

Predictors and consequences of individual differences in cross-linguistic interactions: A model of second language reading skill

Brianna L. Yamasaki^{1,2,3} and Chantel S. Prat^{1,2}¹Department of Psychology, University of Washington; ²Institute for Learning and Brain Sciences, University of Washington and ³Department of Psychology and Human Development, Vanderbilt University

Research Article

Cite this article: Yamasaki BL, Prat CS (2021). Predictors and consequences of individual differences in cross-linguistic interactions: A model of second language reading skill. *Bilingualism: Language and Cognition* **24**, 154–166. <https://doi.org/10.1017/S1366728920000279>

Received: 24 August 2018
Revised: 16 March 2020
Accepted: 1 April 2020
First published online: 18 May 2020

Keywords:

bilingual reading; second-language reading; cross-linguistic interactions; executive attention; individual differences

Address for correspondence: Brianna L. Yamasaki,
E-mail: brianna.l.yamasaki@vanderbilt.edu

Abstract

Previous research has demonstrated that individual differences in conflict management predict second-language (L2) reading skill. The current experiment tested the hypothesis that this relation reflects the need to manage conflict from cross-linguistic interactions (CLI). A novel model specifying the relation between L2 reading skill, CLI, and the predictors of such interactions was tested in 253 L2 English speaking adults, using structural equation modeling. In support of the hypothesis, the findings revealed that stronger CLI was related to poorer L2 reading skill. In addition, variability in non-linguistic conflict management, as measured by executive attention tasks, and relative language dominance reliably predicted CLI. Specifically, better conflict management and lower L1 dominance corresponded to fewer interactions. These results fill a crucial gap by demonstrating for the first time that the ability to manage CLI is critical to L2 reading, and that both cognitive skills and language experience contribute to variability in these interactions.

Introduction

Reading is a complex cognitive skill that is critical to functioning in modern society. Consequently, a substantial body of research has focused on improving our understanding of the nature of individual differences in reading (e.g., Bell & Perfetti, 1994; Daneman & Carpenter, 1980; Just & Carpenter, 1992; Long, Prat, Johns, Morris & Jonathan, 2008). Historically, the majority of this work has focused on monolingual readers; although, recently there has been a rise in research investigating individual differences in bilingual reading. One major limitation of this work, however, is that few existing models of bilingual reading account for processes that might be unique to, or particularly important for, bilinguals (see Yamasaki & Prat, 2014 for a counter example).

For example, a plethora of psycholinguistic research has demonstrated that bilinguals co-activate representations in both their first-language (L1) and second-language (L2) during language use. Furthermore, co-activated non-target language representations have been shown to have behavioral consequences on language processing (i.e., cross-linguistic interactions). Specifically, co-activated L1 representations have been shown to lead to a facilitation or inhibition of response times during L2 auditory and visual word recognition, sentence reading, and Stroop paradigms (e.g., Bijeljac-Babic, Biardeau & Grainger, 1997; Chen & Ho, 1986; Fang, Tzeng & Alva, 1981; Mercier, Pivneva & Titone, 2014; Pivneva, Mercier & Titone, 2014; Preston & Lambert, 1969; van Heuven, Dijkstra & Grainger, 1998). Therefore, it follows that variability in the capacity to manage these cross-linguistic interactions may be a unique source driving individual differences in bilingual reading skill. The current study investigates this hypothesis through the development and testing of a novel model of bilingual reading skill focused on individual differences in cross-linguistic interactions.

The bilingual mental lexicon and the nature of cross-linguistic interactions

The work of Dijkstra and colleagues has been foundational in establishing our understanding of the structure of the bilingual mental lexicon and the systems that support visual word recognition during reading. According to the interactive activation models proposed by Dijkstra and colleagues (Bilingual Interactive Activation Model: Dijkstra & van Heuven, 1998; Grainger & Dijkstra, 1992; Bilingual Interactive Activation + Model: Dijkstra & van Heuven, 2002; MultiLink: Dijkstra, Wahl, Buytenhuijs, van Halem, Al-Jibouri, de Korte & Rekké, 2018), the bilingual mental lexicon is composed of a series of hierarchically organized and interconnected nodes representing L1- and L2-relevant orthographic, phonological, and semantic features. Each node is assumed to have a resting-level of activation driven by one's experience with that particular linguistic feature. When a bilingual individual encounters a

word, the input prompts a spreading of activation to any node representing a feature similar to that of the input or other activated nodes. Lexical candidates that ultimately reach an activation threshold are then fed forward to a decision system. The decision system uses contextual information to modulate the activation levels of the competing candidates to ultimately “select” the most relevant representation and facilitate word recognition. Therefore, under these interactive activation models, cross-linguistic interactions occur when both L1 and L2 representations reach the activation threshold and compete for selection.

The likelihood that a particular representation will reach the activation threshold is driven by an interaction between the resting-level of activation in the nodes underlying that representation and the degree to which those nodes share similarity with the input or other activated nodes. A node’s resting-level of activation reflects an individual’s exposure to and usage of that linguistic feature, such that features that are encountered and used more frequently develop a higher resting-level of activation. Words with underlying nodes representing these higher frequency features more rapidly reach the activation threshold, thus increasing the likelihood that they will compete for selection. For bilingual individuals, these higher frequency features are more likely to be associated with their dominant language, typically their L1. This then leads to an asymmetry in how cross-linguistic interactions are experienced. In particular, bilingual individuals often experience more cross-linguistic interactions from their more quickly accessed L1 during L2 processing, than from their more slowly accessed L2 during L1 processing (e.g., Meuter & Allport, 1999; Peeters, Runnqvist, Bertrand & Grainger, 2014; Preston & Lambert, 1969).

Once L1 and L2 representations have reached the activation threshold and been fed forward to the decision system the likelihood that this cross-linguistic co-activation will influence language processing is then driven by the efficacy of one’s conflict management mechanisms. Much of the research exploring the relation between conflict management and cross-linguistic interactions has focused on cross-linguistic interactions that are experienced during spoken word processing. This work has demonstrated that individual differences in executive attention, a cognitive skill supporting conflict management, relate to one’s ability to manage unwanted non-target language intrusions during spoken language production (e.g., Festman, 2012; Festman & Münte, 2012; Festman, Rodriguez-Fornells & Münte, 2010) and comprehension (e.g., Mercier et al., 2014). To date, relatively few studies have directly explored the relation between conflict management and cross-linguistic interactions experience during L2 reading. Yamasaki and Prat (2014) explored the more general relation between variability in conflict management and individual differences in L1 and L2 reading comprehension. Specifically, consistent with previous research demonstrating stronger L1 to L2 cross-linguistic interactions (e.g., Meuter & Allport, 1999; Peeters et al., 2014; Preston & Lambert, 1969), it was hypothesized that conflict management demands would be greatest during L2 reading. To test this hypothesis, Yamasaki and Prat (2014) investigated the relation between individual differences in conflict management, as measured by a Stroop task, and reading skill in English-speaking-monolingual, and L1- and L2-English-speaking-bilingual readers. In line with their hypothesis, the authors found that conflict management was not related to monolingual or L1 reading skill but was related to L2 reading skill. That is, better conflict management uniquely predicted better L2 reading skill.

Pivneva et al. (2014) were among the first to directly investigate a relation between conflict management and cross-linguistic interactions experienced during L2 reading. Specifically, the authors used traditional executive attention tasks, including the Simon, Spatial Stroop, and antisaccade tasks, to index individual differences in conflict management and eye-tracking measures to investigate variability in cross-linguistic interactions experienced during L2 sentence reading. Among other findings, the results demonstrated that when processing interlingual homographs (e.g., words that share orthography but not meaning across L1 and L2) better executive attention was related to reduced cross-linguistic interference.

