9-21-2018

The Word Frequency Effect: Relationship of Lexical Entries Between the Primary and Secondary Language

Nawal Mustafa
University of Windsor

The Word Frequency Effect: Relationship of Lexical Entries Between the Primary and Secondary Language

by

Nawal Mustafa

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Master of Arts
at the University of Windsor

Windsor, Ontario, Canada

2018

The Word Frequency Effect: Relationship of Lexical Entries Between the Primary and Secondary Language

by

Nawal Mustafa

APPROVED BY:

C. Miller
Department of Psychology

C. Rangan
Department of Physics

L. Buchanan, Advisor
Department of Psychology

August 23, 2018

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution

ABSTRACT

The current study examined the effect of differential word frequency on the relationship of lexical entries between the primary and secondary language. Ninety Urdu-English bilingual participants were used, and their performance was compared to forty-five English monolinguals matched for age and education. The task for both participant groups was a lexical decision task with 60 high and 60 low frequency English words. The stimulus set consisted of four frequency conditions High English-High Urdu, High English-Low Urdu, Low English-High Urdu and Low English-Low Urdu. A general frequency effect was observed – all participants responded faster to high frequency targets than low than low frequency target words. There was also a main effect of language experience with bilinguals producing longer reaction times than monolinguals. In addition, a frequency effect was observed in response times for high frequency English words as a function of their Urdu pair frequency. These results reveal a cross language frequency differential effect consistent with models proposing non-selective access of lexical processing in bilinguals.

ACKNOWLEDGEMENTS

To Dr. Lori Buchanan: Thank you for your valuable feedback and guidance and for all the time you invested in me.

To Dr. Carlin Miller and Dr. Chitra Rangan: Thank you for your contributions and your genuine curiosity towards my work.

To members of the Buchanan Lab and my classmates: Thank you for your support and feedback

To my family: Thank you for your love, unconditional support, and for taking pride in my accomplishments.

To Ragheb: Thank you for always believing in me and encouraging me to be the best version of myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

REVIEW OF LITERATURE

Language comprehension requires the activation and identification of contextually suitable mental representations with speed and accuracy. This may be particularly challenging for bilinguals because their primary and secondary languages may have a potential for interference and opposition (Ibrahim, Cowell, & Varley, 2017). Bilinguals have approximately twice as many words in their mental lexicon compared to monolinguals, but they rarely say a word in the unintended language (Gollan, Sandoval, & Salmon, 2011) and are generally unaware of possible word level competition between the two languages. This has encouraged scholars to investigate the underlying mechanisms of the cross-lingual interaction in bilinguals by comparing attributes of language processing between the mother tongue (L1) and the second language (L2).

One line of research in the bilingual literature focuses on the nature of lexical access. During word comprehension, an individual is presented with a lexical word form and must determine its meaning. In reading, this process begins first with the identification of the letter features of the word, then the letters themselves, and finally the meaning of the word (Litcofsky, Tanner, & van Hell, 2016; Adelman, 2012). Most current models consider lexical information in the form of graphemes (visual form), phonemes (sound), or semantics (meaning) to be engaged during word recognition and comprehension (Adelman, 2012; Balota, Yap, & Cortese, 2006; Dijkstra, 2005, p.180).

A central question of visual word recognition in bilinguals is whether a word (printed or spoken) activates a single meaning representation for both L1 and L2, or whether there are two distinct representations. Bilinguals may manage this cross-language interference using one of two hypothetical cognitive mechanisms. These mechanisms can be broadly thought of as the selective access hypothesis and the non-selective access hypothesis. The language selective view holds that bilinguals activate lexical representations from the contextually appropriate language only, meanwhile inhibiting representations from the non-target language (Gerard & Scarborough, 1989; Dijkstra, 2003). Initially, this highly selective system activates candidates solely from the target language, and contact is established with the non-target lexicon as a last resource when the search for the corresponding unit in the target language does not match (Dijkstra, 2003). The non-selective access hypothesis proposes that bilinguals activate lexical representations from both languages in parallel and subsequently suppress representations from the non-target language (Dijkstra, Timmermans, & Schriefers, 2000). If bilinguals simultaneously activate both languages, then how do they control potential cross-language intrusions to allow fluent performance? A number of theories have been proposed to explain this mechanism.

Theories on Bilingual Language Processing

The earliest account of bilingual representation and translation was proposed by Weinreich (1968) who suggested that the answer may differ depending of the type of bilingual. He proposed three types of bilingualism: coordinative, compound, and subordinative. In coordinate bilinguals, the words of their two languages would be kept in separate memory stores. For example, the word *cat* in English and its French equivalent *chat* would have their own, distinct form and meaning. Compound bilinguals, in contrast, would have only one shared memory store for *cat* and *chat*. Lastly, subordinate bilinguals would process their L2 words through the influence of L1, their stronger language (Weinreich, 1968). Weinreich's (1986) distinction of bilingualism led subsequent research to focus on the modes of interconnection between L1 and L2 at lexical and conceptual levels of representation. Potter, So, Van Eckardt, and Feldeman (1984) submitted two hierarchical models to account for this interaction: Word Association model and Concept Mediation Model, the suitability of which they determined by comparing bilingual performance in picture-naming and word translation tasks. The Word Association translation route indirectly accesses the L2 word after first translating it into the existing L1 language system. So, for example, the French word form *chat* would immediately be identified as the English-dominant word form *cat* via an orthographic or phonological connection. On the other hand, the translation procedure of concept mediation has no direct links between L1 and L2 words, but rather these words are linked through a shared conceptual system; thus, *cat* and *chat* would be accessed directly. Potter et al. (1984) conducted two experiments, one with proficient Chinese-English bilinguals and the second with non-fluent English-French bilinguals. The participants completed word-reading, picture-naming, and word-translation tasks. Results revealed that the participants used approximately the same speed to name pictures and translate words, providing support in favor of the concept mediation model.

Subsequent research, however, presented evidence for the word association link in novice bilinguals. For example, Kroll and Curley (1988) who replicated Potter et. al.'s (1984) study with bilinguals of ranging proficiency found that participants who learned their second language within the last two years performed in line with the Word Association model, whereas more proficient bilinguals showed results in correspondence with the Concept Mediation model. In light of these new findings, Kroll and Stewart (1994) proposed the Revised Hierarchical Model (RHM) which integrates the word association and concept mediation translation routes and accounts for the changes in bilingual lexical representations that are a result of increasing L2 proficiency. According to this model of selective-access, translation from the separate L1 to L2 lexicons may occur indirectly through conceptual representations because of the strong L1 link to meaning, thus taking more time than the translation from L2 to L1 which proceeds via word association and without semantic access (Kroll & Stewart, 1994; Kroll, van Hell, Tokowicz, & Green; 2010).

The RHM is among the few developmental models in the bilingual mental lexicon literature that incorporates a model for less proficient L2 speakers and a transition to a higher level of proficiency. Despite this strength, the model has come under attack because it does not take into account the involvement of additional factors, like cognate status, concreteness, and the nature of the mapping between L1 and L2 (De Groot & Poot, 1997; Sunderman & Kroll, 2006; Grainger, Kiyonaga, & Holocomb, 2006; Brysbaert & Dyck, 2010). Some scholars believe RHM is no longer a useful characterization of language processing mechanisms in bilinguals (Brysbaert & Dyck, 2010). Instead, the Bilingual Interactive Activation Plus (BIA+) model is considered to better capture the complex interaction between the multiple underlying causal mechanisms in bilingualism (Dijkstra & Van Heuven, 2002).

