi: 10.1016/j.cortex.2017.02.004

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. http://hdl.handle.net/10125/66684

Bernhardt, E.B. (2011). *Understanding advanced second-language reading*. New York: Routledge.

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics*, *2*(1), 165-92. doi: 10.1257/app.2.1.165

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. doi: 10.1016/j.jml.2019.104047

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663. doi: 10.1037/0278-7393.33.4.663

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide.* New York: Routledge.

Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of educational psychology*, *95*(4), 719. doi: 10.1037/0022-0663.95.4.719

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160-212. doi: 10.1111/lang.12034

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., ... & Johnson, S. P. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental science*, *18*(1), 90-105. doi: 10.1111/desc.12188

Kim, Y. S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from Grade 1 to Grade 2. *Journal of experimental child psychology*, *113*(1), 93-111. doi: 10.1016/j.jecp.2012.03.002

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific reports*, *6*(1), 1-11. doi: 10.1038/srep19741

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental science*, *14*(5), 1100-1106. doi: 10.1111/j.1467-7687.2011.01056.x

Kyröläinen, A. J., & Kuperman, V. (2021). Predictors of literacy in adulthood: Evidence from 33 countries. *Plos one*, *16*(3), e0243763. doi: 10.1371/journal.pone.0243763

Kuperman, V. Siegelman, N., Schroeder, S. … Usal, K. (2021a). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-Movement Corpus (MECO). Manuscript submitted for publication.

Kuperman, V., Siegelman, N., Schroeder, S. … Usal, K. (2021b). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus (MECO). Manuscript submitted for publication.

McCarron, S. P., & Kuperman, V. (2021). Is the author recognition test a useful metric for native and non-native english speakers? An item response theory analysis. *Behavior Research Methods*, 1-12. doi: 10.3758/s13428-021-01556-y

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114-135. doi: 10.1111/j.1467-9817.2010.01477.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in psychology*, *1*, 31. doi: 10.3389/fpsyg.2010.00031

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes*, *18*(4), 443-468. doi: 10.1080/01690960344000008

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, *105*(1), 125. doi: 10.1037/0033-295X.105.1.125

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon mechanical turk for linguistic research. *Psihologija*, *43*(4), 441-464. doi: 10.2298/PSI1004441S

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining “learning” in statistical learning: What does an online measure reveal about the assimilation of visual regularities?. *Cognitive science*, *42*, 692-727. doi: 10.1111/cogs.12556

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311-327. doi: 10.1111/lnc3.12151

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC neuroscience*, *10*(1), 1-8. doi: 10.1186/1471-2202-10-21

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific studies of reading*, *11*(1), 3-32.

Zadeh, Z. Y., Farnia, F., & Geva, E. (2012). Toward modeling reading comprehension and reading fluency in English language learners. *Reading and Writing*, *25*(1), 163-187. doi: 10.1007/s11145-010-9252-0