The current study aims to extend these previous lines of research by investigating the simultaneous contributions of language experience and conflict management on variability in cross-linguistic interactions, and how this variability then contributes to individual differences in L2 reading. In particular, three specific predictions were generated:

Individual differences in L2 reading skill

If cross-linguistic interactions contribute to individual differences in bilingual reading skill, then individuals who experience greater L1 to L2 cross-linguistic interactions should exhibit poorer L2 reading skill. This prediction is in line with results from previous work illustrating that efficiency in managing conflict uniquely contributes to individual differences in L2 reading skill (Yamasaki & Prat, 2014); however, importantly, this previous research did not explore the role of cross-linguistic interactions in this relation.

Individual differences in L1 to L2 interactions

Previous research has shown that variability in the exposure to and usage of one’s L1 and L2 modulates the degree to which cross-linguistic interactions are experienced (e.g., Meuter & Allport, 1999; Peeters et al., 2014; Preston & Lambert, 1969). Consistent with this work, it is predicted that higher L1 dominance will result in more L1 to L2 interactions, as a higher L1 dominance should result in a higher likelihood of the co-activation of L1 lexical alternatives during L2 reading.

The limited existing research exploring the relation between conflict management and cross-linguistic interactions has demonstrated that better conflict management is related to more successful suppression of non-target language intrusions during bilingual language use (e.g., Festman, 2012; Festman & Münte, 2012; Festman et al., 2010; Mercier et al., 2014; Pivneva et al., 2014). Therefore, it is predicted that a similar relation will be observed in the current study. More specifically, it is predicted that poorer conflict management will result in more L1 to L2 interactions, as individuals poorer in this skill would be less efficient at managing co-activated lexical items.

Methods

Three-hundred and twelve individuals (mean age = 19.70 years, female = 62.18%) received course credit for participation in this Institutional Review Board approved study. All individuals completed informed consent procedures before participation.

Table 1. Demographic Information For Study Sample (N = 253)

	N	Mean (SE)	Minimum	Maximum
Female	157			
Age	253	19.7 (0.1)	18.0	26.0
L1 Average Use (0-100)	253	42.3 (1.0)	2.0	78.3
L1 Speaking Proficiency (1-10)	253	9.1 (0.1)	5.0	10.0
L1 Understanding Proficiency (1-10)	252	9.3 (0.1)	6.0	10.0
L2 Age-of-Acquisition	253	6.9 (0.2)	2.0	19.0
L2 Average Use (0-100)	253	56.2 (1.0)	18.3	98.0
L2 Speaking Proficiency (1-10)	252	7.3 (0.1)	5.0	10.0
L2 Understanding Proficiency (1-10)	253	7.7 (0.1)	5.0	10.0

Participant eligibility

Study inclusion criteria were evaluated against participants' self-reported L1 and L2 history. To be included in the analysis, participants had to be proficient in their L1 and L2 (as indexed by a self-reported proficiency of at least five on a scale from 1 to 10). In addition, each participant had to report an L1 of Mandarin, Korean, Spanish, or Japanese and an L2 of English. A modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007) was used to index each participants' language history and to ensure that all participants met the inclusion criteria. Of the total 312 participants, data from 32 (10.26%) were removed due to a missing or incomplete LEAP-Q, 16 (5.13%) due to speaking an inappropriate L1, 8 (2.56%) due to a lack of proficiency in either L1 or L2, 2 (0.64%) due to simultaneous L1 and L2 acquisition, and 1 (0.32%) due to an experimenter error which resulted in the behavioral tasks not being conducted in the participant's L1. Therefore, 253 sequential bilinguals (75.89% = L1 Mandarin; 13.04% = L1 Korean; 7.51% = L1 Spanish; 3.56% = L1 Japanese) were included in the final analysis. Demographic information for the final study sample is displayed in Table 1.

Materials and procedures

Participation consisted of two 1.5 hour sessions that occurred one day apart (e.g., Monday and Wednesday). Of the 312 total participants, 272 (87.18%) completed both sessions. Over the course of the two sessions, participants completed twelve tasks as well as demographic questionnaires. The twelve tasks consisted of three L2 reading tasks (indexing reading at the word-, sentence-, and discourse-level), three executive attention tasks (indexing conflict management; selected from a comprehensive review of canonical executive functioning tasks, see Diamond, 2013), and six cross-linguistic interaction tasks (indexing both cross-linguistic interference and facilitation). Across participants, the tasks were presented in one of four pseudorandomized orders.

L2 reading tasks

Nelson Denny Reading Test

The Nelson Denny Reading Test (NDRT; Brown, 1960) is a timed two-part test with subtests that index English vocabulary knowledge and English discourse comprehension skill. The vocabulary subtest is composed of 80 multiple choice questions, each

containing a test word embedded in an opening statement (e.g., "A vivid description is: ") and five potential response options. Participants are given 15 minutes to complete the vocabulary subtest by selecting the word that best completes the statement for each question (e.g., "lively"). For the comprehension subtest, participants are given 20 minutes to answer 38 comprehension questions distributed across five passages. After reading each passage, participants work through the associated questions, referencing the passage as needed. Two standardized versions of the NDRT were administered across participants, each consisting of an independent set of questions for both the vocabulary and comprehension subtests. The number of correct answers was calculated independently for each subtest. These scores were then used to determine percentile scores based on normed percentiles for college readers.

Homograph Sentence task

The Homograph Sentence task was modified from Gernsbacher, Varner and Faust, 1990. In this task, participants read English sentences, 3-6 words in length (mean length = 4.24 words), and then made relatedness judgments to probe words presented after a short (100ms; 50% of trials) or long (850ms; 50% of trials) delay following the sentence. Participants completed 160 trials, of which 80 had related probe words and 80 had unrelated probe words. Of the 80 unrelated trials, 40 comprised the critical condition in which the sentence final word consisted of an ambiguous homograph. All homographs were balanced (with equally frequent meanings, see p. 439 in Gernsbacher et al., 1990); however, the sentence context biased a particular meaning (e.g., He dug with the spade.). In the critical condition, the probe word always corresponded to the alternative, context-inappropriate meaning of the homograph (e.g., ACE). The remaining 40 unrelated trials comprised the control condition, in which the sentence final word was unambiguous (e.g., He dug with a shovel.). On each trial, a fixation was presented for 850ms, followed by each word of the sentence presented one word at a time. Each word was presented at a rate of 300ms + 16.7ms multiplied by the number of letters in the word with a 150ms ISI. After a delay (100ms or 850ms), the probe word was presented for 2000ms (or until a response was made). Participants received accuracy feedback (presented for 1500ms) following each response. Two versions of the task were created such that the condition the probe word was associated with (e.g., critical or control) varied across versions. Critical and control probe words in both versions and

across delay conditions were balanced for length, frequency, and number of syllables ($ps > .226$). Average accuracy to probe words was computed and an effect size was calculated by taking the difference between accuracies on control trials (unambiguous trials) and critical trials (inappropriate homograph trials) at the long delay (where individual differences have been shown to be greater; Gernsbacher et al., 1990).