According to the BIA+ model, a word stimulus causes matching lexical representations to resonate in long-term memory regardless of what language the stimulus is presented in; that is when a word is presented, orthographic and phonological information related to that concept is initially activated for both languages (Dijkstra & Van Heuven, 2002). Activation of lexical word form candidates depends on word frequency, language proficiency, and recency of use, iregardless of language membership; for example, the English word *cat* and its French pair *chat* will take the same amount of time to activate in a balanced-bilingual if their frequency and recency of use is equal in both languages. Subsequently, associated semantic representations become active and compete for selection. BIA+'s predecessor, the Bilingual Interactive Activation model, consisted of language nodes with excitatory and inhibitory connections. These nodes identify each word's language membership and when activated, they inhibit all words from the opposite language. The BIA+ model retained these nodes to represent the language membership of the target word, but removed their inhibitory role in lexical processing (Dijkstra & Van Heuven, 2002). BIA+ also assumes that neither the word recognition system (language nodes) nor any task-related processes affect any stage of lexico-semantic processing. Thus, only bottom-up processes (such as frequency and sentence context) initially drive visual word recognition and the influence of top down processes (such as task demands) occurs later.

Research on the Non-selective View

Research investigating the nature of bilingual lexical access has predominantly focused on cross-language ambiguity involving inter-lingual homographs (IHs) and cognates. IHs share orthographic form, but correspond to a different semantic representation (e.g., the English word *ramp* refers to "disaster" in Dutch); whereas, cognate words have an orthographic, semantic, and often phonological overlap across the two languages (e.g., *film* refers to the same concept in both Dutch and English). IHs appear to slow down semantic processing in bilinguals, presumably because the reader makes a choice between the two unique meanings of the word (Dijkstra, Grainger, and van Heuven, 1999). On the other hand, cognates, due to representation of linked memory, show a facilitation effect in word recognition when compared to language unique words with the same frequency.

Such cognate facilitation effects and semantic impedance of homographs have been reported for a variety of experimental tasks, such as word naming, picture naming, and word recognition (De Groot, Borgwaldt, Bos & Van den Eijnden, 2002; Hoshino & Kroll, 2008; Costa, Caramazza, & Sebastian, 2000; Dijkstra, Timmermans, & Schriefers, 2000). Most of these studies have found some degree of parallel activation in proficient bilinguals, as predicted by non-selective access accounts. For example, Costa, Caramazza, and Sebastian-Galled (2000) asked highly proficient Catalan-Spanish and Spanish-Catalan bilinguals to name pictures with cognate names and those with non-cognate names. They found that bilinguals named cognate pictures faster than noncognate pictures in both the non-dominant L2 and the dominant L1. Another study by Brenders, Van Hell, and Dijsktra (2011) also found evidence of cognate facilitation through a lexical decision task, where the participant identifies a string of letters as a word or a non-word as quickly and accurately as possible. They presented cognates and non-cognates in L1 and L2 to native Dutch children with different levels of English (L2) proficiency: beginning classroom learners of English (grades 5 and 6) and advanced classroom learners of English (grade 7 and 9). All participants processed cognates faster than matched control groups for English lexical decision tasks, but not for Dutch lexical decision tasks. The presence of a cognate effect in the participants' L2 but not in L1 suggested that differences in L2 proficiency can modulate the magnitude of cognate effects in word recognition. As for interlingual homographs, Dijkstra and colleagues (2000) used a go/no-go paradigm to test Dutch-English bilinguals. The participants were instructed to press a button only if an English word was presented, and to wait for the next word if a Dutch word appeared. A reliable homograph effect was obtained by the researchers: the participants exhibited slower reactions times to identify English homographs than English non-homographs despite the English reading of the homograph being higher in frequency than the Dutch reading.

There is also an increasing amount of data supporting the non-selective access view in bilinguals of cross-script languages (Gollan, Forster, and Frost, 1997; Kim and Davis, 2003; Mishra and Singh; 2014; Khan and Buchanan, 2013). As an example, Kim and Davis (2003) tested Korean-English unbalanced bilinguals using a lexical decision task to determine cognate and non-cognate priming effects for words. They reported significant effects for both, suggesting that a word in one language still activated its cognate in another, orthographically distinct language. Nakayama, Sears, Hino and Lupker (2012) also found similar results on a masked phonological priming paradigm with cross-script Japanese-English bilinguals where the participants made lexical decisions to L2 English targets. Significant priming effects were observed for both cognate translation primes and phonologically similar primes. Likewise, Khan and Buchanan (2013) found cognate effects for words when Urdu-English bilinguals were presented with English words in a simple lexical decision task, indicating a non-selective lexical access and interconnectivity of the bilingual lexicon.

Research on Selective View

Although most findings are consistent with the non-selective view of language processing in bilinguals, there are some studies that contradict these accounts (Schwartz and Kroll, 2006; Titone, Libben, Mercier, Whitford, and Pivneva, 2011, Hoversten, Brothers, Swaab, and Traxler, 2015). For instance, Schwartz and Kroll (2006) examined L2 word recognition performance of Spanish-English bilinguals with high and intermediate levels of proficiency. They used cognates and interlingual homographs in English high- and low-constraint sentences to determine whether sentence context would modulate cross language, non-selective activation. The authors observed a cognate facilitation effect in low constraint sentences, but not in high-constraint sentences, suggesting the restriction of non-selective access when there is sufficient semantic information available to suppress the non-target language.

Another study by Titone and colleagues (2011) investigated cross-language activation in an L1 (English) reading task as a function of L2 (French) age of acquisition and task demands. They asked participants to read English sentences that consisted of interlingual homographs, cognates, and control words. Their findings suggested that individual differences in L2 acquisition modulated the degree of non-selective access for bilinguals reading in their L1. Similary, Hoversten et al. (2015) found evidence for partially selective access in Spanish-English bilinguals. The subjects performed a language go task, where a go or no-go response was required based on the word's language membership and animacy (living or non-living). In addition, they used EEG to assess the temporal relationship between access to language membership information and animacy information. Their results suggested that the bilingual brain is able to quickly identify the language to which a word belongs and then use this information to selectively modulate the depth of processing in each language accordingly.

Most studies in the past have found similar support for the BIA+ model, providing evidence for a parallel, non-selective operation of lexical activation in bilingual memory, even when the contextual cues call for only one language (Gollan et al., 1997; Kim and Davis, 2003; Mishra and Singh; 2014; Khan and Buchanan, 2013). But, some studies have also demonstrated the possibility of selective or partially selective access under certain conditions (Schwartz & Kroll, 2006; Hoversten et al., 2015; Titone et al., 2011). Thus, the field has not yet reached a consensus about the pervasiveness of nonselective access, nor the specific determinants that may lead to selective access.