Executive attention tasks

Simon task

In the Simon task (Simon & Rudell, 1967), participants respond to visually presented shapes according to a specific task rule (e.g., if circle, then press right). On each trial, a shape was presented on the left (50% of trials) or right (50% of trials) side of the screen. On 75% of the trials, the location of the response indicated by the task rule corresponded to the presentation side of the stimulus (e.g., circle presented on the right-hand side of the screen; congruent trials). On the remaining 25% of trials, the location of the response was opposite of the stimulus presentation side (e.g., circle presented on the left-hand side of the screen; incongruent trials). After 8 practice trials, participants completed 60 experimental trials (45 congruent, 15 incongruent). Each trial began with a fixation for 800ms and a blank preparation screen for 250ms, followed by the stimulus for 3000ms (or until a response was made). Two versions of the Simon task were created by randomizing the trials between versions. Average reaction time was calculated separately for the congruent and incongruent trials. Effect sizes were calculated by subtracting incongruent and congruent reaction times.

Flanker task

The structure, trial composition, and dependent variable calculation on the Flanker task (Eriksen & Eriksen, 1974) was consistent with the Simon task. However, in the Flanker task, participants were presented with a series of five arrow symbols (e.g., < or >). On congruent trials, all five stimuli faced the same direction (e.g., < < < < <), whereas on incongruent trials, all five stimuli excluding the center symbol faced the same direction (e.g., < > < < <). On all trials, participants were instructed to respond with a right or left button press corresponding to the direction of the center symbol.

Spatial Stroop task

The Spatial Stroop task (Shor, 1970) mirrored the Simon (and Flanker) task in structure, trial composition, and dependent variable calculation. In the Spatial Stroop task, participants respond to arrows presented laterally on the screen. Participants were instructed to press the button that corresponded to the side of the screen the arrow was presented on (e.g., arrow on right, press right button). On congruent trials, the arrow was presented such that the orientation of the arrow corresponded to the correct response (e.g., right facing arrow presented on the right-hand side of the screen). On incongruent trials, the arrow was presented such that the orientation was opposite of the correct response (e.g., right facing arrow presented on the left-hand side of the screen).

Linguistic stimuli selection

Linguistic stimuli for all of the cross-linguistic interaction tasks were selected from the University of South Florida Word

Association, Rhythm, and Word Fragment Norm Database (Nelson, McEvoy & Schreiber, 1998). Two-thousand three-hundred and sixty-seven cue-target pairs, with forward strengths between 0.3 and 1.0, were taken from the database. From those, 325 pairs were randomly sampled. A multi-step process was used to acquire the L1 cue for each pair. First, all cues were translated using Google Translate (into Mandarin, Japanese, Korean, or Spanish). Second, a native speaker of each L1 ensured that the translations were correct and removed items that did not have a translation in the L1. Third, a native speaker of each L1 evaluated the relatedness between the L1 cue and English target and categorized each pair into one of 6 categories: (1) cue and target words are related in the L1, (2) cue and target words are not related in the L1, (3) cue and target words are related in the L1 but only in particular contexts, (4) cue and target words have the same translation in the L1 (e.g., YELL - SCREAM, YELL and SCREAM are the same word in Mandarin), (5) target word contains the cue word (e.g., HAND - FINGER, FINGER in Mandarin is "hand point"), or (6) other problem with the cue and target pair (e.g., MAN - WOMAN, WOMAN in Mandarin also means "human"). Word pairs that were categorized under condition 1 were distributed among the related conditions in the linguistic interaction tasks (excluding the Color-Word task, which used four color words and their L1 translations). Word pairs that were categorized under the other conditions were distributed among the filler conditions in the linguistic interaction tasks (excluding Color-Word). Across all linguistic interaction tasks (excluding Color-Word), degree of prime-target relatedness (in English) for the semantically related pairs in the critical condition was balanced across the two task versions ($ps > .186$).

L1 to L2 interaction tasks

Lexical Decision task

In the Lexical Decision task, participants were presented with L1 and English (L2) letter strings. Participants were instructed to press a button according to whether the presented letter string was a word or a nonword (25% English words, 25% L1 words, 25% English nonwords, and 25% L1 nonwords). Nonwords were created based on previous research utilizing nonword stimuli in one of the five languages used in the current study (English, Mandarin, Japanese, Korean, or Spanish). The language of the nonwords was specified by the orthographic and phonologic nature of the nonwords. The two primary conditions of interest consisted of prime-target pairs in which English target words were preceded by either a semantically related L1 prime word or a semantically unrelated L1 prime word or nonword. On each trial, a fixation cross was presented for 750ms followed by a letter string, which remained on the screen until a button response was detected. Two versions of the task were created such that all trials remained the same except for those consisting of a related prime-target pair, in which the language of the prime (L1 or English) was switched across versions. Effect sizes were calculated by subtracting reaction times to English targets preceded by a related L1 prime from English targets preceded by an unrelated L1 prime.

Word Naming task

In the Word Naming task, participants were instructed to verbally name each English (50%) and L1 (50%) word presented on the screen. The two primary conditions of interest consisted of prime-target pairs in which English target words were preceded by either

a semantically related L1 prime word or a semantically unrelated L1 prime word. Each trial consisted of the presentation of the word until a verbal response was detected. Two versions of the task were created in which all trials remained the same, except for the related prime-target trials in which the language of the prime (L1 or English) was switched across versions. Effect sizes were calculated by subtracting reaction times to English targets preceded by a related L1 prime from English targets preceded by an unrelated L1 prime.

Word Identification task

In the Word Identification task (van Heuven et al., 1998) participants were sequentially presented with word pairs. Participants were instructed to press the spacebar when they had identified each English (50%) or L1 (50%) word. After a 1000ms fixation, the first word in the pair (e.g., the prime) was presented for 2000ms (or until a button response was made). Immediately following a button press, participants were prompted to verbally name the prime word. Then, the second word in the pair (e.g., the target) was presented in an alternating pattern with a mask (#####) for 300ms. The word was initially presented for 15ms followed by a 285ms mask. On each successive presentation of the word and mask, the duration of the word increased (in 15ms increments) and the duration of the mask decreased (in 15ms increments). This alternating pattern continued until the word was presented for 300ms or a button response was made. Following the target word, a 500ms mask was presented before participants were prompted to verbally identify the previous target word. Prime-target pairs in which an English target word was preceded by either a semantically related L1 prime or a semantically unrelated English prime comprised the two primary conditions of interest. The language of the related prime (L1 or English) was switched across two versions of the task (with all other trials remaining the same across versions). Effect sizes were calculated by subtracting average reaction times to English targets preceded by a related L1 prime from English target preceded by an unrelated English prime.

Color-Word task

In the Color-Word task (Preston & Lambert, 1969), participants verbally responded (in English) to visually presented letter strings in accordance with their font color. On 20% of the trials, the letter strings consisted of a series of X's (e.g., neutral trials). On 40% of the trials, the letter strings consisted of English color words that were incongruent with the to-be-named font color. On the remaining 40% of the trials, the letter strings consisted of L1 color words that were incongruent with the to-be-named font color. On each trial, the letter string was presented until a verbal response was detected. Two versions of the task were created in which the trials were randomized across versions. Average reaction time was calculated for each condition and effect sizes were calculated by subtracting reaction times to neutral trials from incongruent trials with an L1 color word.

Word-Word task

In the Word-Word task, participants were presented with pairs of letter strings and were instructed to name aloud the lowercase English word (regardless of the language or composition of the other letter string presented). On 33% of the trials, the lowercase English target word was presented with a letter string that consisted of a series of X's (e.g., neutral trials). On 33% of the trials, the lowercase English target word was presented with a

semantically related uppercase English word. On the final 33% of trials, the lowercase English target word was presented with a semantically related L1 word. The lowercase English target word was presented equally often on the left and right side of the letter string pairs. Each trial consisted of the simultaneous presentation of the letter string pairs, and was not terminated until a verbal response was detected. The language of the semantically related distractor (L1 or English) was switched across two task versions. Average reaction time was calculated for each condition and effect sizes were calculated by subtracting reaction times to neutral trials from trials with a semantically related L1 distractor word.