The Word Frequency Effect

A potential way to advance towards a more complete formalization of the bilingual lexical access theory is to find a word level variable that is known to exert a strong influence on a bilingual's word recognition ability. There are many such variables including orthographic neighborhood, concreteness, word length, and word frequency (Andrews, 1997; Murray & Forster, 2004). Of the word level variables known to influence performance on visual word recognition, word frequency is arguably the most powerful and robust (Murray & Forster, 2004). Word frequency is a determinant of the speed with which lexical information can be retrieved. As one might expect, words that appear more commonly in print are recognized faster and more efficiently than words that are less frequent (Preston, 1935; Brysbaert, Stevens, Mandera, & Keuleers, 2016). This is known as the word frequency effect.

Previous studies show that word frequency especially affects lexical accessibility in bilingual language production, and disadvantages related to bilingualism are most apparent during the retrieval of low frequency words (Gollan, Montoya, Cera, & Sandoval, 2008; Ivanova & Costa, 2008; Mendez & Gollan, 2010). For instance, Gollan and colleagues (2008) demonstrated a larger frequency effect in English-Spanish bilinguals than English monolinguals when they were asked to name pictures of L1 low-frequency words (e.g. *pharmacy*) as compared to L1 high-frequency words (e.g. *church*). These findings are suggestive of a bilingual disadvantage in picture naming because of frequency.

In another example, Miwa, Djikstra, Bolger, and Baayen (2013) conducted a lexical decision eye-tracking study that examined the influence of word frequency, phonology, and semantics of Japanese (L1) words on English (L2) word recognition processes. They observed bilingual-specific effects in the reaction times, including the interaction between L1 and L2 frequencies which was found early on at the first subgaze. Providing that it takes approximately 200 ms for response planning and execution (Schmidt, 1982), the cross-language competition was presumably a part of the central lexical processing mechanism as explained through the BIA+ model, according to which a cross-script word's lexical orthographic representation occurs via top-down processing from the conceptual representation (Miwa et. al., 2013; Djikstra & Van Heuven, 2002). The BIA+ also posits that differences in word frequency can be seen in the resting-level activation of the word, so high frequency words are activated more rapidly and reach the recognition threshold earlier because they have a higher resting level activation than low frequency words (Djikstra & Van Heuven, 2002).

In the case of bilingualism research, however, most studies use participants who acquired their second language later in life through school, and thus are not considered balanced-bilinguals. Consequently, a larger L2 word frequency effect is observed for these participants as compared to their L1, especially for high frequency L2 words (Diependaele, Lemhofer, & Brysbaert, 2013; Gollan et al., 2008; Van Wijnendale & Brysvaert, 2002). In one large-scale study, Lemhofer, Dijkstra, Schriefer, Baayen, Grainger, & Zwitserlood (2008) compared English word recognition in Dutch-English bilinguals, French-English bilinguals, German-English bilinguals, and English monolinguals in a progressive demasking paradigm which is a perceptual identification task where a word that slowly emerges from a pattern must be identified as quickly as possible and subsequently be typed in after identification. They observed many similarities between the four groups of participants in reaction time patterns, suggesting a generalization of the L2 word recognition processes. In addition, Lemhofer and colleagues (2008) found cognate status and frequency effect of words to be the only robust differences between the English monolinguals and bilinguals. With respect to word frequency, bilinguals (English = L2) required more time to process low-frequency words as compared to monolinguals (English = L1).

Diependaele, Lemhofer, and Brysbaert (2013) found similar results showing a larger frequency effect for second-language processing than for native-language processing. However, they argue that this frequency effect is irrespective of whether one is bilingual, more proficient in one language than the other, or if there is similarity between the L1 and L2 languages. They hypothesize these effects to be a result of within-language characteristics, like vocabulary size. Due to a generally low proficiency of L2 in unbalanced bilinguals, lexical processing may require more time because of weaker representations; Diependaele et al. (2013) called this the “lexical entrenchment account”. Once they controlled for English vocabulary size, all differences between bilinguals and native speakers disappeared. In a more recent study, Brysbaert, Lagrou, and Stevens (2017) tested this effect of the lexical entrenchment hypothesis using a lexical decision task, a more commonly used tool in word recognition because it shows a very clear word frequency effect (Keuleers, Lacey, Rastle, & Brysbaert, 2012). In line with previous research, once the participant’s vocabulary size was taken into account, the differences between L2 and L1 word frequency effects dramatically decreased.

Objectives

The present paper aims to determine the lexical processing mechanism in bilinguals using word frequency as the independent variable. For example, if the English word *‘cat’* were high frequency and the French word *‘chat’* were low frequency, it would be interesting to see whether the French word *‘chat’* would be recognized faster in French-English bilinguals compared to another word with the same low frequency. If such facilitation occurs, it would be inferred that the advantage came from the higher frequency English translation. Although the question of interest is fairly straight forward, the implementation of this experiment is far more complicated. For most language pairs, there exist very few words that differ in their word frequency – *'cat'* is about as frequent in English as '*chat'* is in French (Davis & Perea, 2004; New, Pallier, Brysbaert, & Ferrand, 2005). Because of this complication, there is a paucity of research in this area.

Fortunately, however, Khan and Buchanan (2006) have produced a word frequency database for written Urdu, a Persian-Arabic language script for which the frequencies of some words are observed to be very different from their English counterpart. The large number of high-low and low-high frequency pairs of English and Urdu provide a unique opportunity for investigating how word frequency in one language influences the processing of the word presented in the other language. In the research reported here, I capitalized on this fairly unique decoupling of frequencies to test the extent to which word frequency influences the cross-lingual interaction in bilinguals. This was done using a standard lexical decision task where the participant identified a string of letters as a word or a non-word as quickly and accurately as possible (Meyer & Schvaneveldt, 1971). Target words belonged to only one language; as in, there were no cross-language homographs, homophones, or cognates within a stimulus set. The delay in response time (RT) was then used to interpret the ease of word recognition.

It is hypothesized that for bilinguals (but not monolinguals).reaction times for L2 words will differ as a function of the frequency of L1 translation. For example, a low frequency L2 word with a high frequency L1 translation should be recognized faster than low frequency L2 words with low frequency L1 translations. This is only expected for the bilinguals of course since they would not have exposure to the L1 translations. Such an effect would be consistent with the non-selective access model. In addition, I expected to see a larger L2 word frequency effect in bilinguals correlated with lower levels of English proficiency when compared to monolinguals.

CHAPTER II

DESIGN AND METHODOLOGY

Participants

Forty-five English monolinguals were recruited from the University of Windsor, Canada, through a participant pool (N=45); they received a partial credit of 1.00 in exchange. The University of Windsor student population consisted of a very small number of Urdu-English bilingual participants. For this reason, ninety Urdu-English bilinguals were recruited from the University of Agriculture Faisalabad, Pakistan. The monolingual participants consisted of 8 males (0.18%) and 37 females (0.82%) ranging between ages 18 to 36 (*M* = 22.8 years). The bilingual participants consisted of 64 males (0.74%) and 26 females (0.26%) ranging between ages 20 to 31 (*M* = 23.2 years). All participants had normal or corrected-to-normal vision and knew the experiment involved English word recognition. The bilingual participants were native speakers of Urdu who learned English early in life with a mean age of acquisition of 10.6 years. The bilingual participants were also fluent in speaking Punjabi (the provincial language in that area of Pakistan); but this knowledge is not expected to affect the results in a way that invalidates the conclusions.