Picture-Word task

In the Picture-Word task (Hentschel, 1973) participants were presented with black-and-white line drawn pictures with red letter strings printed in the upper right-hand corner of the image. Participants were instructed to verbally name the item depicted in the line drawn picture. On 50% of the trials, the line drawn picture was presented with a letter string that consisted of a series of X's (e.g., neutral trials). On 25% of the trials, the line drawn picture was presented with a semantically related English word. On the remaining 25% of trials, the line drawn picture was presented with a semantically related L1 word. Each picture-word pair was presented until a verbal response was detected. Two task versions were created in which the language of the semantically related distractor (L1 or English) was switched across versions. Average reaction time was calculated separately for each condition. Effect sizes were calculated by subtracting reaction times to neutral trials from trials with a semantically related L1 distractor word.

Language Experience and Proficiency Questionnaire (LEAP-Q)

In addition to being used to determine study eligibility, responses on the LEAP-Q (Marian et al., 2007) were also used to index three language experience variables. A participant's self-reported speaking and understanding proficiency in their L1 and L2 were used to calculate two proficiency ratios. More specifically, relative speaking proficiency was measured by subtracting self-reported L2 speaking proficiency from L1 speaking proficiency. Similarly, relative understanding proficiency was measured by subtracting self-reported L2 understanding proficiency from L1 understanding proficiency. Across participants, there was a significant difference in both speaking and understanding proficiency between L1 and L2, with L1 proficiency being reportedly higher than L2 ($ps < .001$). Thus, higher values on either of the proficiency ratio measures indicated a more L1 (as compared to L2) proficient profile. In addition, participants' self-reported percentage of average use in L1 was used to index relative language usage (participants were instructed to report average percentage of use for each language and to ensure that values summed to 100% across languages). Higher values on this usage measure indicated a more L1 dominant language use profile.

Data analysis

Data cleaning

For all computerized behavioral tasks (e.g., all tasks excluding the NDRT), a three-step data cleaning procedure was conducted. First, at the individual trial level, reaction times to incorrect trials and reaction times that exceeded 3 +/- standard deviations from the mean (calculated on correct trials only) were removed before

Table 2. Task Descriptives

	N	Mean (SE)	Minimum	Maximum
Nelson Denny Comprehension*	235	39.6/207.1 (1.5/1.3)	1.0/141.0	96.0/245.0
Nelson Denny Vocabulary*	238	27.6/199.4 (1.4/1.2)	1.0/151.0	95.0/252.0
Homograph Sentence Task [^]	228	19.7 (0.8)	-20.0	50.0
Simon Task [^]	246	79.9 (2.3)	-23.4	201.8
Spatial Stroop Task [^]	243	92.0 (2.7)	7.1	296.6
Flanker Task [^]	238	57.7 (2.7)	-59.5	299.5
Word-Word Task [^]	244	-33.5 (3.8)	-207.9	162.4
Color-Word Task [^]	237	86.8 (4.2)	-69.9	280.2
Picture-Word Task [^]	245	70.8 (6.3)	-199.5	432.7
Word Naming Task [^]	246	18.3 (2.9)	-250.6	143.8
Word Identification Task [^]	228	12.3 (0.9)	-16.0	50.8
Lexical Decision Task [^]	245	20.9 (5.2)	-234.0	396.7

Note. * = Percentile Score/Scaled Score; += Accuracy Effect; ^ = Reaction Time Effect (in ms).

calculating condition means. Next, at the individual participant level, a participant's data for a particular task was removed from further analysis if any of the relevant conditions for a dependent variable had less than three, usable, individual trials (this resulted in the exclusion of 0.5% of the data) or the participant's overall task (collapsed across conditions) performance was below chance (for naming tasks, in which it is difficult to accurately calculate chance, a 50% overall accuracy cutoff was used; this resulted in the exclusion of 0.3% of the data). Finally, at the group level, a participant's data for a particular task was removed from further analysis if the value of their dependent variable for that task was greater than 3 +/- standard deviations from the group mean for that task (this resulted in the exclusion of 1.1% of the data).

Structural equation modeling

The hypothesized relations between L2 reading skill, L1 to L2 cross-linguistic interactions, and the predictors of those interactions were tested through a structural equation model. Structural equation modeling is beneficial as a statistical approach as constructs of interest are analyzed as latent variables. Estimation using latent variables, unlike other statistical methods (e.g., bivariate correlations), allows for an examination of estimated relations unaffected by random measurement error. While previous research has utilized similar statistical techniques to understand individual differences in L2 reading skill, no studies to date have examined the role of the primary construct of interest in the current study (e.g., L1 to L2 interactions). In the present study, four factors (and their associated paths) were estimated: (1) Relative Language Dominance, (2) Executive Attention, (3) L1 to L2 Interactions, and (4) L2 Reading. To control for differences in variance scale, all dependent variables were normalized before being entered into the model. The lavaan package (Rosseel, 2012) in R (R Core Team, 2013) was used to conduct a confirmatory factor analysis and estimate the path model. Parameter estimations were made using full information maximum likelihood. Five fit indices were examined to evaluate overall model fit: Chi-Square (X^2), Comparative Fit Index (CFI), Tucker-Lewis Index (TFI), Root Mean Square Error of

Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). Good model fit is indicated by a Chi-Square with a significance value greater than .05, CFI value greater than .95, TFI value greater than .90, RMSEA value less than .06, and SRMR value less than .08 (Hu & Bentler, 1999).

Results

Task descriptives and bivariate correlations

Means, standard errors, and ranges for all task dependent variables are displayed in Table 2. Bivariate correlations for all normalized task dependent variables are displayed in Table 3.

Structural equation model

To test the study hypotheses, the structural equation model displayed in Figure 1 was estimated. All of the goodness-of-fit indices (X^2 (86) = 106.36, p = .068; CFI = .97; TLI = .96; RMSEA = .03; SRMR = .06) revealed a well-fitting model. An evaluation of the factor loadings (see Figure 2) revealed that higher values on the Relative Language Dominance factor reflected a higher degree of relative L1 dominance over L2, higher values on the Executive Attention factor reflected a larger degree of conflict experienced (or poorer conflict management), higher values on the L1 to L2 Interaction factor reflected less influence from L1 (or greater L2 autonomy), and higher values on the L2 Reading factor indicated better L2 reading skill. Thus to facilitate comprehension, the factors have been relabeled in Figure 2 to reflect direction, with greater scores always reflecting more of what is labeled: Relative L1 Dominance, Non-linguistic Conflict, L2 Autonomy, and L2 Reading Skill. These labels will additionally be maintained throughout the discussion. Finally, the path analysis revealed a significant relation for all estimated paths. Specifically, consistent with the hypotheses tested herein, L2 Autonomy was found to strongly positively predict L2 Reading Skill (β = 0.767) and both Relative L1 Dominance and Non-linguistic Conflict were found to negatively influence L2 Autonomy (β = -0.790 and β = -0.384, respectively).

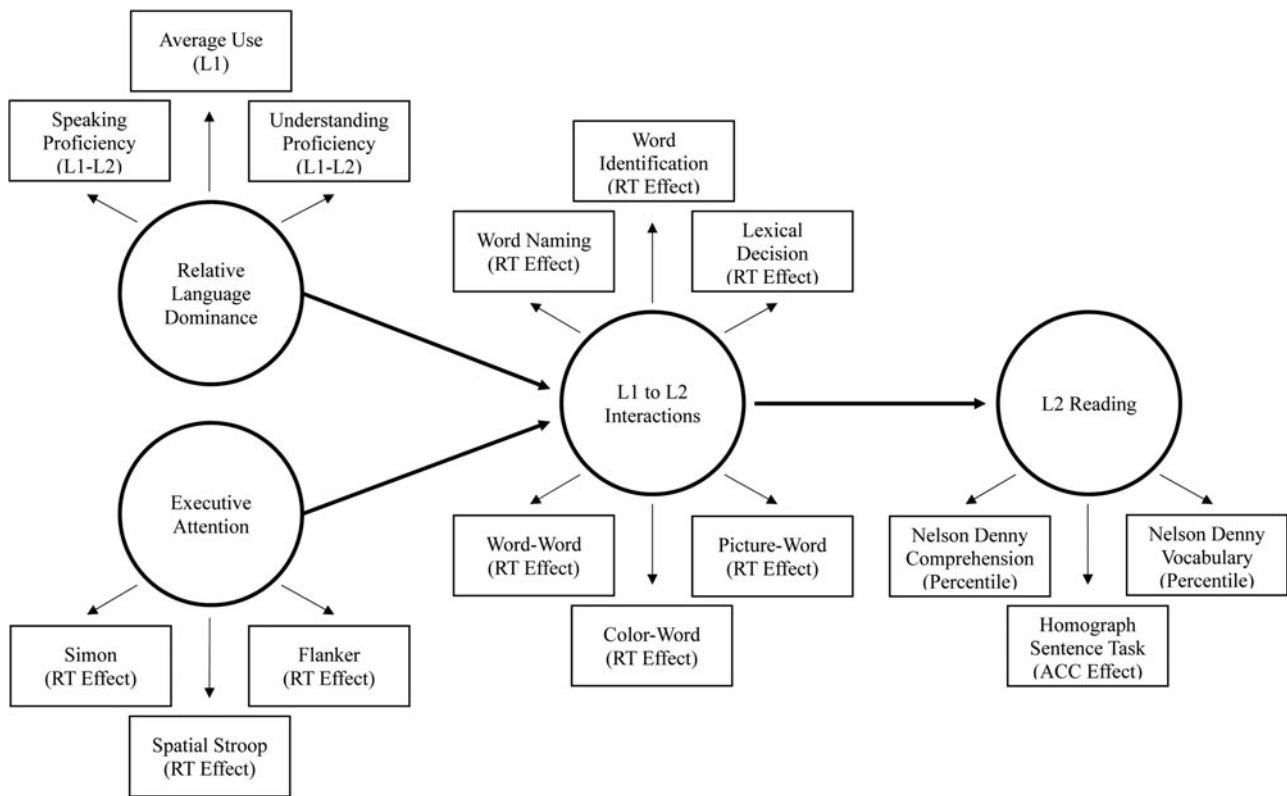


Fig. 1. Hypothesized structural equation model. Latent variables indicated by circles, measured variables indicated by rectangles, and error terms omitted. RT = reaction time; ACC = accuracy.

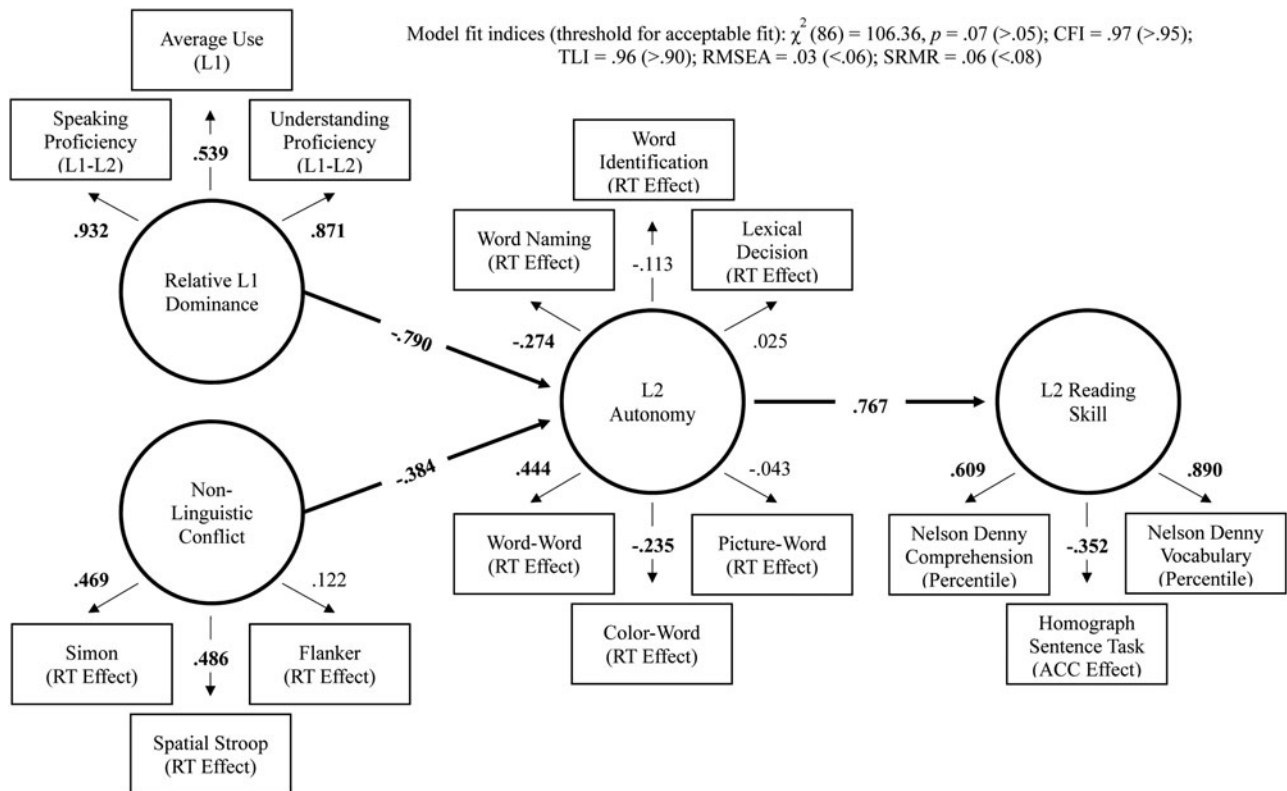


Fig. 2. Standardized coefficients presented for hypothesized structural equation model (significant coefficients are bolded). Latent variables indicated by circles, measured variables indicated by rectangles, and error terms omitted. RT = reaction time; ACC = accuracy; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table 3. Bivariate Correlations

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.539***	-.132*	-.128 [†]	-.177**	-.085	.200**	-.131 [†]	.001	-.118 [†]	.065	.069	-.244***	-.310***	-.297***
2		-.319***	-.065	-.110 [†]	-.142*	.325***	-.119 [†]	-.027	-.158*	-.035	-.015	-.340***	-.478***	-.453***
3			.081	.069	.088	-.157*	.189**	.066	.097	.034	-.017	.206**	.191**	.183**
4				.225***	.073	-.056	.155*	.052	.110 [†]	-.126 [†]	.037	.025	-.034	.031
5					.040	.032	.167*	-.017	.060	-.015	.059	.055	-.021	-.017
6						-.141*	.041	-.070	.030	-.075	.028	.092	.211**	.176**
7							.048	.084	-.231***	-.121 [†]	.013	-.184**	-.353***	-.295***
8								.191**	-.060	-.012	.004	.131*	.176**	.115 [†]
9									-.079	-.087	-.071	.007	.044	.067
10										.115 [†]	-.090	.140*	.182**	.153*
11											.028	.044	.141*	.072
12												.064	-.057	-.023
13													.496***	.458***
14														.808***

Note. 1 = Comprehension, 2 = Vocabulary, 3 = Homograph Sentence, 4 = Simon, 5 = Spatial Stroop, 6 = Flanker, 7 = Word-Word, 8 = Color-Word, 9 = Picture-Word, 10 = Word Naming, 11 = Word Identification, 12 = Lexical Decision, 13 = L1 Use, 14 = L1-L2 Speaking Proficiency, 15 = L1-L2 Understanding Proficiency; [†] = $p \leq .10$, * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$.