The above-mentioned criteria have been implemented in previous studies investigating similar ideas (Khan & Buchanan, 2013; Mishra & Singh, 2014; Kim& Davis, 2003). The experiments were conducted in accordance with the ethical guidelines of the Tri-Council policy statement and the Research Ethics Board.

Method of evaluating language proficiency in bilinguals

A variety of methods are used to evaluate the level of language proficiency and dominance in individuals, including self-report measures, or certain experimental or standardized tasks. However, subjective measures like self-reports are not always accurate or practical. They may consist of biases (Matsuno, 2009) as well as inaccurate responses due to cultural or linguistic differences (Beaton et al, 2000).

A more thorough perspective of an individual's language profile can be measured through detailed language questionnaires that ask participants to report the current amount of use and the context of use for each language (Birdsong, Gertken, Amengual, 2015; Marian, Blumenfeld, Kasuhanskaya, 2007). Such measures have been demonstrated to have good reliability and validity (Marian, Blumenfeld, Kasuhanskaya, 2007).

For the current study, a Language Experience and Proficiency Questionnaire (LEAP-Q, see Appendix B) that consists of questions about the participant's use of Urdu and English was used. This tool was developed by Marian, Blumenfeld, and Kaushanskaya (2007). In this questionnaire, participants rate their proficiency levels based on language competence (including proficiency, dominance, and preference ratings); age of acquisition; modes of acquisition; prior language exposure; and preference ratings (Marian et. al, 2007). All bilingual participants filled out this questionnaire.

Material and Apparatus

The experimental stimuli were selected such that four frequency categories were formed using high and low frequency words in Urdu and English. Urdu frequency counts were derived from a database constructed previously by Khan and Buchanan (2006). English frequency counts were based on the Wordmine database developed by Durda and Buchanan (2006).

The four frequency categories consisted of: 1) low-frequency English words with their high frequency Urdu pairs (LFE-HFU); for example, English word *nanny* (LFE, 3.3 ppm) and its Urdu pair آیا (HFU, 617.0 ppm); 2) low-frequency English words with low-frequency Urdu pairs (LFE-LFU), for example, the English word *yam* and its Urdu pair شکرقند (LFU, 1.0 ppm); 3) high-frequency English words with low-frequency Urdu pairs (HFE-LFU), for example, the English word *history* (HFE, 177.9 ppm) and its Urdu pair سرگذشت (LFU, 6.0 ppm) 4) high-frequency English with high-frequency Urdu pairs (HFE-HFU), for example, the English word *fish* (HFE, 106.2 ppm) and its Urdu pair مچھلی (HFU, 156.0 ppm). There were 60 items in each of these categories (see Appendix A).

Words with an orthographic frequency of 10 or less occurrences per million were defined as low-frequency words while words with an orthographic frequency of 95 occurrences per million or more were defined as high-frequency words. Orthographic frequency is "the frequency with which a word is encountered in written text, expressed as the number of encounters per million words" (Westbury, Buchanan, Sanderson, Rhemtulla, & Phillips, 2003, p. 205)

Word length was determined by counting the number of letters in each word. Urdu is more complex orthographically than English. It comprises of diacritics that represent vowels, such that many written words have two or more readings. For instance, the Urdu spelling alif-laam-seen (سلا) may be read /sila¯/ (cause to be sewn) or /sula¯/ (put to sleep; Rao, Vaid, Srinivasan, & Chen; 2011). The letters also have graphemic complexity where one letter may have multiple graphemic forms depending on their position within a word (Rao et. al, 2011). For this study, the length of Urdu words was determined by distinguishing each alphabet from the word and counting them. The diacritics were also counted as an alphabet. For example, the Urdu word "آلو" was separated into its three alphabets ل, ا, and و. Then the diacritic was counted as a single letter, with the resultant word length of 4.

All words were nouns and the researcher evaluated concreteness by selecting words that were most concrete and excluding obscure and archaic words. English non-words were created using words that matched the stimulus words in length and bigram frequency. Two letters were changed in each of these words to create orthographically and phonologically legal non-words. For both participants, the English target words and non-words were presented in a random order in a single block.

Participants were tested using a Dell PC. All stimuli were presented in the center of the screen in black against a white background. A 42-point Times New Roman font was used to present the words. The participants' responses were recorded using the Direct RT software (Jarvis, 1999).

Procedure

The study was approved by the University of Windsor Research Ethics Board (REB# 17-144). Participants were tested over one single session that lasted approximately 20 minutes for monolinguals and 40 minutes for bilinguals. The English monolinguals and Urdu-English bilinguals were presented with stimuli from all four conditions: 30 HFE words with LFU pairs, 30 LFE words with HFU pairs, 30 HFE words with HFU pairs, and 30 LFE words with LFE pairs. After signing the consent form, instructions were given at the beginning of each trial in English. Participants were asked to identify the string of letters as a word or a non-word as quickly and accurately as possible. The items appeared in random order in the center of the computer screen.

All participants completed the study in a quiet room with adequate lighting. The task was performed on a Dell laptop and the participants faced a wall to minimize distractions. Each participant pressed the “?” key if the string of letters was a word and the “Z” key if it was a non-word, with their index fingers placed on these keys throughout the task. The items appeared in random order in the center of the computer screen. After each response, the stimulus was cleared from the screen and the next stimulus appeared. For both groups of participants, the practice trial consisted of 10 practice stimulus items, followed by the experimental trial with 240 stimuli. The RTs were measured from the onset of the stimulus to when the subject pressed the response button. After the lexical decision task, all bilingual participants also filled out the LEAP Questionnaire on an Acer PC.

CHAPTER III

STATISTICAL ANALYSIS AND RESULTS

LEAP Questionnaire Results

The LEAP Questionnaire results indicated that all bilingual participants (N=90) identified Urdu as their primary language and English as their secondary language. The mean age of acquisition was 4.2 years for Urdu and 10.6 years for English. The participants reported their current exposure to be on an average of 46.6% to Urdu, 22.5% to English, and the remaining 30.6% to Punjabi (the provincial language). In addition, they rated their average proficiency score as 9 (excellent) for Urdu and 7 (good) for English on a 10-point Likert-type rating scale (see Appendix B).

Statistical Analysis Results

Multiple analyses were conducted to determine the relationship of lexical entries between the primary and secondary language. The monolingual participants (*N*=45) had no previous exposure to Urdu and thus were used as the control group. The bilingual participants (*N*=90) were randomly divided into two groups of forty-five (group-A and group-B) and were contrasted with the monolingual group. In the first analysis, monolingual participants and bilingual group-A were examined together and fitted in a linear mixed effects regression (lmer) model. An lmer analysis was conducted because of the linear relationship that exists between the word frequency and reaction times and to resolve the non-independencies in the data. The results of this analysis were then replicated using bilingual group-B. The purpose of this second analysis was to ensure reliability of the results. A third analysis was done to observe the effect of bilingual proficiency level on the word recognition performance in all four frequency conditions. In this analysis, all bilingual participants (*N*=90) were divided into two groups based on their proficiency ratings (high or low) and fitted in a linear mixed effects regression model. All statistical analyses were performed in R (Version 3.5.0) using the lme4 package. Lmer analyses were conducted because of the linear relationship that exists between the word frequency and reaction times and to resolve the non-independencies in the data. For all three lmer analyses, the assumption of multivariate normality was met with the residuals being approximately normally distributed. The assumption of homoscedasticity was also met as determined by the constant spread of the residual against the fitted values.