Table 4. Unstandardized Factor Loadings (Standard Error)

	L2 Reading Skill	Non-Linguistic Conflict	L2 Autonomy	Relative L1 Dominance
Comprehension	1.00			
Vocabulary	1.47 (0.20)			
Homograph Sentence	−0.58 (0.13)			
Simon		1.00		
Spatial Stroop		1.04 (0.50)		
Flanker		0.26 (0.24)		
Word-Word			1.00	
Color-Word			−0.53 (0.21)	
Picture-Word			−0.10 (0.17)	
Word Naming			−0.62 (0.19)	
Word Identification			−0.26 (0.19)	
Lexical Decision			0.06 (0.18)	
L1-L2 Speaking Proficiency				1.00
L1-L2 Understanding Proficiency				0.95 (0.06)
L1 Average Use				0.58 (0.06)

Unstandardized factor loadings and associated standard errors are displayed in [Table 4](#).

Discussion

Results from this experiment support the novel hypothesis that variability in the strength of L1 to L2 cross-linguistic interactions will predict individual differences in L2 reading. Specifically, it was demonstrated that higher L2 autonomy (or lower L1 to L2 interactions) is correlated with better L2 reading skill. In addition, in line with the predictions generated from the interactive activation models of bilingual language processing (Dijkstra & van Heuven, 1998; Dijkstra & van Heuven, 2002; Dijkstra et al., 2018; Grainger & Dijkstra, 1992), the data also showed that both linguistic and non-linguistic cognitive factors predict variability in cross-linguistic interactions. In particular, it was found that poorer conflict management, as measured by standard, non-linguistic executive attention tasks, predicted stronger L1 to L2 interactions and greater relative L1 over L2 dominance was associated with stronger L1 to L2 interactions.

Understanding the contributions of conflict management on L1 to L2 interactions

According to the interactive activation models of bilingual language processing (Dijkstra & van Heuven, 1998; Dijkstra & van Heuven, 2002; Dijkstra et al., 2018; Grainger & Dijkstra, 1992), when multiple lexical candidates become co-activated, selection mechanisms are necessary to “select” the most relevant alternative. Therefore, it was predicted that better conflict management skills would result in more efficient suppression of the non-target L1 representations and therefore smaller L1 to L2 interactions. In support of this prediction and consistent with the limited research in this area (Pivneva et al., 2014), the results of the model estimated in the current study demonstrated that better conflict management on the executive attention tasks was associated with

greater L2 autonomy (or less L1 to L2 cross-linguistic interactions). It should be noted that the interactive activation models of bilingual language processing do propose that conflict management mechanisms support the word recognition process. However, these models argue, more specifically, that conflict management mechanisms operate on the output of the language system and thus do not directly influence lexical activation with the language system. Unfortunately, it is not possible to determine whether the relation observed in the current study is driven by an influence of executive attention on the spreading of lexical activation between L1 and L2 representations in the language system or on the resolution of conflict between the L1 and L2 outputs of the language system. However, given that at least one other study has shown evidence that individual differences in executive attention predict reading-related measures indexing lexical activation (Pivneva et al., 2014), the current study may provide further evidence for a more robust role of executive attention during reading. This interpretation would be in line with the Inhibitory Control Model (Green, 1998), a more general model that argues for a global role of executive attention during bilingual language use.

Understanding the contributions of language experience on L1 to L2 interactions

Motivated by previous research demonstrating that language experience can modulate the nature of cross-linguistic interactions (e.g., Beauvillain & Grainger, 1987; Bijeljac-Babic et al., 1997; Chen & Ho, 1986; Mägiste, 1984; Preston & Lambert, 1969; Tzelgov, Henik & Leiser, 1990; van Heuven et al., 1998), it was predicted that a bilingual’s relative L1 to L2 dominance would contribute to the strength of cross-linguistic interactions experienced during L2 reading. Consistent with this prediction, it was found that individuals who were more L1 dominant experienced fewer cross-linguistic interactions.

Given that linguistic representations in a bilingual’s L1 and L2 can develop semi-independently, it is possible that absolute

language experience in either L1 or L2 is driving the observed relative language experience effect. This would be consistent with previous models of reading which have shown that both L1 and L2 proficiency relate to L2 reading comprehension (e.g., Hdstijn & Bossers, 1992). Therefore, to test this alternative hypothesis, two additional models were run in which the L1 to L2 proficiency and use variables were replaced by L1 or L2 variables individually. Both models were able to be estimated: however, in both cases only 2 or 3 of the 5 goodness-of-fit indices indicated acceptable model fit (L1 model: $X^2(86) = 121.98$, $p = .007$; CFI = .92; TLI = .90; RMSEA = .04; SRMR = .07; L2 model: $X^2(86) = 117.77$, $p = .013$; CFI = .94; TLI = .93; RMSEA = .04; SRMR = .06). Given the poor model fit, interpretability is limited; however, a preliminary evaluation of the relation between the Language Proficiency (formerly Relative L1 Dominance) factor and the L2 Autonomy factor revealed a significant association for both models. As might be expected, the L1 model indicated that higher L1 proficiency was associated with stronger L1 to L2 interactions and the L2 model indicated that higher L2 proficiency was associated with weaker L1 to L2 interactions. Taken together, the fact that the model incorporating relative dominance of L1 and L2 fit the data well, while the models with either L1 or L2 proficiency alone were poor fitting suggests that the relative language dominance profile seems to explain cross-linguistic interactions better than individual proficiency in either language alone.

Understanding the role of linguistic interactions in L2 reading

Only a handful of studies have used multivariate analyses or latent variable modeling to understand individual differences in L2 reading skill; while some models have included both L1 and L2 variables (e.g., L1 or L2 proficiency), none of the previous work has considered interactions *between* L1 and L2. Thus, to the best of our knowledge, the current study is the first to consider how these cross-linguistic interactions contribute to individual differences in L2 reading. As predicted, the results of the model tested herein support the hypothesis that individuals who experience stronger L1 to L2 interactions display poorer L2 reading skills. Given that cross-linguistic interactions can manifest behaviorally as either interference (e.g., as indexed by slower reaction times on conflict trials on the Stroop task) or as facilitation (e.g., as indexed by faster naming times for related targets on the Word Naming task), one might have predicted that only interference effects would contribute to poorer L2 reading skill. However, this was not the case. Importantly, both tasks eliciting interference and facilitation (priming) effects loaded significantly onto the L1 to L2 interaction factor suggesting that any type of cross-linguistic interaction can impair L2 reading.