Only responses to target words were included in the statistical analyses. First, participants and stimulus items with less than 70 % accuracy rates were excluded from the study. This resulted in 13 out of 120 stimulus items to be removed (7 from LFEHFU condition and 6 from LFELFU condition). No participants responded below the accuracy cut-off, so they were all included in the analysis. Next, all 363 incorrect responses (3.8 %) were removed from the data. Of these erroneous clicks, 140 (1.5%) were from the English monolingual group and 223 (2.3 %) were from bilingual group-A. Reaction times less than 200ms and more than 5000ms were regarded as invalid responses to avoid inflating individual means. The maximum cut-off was chosen to 5000ms because English is a secondary language for bilinguals, thus slower reaction times are expected from this population. This resulted in 14 responses to be removed from the analysis. In total, 9.8% of the responses were removed from the original data set. Participant mean response times, standard deviations, and error rates per condition for the final data set are displayed in Table 1 for bilingual participants and Table 2 for monolingual participants.

Table 1

Mean Reaction Times (RTs) with Standard Deviations, Number of Responses, and Standard Error Rates for Bilingual Participants

Condition	Mean RT (msec)	SD	N	SE(%)
HFEHFU	785	414	1328	11.4
HFELFU	876	490	1314	13.5
LFEHFU	1073	630	955	20.4
LFELFU	1085	631	984	20.1

Table 2

Mean Reaction Times (RTs) with Standard Deviations, Number of Responses, and Standard

Condition	Mean RT (msec)	SD	N	SE(%)
HFEHFU	594	244	1319	6.72
HFELFU	598	179	1325	4.93
LFEHFU	681	231	987	7.34
LFELFU	658	305	904	9.45

Error Rates for Monolingual Participants

First, the difference between reaction times for high and low frequency words were compared in the monolingual group and bilingual group-A with paired *t*-tests. As expected, these results showed a frequency effect on word recognition performance. Monolinguals recognized high frequency words more quickly than low frequency words [$t(4124) = -14.05$, $p < .001$]. Similarly, in bilinguals, the reaction times were faster for high frequency words than low frequency words [$t(3784) = -19.05$, $p < .001$, see Figure 1]. Paired *t*-tests were also used to compare the differences between the reaction times of monolingual and bilingual participants. Results showed an L2 disadvantage, with bilinguals responding significantly slower to the English target words than the monolingual group [$t(6278_{)} = -34.69$, $p < .001$, see Figure 2].

Next, mean reaction times (RT) for each condition were examined to determine the influence of L1 word frequency on the recognition of the L2 word pairs. Forty-five of the ninety bilingual participants were randomly selected for this analysis to match the sample size of the forty-five monolingual participants. Data from monolingual and bilingual participants were analyzed together and fitted in a linear mixed effects regression model (lmerTest package in R) to the logit transformed proportion of RTs. As fixed-effects predictors, the word frequency Condition (HFEHFU, HFELFU, LFEHFU, LFELFU) and Group (monolingual or bilingual) were entered into the model. The model was fitted with a backwards step-wise elimination procedure where the predictor variables that did not significantly improve the model as indicated by likelihood ratio testing were removed one by one. The inclusion of the random slope for Subject by Condition was justified by the likelihood tests. In this fitted model, outliers with a standardized residual at a distance greater than 3.0 standard deviations from the mean were excluded. This resulted in the removal of 178 responses (1.9% of the data) which were associated with large residual values. Of these, 40 responses were from the monolingual group and 138 responses were from the bilingual group-

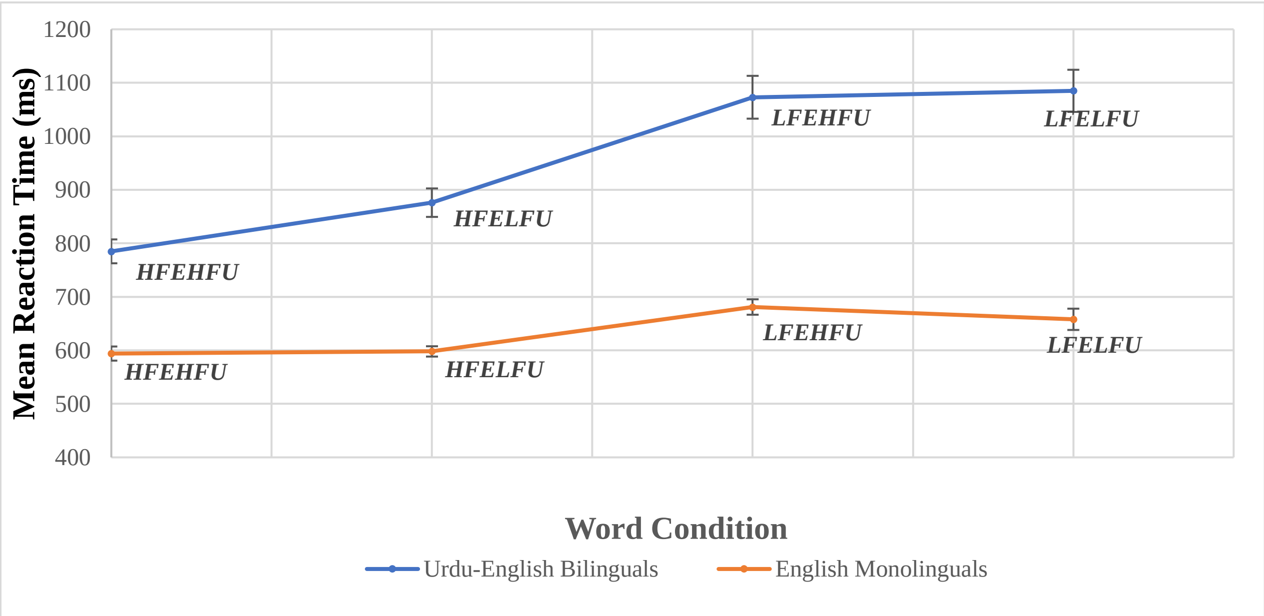

Figure 1. The Effect of Frequency in Monolinguals and Bilinguals

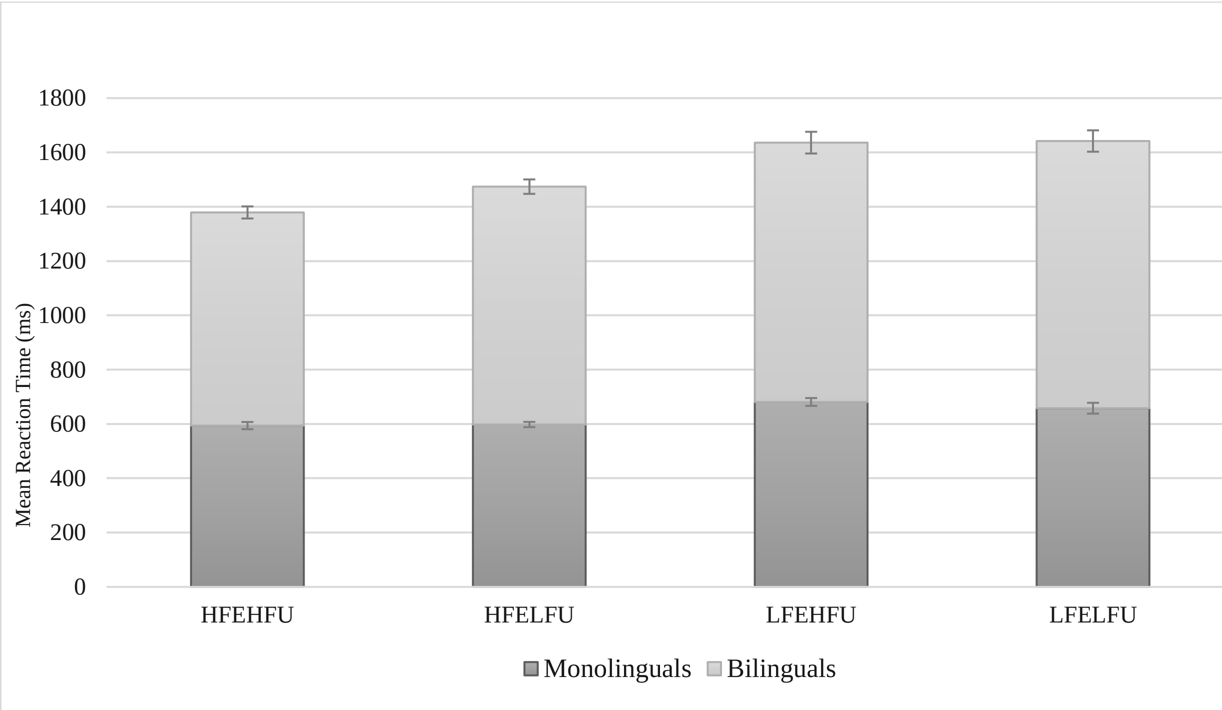

Figure 2. Comparison of Mean RTs in Lexical Decision Task between Monolingual and Bilingual Participants for Each Condition.

P-values were obtained for the fixed effects using the lmerTest package with Satterthwaite approximations to degrees of freedom. Unsurprisingly, the model with the best fit showed a main effect for Group [estimate = -0.22, $t(88.07) = -7.89$, $p < .001$] with bilinguals producing longer RTs than the monolinguals (see Figure 1 and 2).

Post-hoc group comparisons with Tukey's adjustments were then performed using the emmeans package. In monolingual participants, results showed no difference between the RTs in the high frequency conditions [HFEHFU and HFELFU; $t(120) = -0.794$, $p = 0.993$]. There was also no difference observed between low frequency condition responses [LFEHFU and LFELFU; $t(123) = 1.906$, $p = 0.546$]. Although the difference between the low frequency condition was not significant, the monolingual participants took longer to respond to the LFEHFU condition ($M = 681$, $SD = 231$) than the LFELFU condition ($M = 658$, $SD = 305$). This is important to note when analyzing results from the bilingual group data as this difference may result in obscurity of the frequency effect in this population.

In bilingual participants, post-hoc multiple comparisons showed a difference between the RTs for the two high frequency conditions [$t(121) = 3.608$, $p < 0.01$)]. The lower frequency of the Urdu translation pair in the HFELFU condition resulted in longer response times compared to words in the HFEHFU condition. However, this difference was not observed between RTs for the two low frequency conditions (LFEHFU and LFELFU) in the bilingual group [$t(126) = 0.467$, $p = 0.999$]. Results of monolingual and bilingual group-A mean logged responses are displayed in Table 3 below.

Table 3
Difference in Mean Log Reaction Times Between Each Condition per Group with Standard Error, z-values, and p-values (MG=monolingual group, BA=bilingual group A).

Condition	Mean logRT Difference (msec)	SE	z-value	p-value
MG: HFEHFU – HFELFU	-0.01	0.02	-0.79	0.994
MG: LFEHFU – LFELFU	-0.01	0.03	1.91	0.546
BA: HFEHFU – HFELFU	-0.09	0.02	-3.61	0.007
BA: LFEHFU – LFELFU	-0.01	0.03	-0.47	0.999

A second analysis was then conducted using bilingual group-B (N=45). Similar to the first analysis, stimulus items and participants with less than 70% accuracy were removed. This resulted in the same 13 out of 120 of stimulus items as above to be excluded. In the bilingual group-B, 2 participants responded below the accuracy cut-off and thus were also excluded from the analysis. All 364 incorrect responses (3.8%) were then removed from the dataset. Next, 11 responses (3: monolingual; 8: bilingual group-B) with reactions times considered as outliers were removed using the same criteria as above. Overall, 9.8% of the responses were removed from the original data set.

Like the procedures used above, the monolingual control group and bilingual group-B were fitted in an LMER model to reaction times. Subjects and Items were used as a crossed-random factor, and Condition and Group were used as fixed-effect predictors. The model fitting procedure was the same as above. A random slope for Subject by Condition was included as justified by likelihood tests. Next, residual outliers at a distance greater than 3.0 standard deviations from the mean were excluded. This resulted in a total of 150 (1.7 %) responses to be removed (30: monolingual; 83: bilingual group-B).

As in the first analysis, the model with the best fit showed a main effect for Group [estimate = 0.14, t(86.6) = 5.06, p < .001] with monolinguals responding to the target words at a faster rate than the bilinguals in all four conditions. Post-hoc multiple comparisons also showed results similar to the first analysis. Monolingual participants did not differ in response times between the two high frequency conditions [HFEHFU and HFELFU; $t(118) = -0.83$, $p = 0.992$] and between the two low frequency conditions [LFEHFU and LFELFU; $t(116) = 1.72$, $p = 0.675$].

In contrast, bilingual participants showed a difference in reaction times between the two high frequency conditions. Again, they responded faster for the HFEHFU condition than the HFELFU condition [$t(119) = -3.89$, $p = 0.003$]; whereas, this difference was not observed in the LFEHFU and LFELFU conditions [$t(119) = -0.84$, $p = 0.991$]. These results replicated the findings from the initial analysis, increasing the accuracy of the current experiment’s findings. See Table 4 for the differences in mean logged reaction times between conditions for the monolingual group and bilingual group-B.

Table 4
Difference in Mean Log Reaction Times Between Each Condition per Group with Standard Error, z-values, and p-values (MG=monolingual group, BA=bilingual group B).

Condition	Mean logRT Difference (msec)	SE	z-value	p-value
MG: HFEHFU – HFELFU	-0.02	0.03	-0.83	0.992
MG: LFEHFU – LFELFU	-0.05	0.03	1.72	0.675
BB: HFEHFU – HFELFU	-0.09	0.02	-3.89	0.003
BB: LFEHFU – LFELFU	-0.02	0.03	-0.84	0.991

Results from the above analyses show that, in bilinguals, the frequency of the Urdu (L1) word pair influenced word recognition ability for the target English (L2) word in high frequency conditions, but this effect was masked in the low frequency conditions. Previous studies have demonstrated that a difference in L2 proficiency can modulate the size of the frequency effect (Lemhofer et al., 2008; Diependaele et al., 2013). Accordingly, to assess whether level of English proficiency in bilinguals influenced the response times for each condition, a third analysis was conducted.