In the current study, the observed relation between L2 Autonomy and L2 Reading Skill was interpreted under the hypothesis that managing cross-linguistic interactions places conflict management demands on the reader. Thus, individuals who experience more L1 to L2 interactions have fewer resources available to perform other reading-related processes and therefore display poorer L2 reading skill. However, an alternative interpretation of this relation is that it is driven not by increased demands associated with managing L1 to L2 interactions, but instead by individual differences in reading speed. In particular, slower reading speeds on the L1 to L2 interaction tasks would result in slower reaction times and therefore more time for L1 and L2 co-activation to occur and cross-linguistic interactions

to emerge. Similarly, slower reading speeds on the L2 reading tasks would result in less time to complete the tasks and therefore overall poorer performance. Thus, the significant relation observed between more L1 to L2 interactions and poorer L2 reading skill could have been a consequence of variability in reading speed. To evaluate this alternative interpretation, additional correlational analyses were conducted in which two measures of reading speed were correlated with a measure of L1 to L2 interactions. Specifically, participants' reading rate (based on the number of words read in one minute) was measured at the start of the comprehension subtest of the Nelson Denny Reading Test. This reading rate measure served as the first measure of reading speed in the correlation analyses. In addition, among the cross-linguistic interaction tasks, the trials in which participants named English words preceded by unrelated stimuli on the Word Naming task provided the purest and most direct measure of online L2 (English) reading speed and thus this naming speed measure was used as the second measure of reading speed in the correlation analyses. Given that one measure of reading speed was estimated from the Word Naming task, the cross-linguistic interaction dependent variable from this task was selected to serve as the L1 to L2 interaction measure in the correlation analyses. Interestingly, inconsistent with the hypothesis that individual differences in reading speed were driving the observed relation between L2 Autonomy and L2 Reading Skill, results from the correlational analyses demonstrated no significant relation between either measure of reading speed and the degree of L1 to L2 interactions experienced on the Word Naming task (reading rate: $p = .352$; naming speed: $p = .121$).

While the central hypothesis in the current study focuses on cross-linguistic interactions, it is possible that the observed relation between variability in cross-linguistic interactions and L2 reading skill relates more generally to linguistic conflict management (both within and across languages). In fact, considerable research has linked the ability to efficiently select the correct word forms and syntactic structures in the face of competition within a language to reading skill (e.g., Gernsbacher et al., 1990). To directly test the hypothesis that variability in within-L2 interactions also contributes to individual differences in L2 reading skill, an additional model was estimated in which each of the L1 to L2 interaction dependent variables were replaced with a homologous within-L2 interaction variable. Although this model did converge, an examination of the goodness-of-fit indices indicated that the model was of poor fit. In fact, only 2 of the 5 fit indices indicated acceptable fit ($X^2(86) = 140.98$, $p < .001$; CFI = .91; TLI = .89; RMSEA = .05; SRMR = .07) and therefore interpretability of the results is limited. Nonetheless, a preliminary evaluation of the associations between the four factors revealed that greater within-language interference (only tasks that elicited interference loaded onto the Within-L2 Interaction factor) was also associated with better L2 reading. Individual differences in within-language interference were primarily driven by differences in language dominance (the relation between within-L2 interactions and conflict management was marginal), with more L2 dominant individuals experiencing more within-L2 interference. In contrast to the interpretation of the cross-linguistic interaction model, the within-language interaction model appears to reflect the relation between quality of lexical representations and reading skill. According to the Lexical Quality Hypothesis (Perfetti & Hart, 2002), the quality of lexical representations, which contributes to the ease with which lexical items can be accessed, scaffolds up to influence reading skill (Perfetti, 2007). It may be

presumed that individuals who are more dominant in their L2, and thus have more experience using their L2, develop higher quality L2 lexical representations. High quality representations are more quickly accessed, and thus co-activations among related lexical alternatives are more likely to occur. This co-activation could then contribute to higher levels of within-language interactions, as was observed for the individuals in the current study. While this interpretation is in line with a prominent model of reading (e.g., the Lexical Quality Hypothesis; Perfetti & Hart, 2002), it is important to note that the interpretation of this model must be taken with caution as the model was of poor fit. Nonetheless, this follow-up analysis suggests that both within- and between-language interactions may drive L2 reading skill, consistent with the growing body of literature suggesting that conflict resolution processes are particularly important for L2 reading (e.g., Yamasaki & Prat, 2014).

Modeling individual differences in L2 reading in young adults

Of the few existing studies that have used multivariate or latent variable models to understand individual differences in L2 reading skill, many have been conducted with children (e.g., Babayiğit, 2015; Gottardo & Mueller, 2009; Lesaux, Crosson, Kieffer & Pierce, 2010; Proctor, Carlo, August & Snow, 2005; Uchikoshi, 2013). The results of the current study extend this work by examining individual differences in L2 reading skill among relatively proficient young adult readers. In contrast to developing English readers, all participants included in the current study are assumed to have a much higher level of English proficiency as they are all being educated at an English-speaking university. Given their level of English proficiency, it may have been predicted that individual differences in reading would be more limited in this population. However, large individual differences in L2 reading skill were observed in the current study (e.g., performance on the Nelson Denny Comprehension Test ranged from 1-96% based on native-English speaking norms). This finding highlights the fact that individual differences in L2 reading occur not only during L2 reading acquisition, but also during much later stages of L2 reading development. It is unclear, however, whether the sources of individual differences in developing and proficient speakers and readers are the same. Previous research, using auditory and pictorial tasks to examine lexical and semantic organization in bilingual children, has shown that even before a child begins to learn to read, cross-linguistic interactions are observable (e.g., Singh, 2014; Von Holzen & Mani, 2012). Additionally, comparisons between adults and children have shown that both groups show comparable levels of semantic interference within a language (e.g., Rosinski, Golinkoff & Kukish, 1975). Therefore, it might be predicted that L2 reading in children, like the adults tested in the current study, is constrained by cross-linguistic interactions. However, additional work is necessary to confirm this hypothesis, and to determine how cross-linguistic interactions contribute to individual differences in the early developmental stages of L2 reading.

Limitations

Certain facets of this study limit the interpretations that can be drawn from it. First, it is important to note that cross-linguistic interactions were only measured from L1 to L2 in the current study. While this decision was made based on the specific hypothesis being tested, it is unclear from the current experiment the

extent to which cross-linguistic interactions as a whole (including L2 to L1 interactions) predict reading skill. Thus, future models aiming to extend the findings of the current study would benefit from a more complete characterization of cross-linguistic interactions, as well as assessment of both L1 and L2 reading. Additionally, not all of the tasks used herein loaded onto their respective factors. This could have been due to low reliability in a subset of the behavioral tasks. Given the large number of tasks that had to be completed by each participant, there were limitations in the length of time each task could take. Thus, some tasks may have had fewer trials per condition than necessary to establish a reliable estimate of the construct of interest. Alternatively, non-significant factor loadings might have resulted from low construct validity. For example, the Flanker task (which had relatively good reliability) may not have loaded onto the Non-Linguistic Conflict factor because performance on this task could reflect a different executive attention mechanism than that of which is recruited on the Simon and Spatial Stroop tasks. While multi-task models, such as the one tested herein, are highly beneficial in improving our understanding of the complex associations between multiple constructs, future studies may benefit from a more directed analysis of fewer constructs that allows for more reliable estimations of each construct. Finally, language similarity has been demonstrated to contribute to variability in cross-linguistic interactions (e.g., Dyer, 1971; Fang et al., 1981; van Heuven, Conklin, Coderre, Guo & Dijkstra, 2011). In an effort to model this source of variability, L2 English bilinguals with four different language profiles (Mandarin-, Korean-, Spanish-, and Japanese-English bilinguals) were recruited for the current study. However, primarily driven by the demographics of the participant pool, 76% of participants in the current study were Mandarin-English bilinguals. Therefore, the ability to understand the role of languages similarity in cross-linguistic interactions was limited in the current study. Additionally, the extent to which these results are true of all L2 English readers, or are driven by the particular characteristics of the languages spoken by our participants, remains to be seen. To resolve this uncertainty, future investigations should consider testing the present model with a more heterogeneous group of L2 readers.