The level of proficiency for each participant was determined by taking their LEAP-Q ratings on Proficiency in Speaking English, Proficiency in Understanding Spoken English, and Proficiency in Reading English. The ratings were provided on a 10-point Likert-type scale from 0 being *none* and 10 being *perfect* (see Appendix B). These three scores were summed together out of a total score of 30. Participants were then divided into two groups based on their total score – high proficiency (HP; *N*=44) and low proficiency (LP, *N*=44).

The data cleaning and statistical procedures were identical to the above analyses. The same 13 out of 120 stimulus items that did not meet the accuracy cut-off were removed and 2 bilingual participants with less than 70% accuracy in their responses were excluded for this analysis (N=88). In addition, 447 (4.7%) incorrect responses were removed from the dataset.

From these, 199 were from the HP group and 248 were from the LP group. There were also 19 (0.2%) responses that were considered outliers and thus excluded from the analysis. This resulted in a total of 13.6% of the responses to be removed from the original data set. Participant mean response times, standard deviations, and error rates per condition for the final data set are displayed in Table 5 for high proficiency bilinguals and Table 6 for low proficiency bilinguals.

Table 5

Mean Reaction Times (RTs) with Standard Deviations, Number of Responses, and Standard Error Rates for High Proficiency Bilingual Participants

Condition	Mean RT (msec)	SD	N	SE(%)
HFEHFU	713	325	1296	9.04
HFELFU	790	382	1291	10.6
LFEHFU	1012	551	943	17.9
LFELFU	977	469	969	15.1

Table 6

Mean Reaction Times (RTs) with Standard Deviations, Number of Responses, and Standard Error Rates for Low Proficiency Bilingual Participants

Condition	Mean RT (msec)	SD	N	SE(%)
HFEHFU	784	404	1304	11.2
HFELFU	868	482	1271	13.5
LFEHFU	1066	619	920	20.4
LFELFU	1072	618	956	20.0

First, paired samples *t*-tests were conducted to analyze differences in mean reaction times between the HP and LP groups. Results showed that the bilinguals who were higher in proficiency recognized the target words more quickly than bilinguals who were lower in proficiency [$t(8851) = -7.26$, $p < .001$].

Next, reaction times for both groups were analyzed with LMER-models. Condition and Proficiency were used as fixed-effect predictors, and Subject and Items were used as a crossed-random factor. The model fitting procedure was the same as above. The likelihood ratio test showed that random slopes for Condition by Subject and Proficiency by Subject improved the model fit. Outliers with a standardized residual greater than 3.0 standard deviations from the mean were then excluded, resulting in a total of 168 responses to be removed (78 from HP and 68 from LP). The model with the best fit showed no effect for proficiency [estimate = 0.07, $t(84.6) = 2.04$, $p = 2.04$], indicating that both high and low proficiency groups performed similarly on each condition.

Because the level of proficiency did not modulate the effect of frequency on word recognition performance, a Pearson product-moment correlation coefficient was computed to investigate the association between the mean RTs and proficiency ratings for the bilingual participants. Results showed no correlation between the two variables [$r(86) = -0.05$, $p = 0.6$].

CHAPTER IV

DISCUSSION

General Discussion

The present study tested the effect of differential word frequency on the relationship of lexical entries between the primary and secondary language. Although the frequency effect is ubiquitous in visual word recognition, no study has explicitly focused on how the frequency of a word in a primary language (L1) might influence recognition of that word in the secondary language (L2). Therefore, the main objective of this study was to investigate the frequency effect of L2 on the word recognition of L1 in bilinguals.

The results provided support for the word frequency effect in the primary language and the secondary language (Preston, 1935; Brysbaert et al., 2016). Both monolinguals (L1) and bilinguals (L2) produced an English frequency effect – they responded faster to high frequency target words than low frequency target words.

In addition, as hypothesized, the bilingual group produced longer RTs than the monolingual group across all four conditions. This is in accordance with previous studies which have found an effect of language experience on word recognition response times due to a lower level of proficiency and/or smaller vocabulary size of the secondary language (Lemhofer et al., 2008; Diependaele et al., 2013; Brysbaert et al., 2017). The BIA+ model also touches upon this topic by indicating that the time of activation for a target word will be similar if the word frequency, language proficiency, and recency of use are equal in both languages (Dijkstra & Van Heuven, 2002). Although the word frequency of the target words was identical for the monolingual and bilingual groups in the present study, the bilinguals were less proficient in English (L2) than the monolingual group (L1), thus taking longer to respond to the target words.

As expected, the reaction times for the control group (monolinguals) did not differ between the two high frequency word conditions (HFEHFU and HFELFU) and between the two low frequency word conditions (LFEHFU and LFELFU). However, it is worth noting that although not significant, the monolingual participants took longer to respond to the LFEHFU condition (M = 681, SD= 231) than the LFELFU condition (M=658, SD = 305). This difference is the opposite of the expected effect as LFELFU words were predicted to take equal, if not longer, time than LFEHFU to recognize a target word. Since the monolingual participants had no exposure to Urdu, the higher mean reaction time for the LFEHFU condition may be attributed to the English target words not balanced uniformly between the two low frequency conditions.

Even though the stimulus items were matched for word frequency, there are a few other variables that may have influenced word recognition performance, like phonological irregularity, which is the inconsistency between the spelling to sound correspondence of a word. Previous studies have shown a latency effect for recognizing words with unusual spelling to sound contrast (e.g. *lever, chaos*) as compared to regular words (e.g. *smug, hack*; Stanovich, 2009). Another such variable is orthographic neighborhood structure, where orthographic neighbors refers to the number of words obtained by changing a single letter in the target word (e.g. *cat's* neighbors include *cut*, *bat*, and *cap*). Research shows that words with many neighbors are recognized faster than those with a few neighbors (Yap, Balota, Sibley, & Ratcliff, 2012). Emotional valence might also have impacted word recognition performance, with negative words eliciting slower responses than neutral or positive words (Kuperman, Estes, Brysbaert, &

Warriner, 2014). It is possible that words in the LFEHFU condition may have had more phonological irregularity, fewer orthographic neighbors, or a greater number of negative words than those in the LFELFU condition.

In comparison, bilinguals took longer to recognize high frequency English words with low frequency Urdu pairs (HFELFU) than high frequency English words with high frequency Urdu pairs (HFEHFU). Because the monolingual control group did not show a difference between these two conditions, it can be inferred that this advantage came from the higher frequency Urdu translation. Consistent with our hypotheses, this result provides support for the non-selective view of language processing in bilinguals like the BIA+ model which posits that sensory inputs from both languages are activated simultaneously in the brain during word recognition (Dijkstra & Van Heuven, 2002).

Although the bilinguals responded slightly faster for the LFEHFU condition than the LFELFU condition, a significant frequency effect was not observed in the low frequency conditions. It is likely that this frequency effect was obscured by the unequal selection of the English stimulus items as demonstrated by results from the control condition where monolingual participants took slightly more time to recognize words in the LFEHFU condition as compared to the LFELFU condition.

Alternatively, it is possible that the bilingual participants experienced a ‘floor effect’ for the low frequency conditions considering the lower level of English proficiency in this group as compared to monolinguals. Previous literature postulates that disadvantages related to bilingualism are most apparent during the retrieval of low frequency words (Gollan et al., 2008; Miwa et al., 2013). Further, according to the BIA+ model, differences in word frequency can be seen in the resting-level activation of the word, so high frequency words are activated more rapidly and reach the recognition threshold earlier because they have a higher resting level activation than low frequency words (Djikstra & Van Heuven, 2002). As such, being an unbalanced bilingual may result in having more English target words at lower levels of resting activation relative to monolinguals, ultimately reaching a plateau for words that have the lowest frequencies. This may render the bilingual group relatively insensitive to differences between low frequency English words as they could not have physically responded any slower regardless of the frequency of the associated Urdu word pair.