Conclusion

The model tested herein represents the first model of L2 reading skill to be centered on understanding the role of cross-linguistic interactions. The findings from this study demonstrate that L1 to L2 interactions constrain L2 reading skill, and that variability in these interactions is driven by both non-linguistic conflict management skills and relative language dominance. These results are important because they highlight the centrality of a novel, widely understudied demand that is unique to bilingual readers, and may be particularly burdensome on L2 readers. Additionally, this work provides a foundation on which assessments of L2 reading skill can be made in a way that allows for targeted recommendations for remediation. For instance, the results of this study suggest that an individual might be a poor L2 reader because they are less proficient in their L2 as compared to their L1, or because of deficits in conflict management skills. These two sources of difficulty would call for markedly different interventions. In summary, this research provides a new window for understanding the nature of L2 reading, providing evidence of the centrality of cross-linguistic interactions and outlining ample questions for future research.

Acknowledgements. We would like to thank Brian Flaherty for his guidance on the statistical analyses. The research described in this paper was supported in part by a grant awarded to Brianna L. Yamasaki from the American Psychological Association, a Royalty Research Fund from the University of Washington awarded to Chantel S. Prat, and a grant from the Office of Naval Research awarded to Chantel S. Prat (Grant #N000141410344).

References

- Babayigit S** (2015) The relations between word reading, oral language, and reading comprehension in children who speak English as a first (L1) and second language (L2): A multigroup structural analysis. *Reading and Writing*, **28**, 527–544.
- Beauvillain C and Grainger J** (1987) Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, **26**, 658–672.
- Bell LC and Perfetti CA** (1994) Reading skill: Some adult comparisons. *Journal of Educational Psychology*, **86**, 244–255.
- Bijeljac-Babic R, Biardeau A and Grainger J** (1997) Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, **25**, 447–457.
- Brown JI** (1960) *The Nelson-Denny Reading Test*. Oxford, England: Houghton Mifflin.
- Chen HC and Ho C** (1986) Development of Stroop interference in Chinese–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 397–401.
- Daneman M and Carpenter PA** (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**, 450–466.
- Diamond A** (2013) Executive functions. *Annual Review of Psychology*, **64**, 135–168.
- Dijkstra T and van Heuven WJB** (1998) The BIA model and bilingual word recognition. In Grainger J and Jacobs A (eds), *Localist connectionist approaches to human cognition*. Hillsdale, NJ: Erlbaum, pp. 189–225
- Dijkstra T and van Heuven WJ** (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, **5**, 175–197.
- Dijkstra T, Wahl A, Buytenhuijs F, van Halem N, Al-Jibouri Z, de Korte M and Rekké S** (2018) Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, **22**, 657–679.
- Dyer FN** (1971) Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, **10**, 297–302.
- Eriksen BA and Eriksen CW** (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143–149.
- Fang SP, Tzeng OJ and Alva L** (1981) Intralanguage vs. interlanguage Stroop effects in two types of writing systems. *Memory & Cognition*, **9**, 609–617.
- Festman J** (2012) Language control abilities of late bilinguals. *Bilingualism: Language and Cognition*, **15**, 580–593.
- Festman J and Münte TF** (2012) Cognitive control in Russian–German bilinguals. *Frontiers in Psychology*, **3**, 1–12.
- Festman J, Rodriguez-Fornells A and Münte TF** (2010) Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, **6**, 1–12.
- Gernsbacher MA, Varner KR and Faust ME** (1990) Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 430–445.
- Gottardo A and Mueller J** (2009) Are first-and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology*, **101**, 330–344.
- Grainger J and Dijkstra T** (1992) On the representation and use of language information in bilinguals. In Harris RJ (ed), *Advances in psychology*. Amsterdam, The Netherlands: North-Holland, pp. 207–220.
- Green DW** (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**, 67–81.
- Hentschel U** (1973) Two new interference tests compared to the Stroop Color-Word test. *Psychological Research Bulletin*, **13**, 1–24.
- Hdstijn J H and Bossers B** (1992) Individual differences in L2 proficiency as a function of L1 proficiency. *European Journal of Cognitive Psychology*, **4**, 341–353.
- Hu LT and Bentler PM** (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, **6**, 1–55.
- Just MA and Carpenter PA** (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122–149.
- Lesaux NK, Crosson AC, Kieffer MJ and Pierce M** (2010) Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology*, **31**, 475–483.
- Long DL, Prat C, Johns C, Morris P and Jonathan E** (2008) The importance of knowledge in vivid text memory: An individual-differences investigation of recollection and familiarity. *Psychonomic Bulletin & Review*, **15**, 604–609.
- Mägiste E** (1984) Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 304–315.
- Marian V, Blumenfeld HK and Kaushanskaya M** (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, **50**, 940–967.
- Mercier J, Pivneva I and Titone D** (2014) Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition*, **17**, 89–117.
- Meuter RF and Allport A** (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, **40**, 25–40.
- Nelson DL, McEvoy CL and Schreiber TA** (1998) *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from <http://www.usf.edu/FreeAssociation/>.
- Peeters D, Runnqvist E, Bertrand D and Grainger J** (2014) Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**, 284–292.
- Perfetti C** (2007) Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, **11**, 357–383.
- Perfetti CA and Hart L** (2002) The lexical quality hypothesis. In Verhoeven L, Elbro C and Reitsma P (eds), *Precursors of functional literacy*. Amsterdam: John Benjamins Publishing Co, pp. 67–86.
- Pivneva I, Mercier J and Titone D** (2014) Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**, 787–796.
- Preston MS and Lambert WE** (1969) Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, **8**, 295–301.
- Proctor CP, Carlo M, August D and Snow C** (2005) Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, **97**, 246–256.
- R Core Team** (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>.
- Rosinski RR, Golinkoff RM and Kukish KS** (1975) Automatic semantic processing in a picture-word interference task. *Child Development*, **46**, 247–253.
- Rosseel Y** (2012) Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, **48**, 1–36.
- Shor RE** (1970) The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, **32**, 346–365.
- Simon JR and Rudell AP** (1967) Auditory SR compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, **51**, 300–304.
- Singh L** (2014) One world, two languages: Cross-language semantic priming in bilingual toddlers. *Child Development*, **85**, 755–766.

- Tzelgov J, Henik A and Leiser D** (1990) Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 760–771.
- Uchikoshi Y** (2013) Predictors of English reading comprehension: Cantonese-speaking English language learners in the US. *Reading and Writing*, **26**, 913–939.
- van Heuven WJ, Conklin K, Coderre EL, Guo T and Dijkstra T** (2011) The influence of cross-language similarity on within-and between-language Stroop effects in trilinguals. *Frontiers in Psychology*, **2**, 1–15.
- van Heuven WJ, Dijkstra T and Grainger J** (1998) Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, **39**, 458–483.
- Von Holzen K and Mani N** (2012) Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, **113**, 569–586.
- Yamasaki BL and Prat CS** (2014) The importance of managing interference for second language reading ability: An individual differences investigation. *Discourse Processes*, **51**, 445–467.