Results also showed that although bilinguals who rated themselves as highly proficient in English had overall faster response times, the level of proficiency did not modulate the effect of frequency between the four conditions. All bilingual participants responded similarly between the high frequency conditions (HFEHFU and HFELFU) and the low frequency conditions (LFEHFE and LFELFU) regardless of proficiency level. More importantly, however, correlation analysis results showed no relationship between the level of proficiency and mean reaction times in bilinguals. Hence, these findings must be interpreted with caution as the proficiency rating amongst bilinguals were subjective and do not seem to be a true measure of their level of English proficiency.

Limitations and Future Directions

One limitation to this study is that the bilingual participants were unbalanced in their proficiency between the two languages – they were more proficient in their primary language (Urdu, L1) than their secondary language (English, L2). Therefore, an overall latency in word recognition performance was observed in this group and they may have been at a lexical processing disadvantage, especially for low frequency words. In the future, it would be interesting to examine bilinguals with the same level of Urdu and English proficiency to determine the extent of the frequency effect without the influence of this confounding variable.

In addition, the proficiency measure used in the current study was a self-report questionnaire (LEAP-Q) where participants provided subjective ratings on their level of proficiency. To get a more objective measure of proficiency, standardized language tests should be considered. For example, the Peabody Picture Vocabulary Test (PPVT) is a widely-known measure of receptive vocabulary for English and provides a more accurate measure of an individual's language proficiency (Dunn & Dunn, 2007). Using such a measure would allow researchers to analyze the frequency effect as a function of proficiency level.

The present study also observed some inconsistency in the reaction times between the low frequency conditions for the monolingual control group. As previously mentioned, although the stimulus set words were matched for frequency, there may have been other word recognition variables that resulted in an unbalanced stimulus set in the low frequency conditions. Next steps should consider controlling for these variables to prevent obscurity of the results.

In addition, this study only presented target words from the bilinguals' L2 and the influence of L1 word frequency was observed on word recognition performance. Future research should continue to explore the impact of word frequency between the primary and secondary language but also show target words from the bilinguals' L1. It would be interesting to assess if the L2 translation pair results in a smaller frequency effect since the secondary language is acquired later in life.

Lastly, in the present paper, there were no neuroimaging or electrophysiological tools incorporated to assess the structural and functional mechanisms that underly bilingual language processing. Future studies should consider using such measures to provide physiological evidence for the secondary-language word frequency effect in bilingual readers and further advance on the development of the non-selective access theory. For example, using an electroencephalogram, one can compare the N400 and P300 component of ERP waves between monolinguals and bilinguals. A difference in amplitude of these ERP waves may be suggestive of the distinct mechanisms for lexical word processing between the two groups.

Summary

The current study provides the first evidence for the relation of lexical entries between the primary and secondary language as a function of word frequency. A bilingual's word recognition ability in one language was influenced by the frequency of its translation pair in another language. These results suggest that recognition of words in different languages is sub-served by one system in bilinguals and have implications for advancing more generalized models of non-selective bilingual language processing, such as the BIA+ model. In addition, this research raises issues not present in monolinguals by highlighting the question of whether recognition of words in one language can be studied separately from the other language in bilinguals.

REFERENCES

Adelman, J. S. (Ed.). (2012). *Visual Word Recognition volume 2: Meaning and context, individuals and development*. Psychology Press.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of psycholinguistics, 2*, 285-375.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25(24),* 3186- 3191.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. COERLL, University of Texas at Austin.

Brenders, P., Van Hell, J. G., & Dijkstra, A. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109*, 383–396.

Brysbaert M, Duyck W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after 15 years of service*? Bilingualism: Language and Cognition, 13(3),* 359-371. doi:10.1017/S1366728909990344

Brysbaert, Stevens, Mandera, & Keuleers. (2016). The impact of word prevalence on lexical decision times: Evidence from Dutch lexicon project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42(3),* 441-458.

Costa, A., Caramazza, A., & Sebastian-Gall, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1283–1296.

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

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistics indices in Spanish. *Behavior Research Methods, 37*(4), 665-671. doi:10.3758/BF03192738

De Groot, A. M. B., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47,* 91–124.

De Groot AMB, Poot R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47,* 215–264.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *The Quarterly journal of experimental Psychology*, *66*(5), 843-863.

Dijkstra, T. & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5,* 175– 197

Dijkstra, T. (2003). *Lexical processing in bilinguals and multilinguals: The word selection problem.* In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), The Multilingual Lexicon (p 11-26). Netherlands: Kluwer Academic Publishers.

Dijkstra, T. (2005). *Bilingual visual word recognition and lexical access*. In J. F. Kroll & A. M. B. De Groot (Eds.), Handbook of Bilingualism: Psycholinguistic Approaches (p. 531 – 553). New York, NY: Oxford University Press, Inc.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language, 41*, 496–518.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*, 445-464. doi: 10.1006/jmla.1999.2697

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Gerard, L. D. & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15 (2),* 305-315. doi: 10.1037/0278-7393.15.2.305

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science, 22,* 1155–1164

Gollan, T. H., Montoya, R. I., Cera, C. M., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). The bilingual effect on Boston naming test performance. *Journal of the International Neuropsychological Society, 13*, 197-208

Grainger, J., Kiyonaga, K. & Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychological Science, 17*, 1021–1026

Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition, 15(03)*, 594-615.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross language activation survive a change of script? *Cognition, 106*, 501–511.

Hoversten, L. J., Brothers, T., Swaab, T. Y., Traxler, M. J. (2015). Language membership identification precedes semantic access: Suppression during bilingual word recognition. *Journal of Cognitive Neuroscience, 27:11*, 2108-2116. doi:10.1162/jocn_a_00844

Khan, Q. H., & Buchanan, L. (2013). Lexical respresentation and processing in cross-script Urdu-English bilinguals: The case of frequency-balanced and frequency-unbalanced cognates and noncognates. Retrieved from Dissertations and Theses database.

Khan, Q. H., & Buchanan, L. (2006). A study of word frequency in written Urdu. Retrieved from Dissertations and Theses database

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*, 484-499

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149-174. doi: 10.1006/jmla.1994.1008

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morroes, & R. Sykes (EDs.), *Practical aspects of memory,* (Vol. 2, pp. 389-395). London: John Wiley & Sons

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism (Cambridg, England), 13(3),* 373-381. http://doi.org/10.1017/S136672891000009X

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology, 143(3),* 1065-1081. doi: 10.1037/a0035669

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwisterlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. Journal of *Experimental Psychology: Learning, Memory & Cognition, 34,* 12-31.

Litcofsky, K. A., Tanner, D., van Hell, J. G. (2016). Effects of language experience, use, and cognitive functioning on bilingual word production and comprehension. *International Journal of Bilingualism, 20(6),* 666-683. doi: 10.1177/1367006915579737

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience 38 and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50(4),* 940-967.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26(1),* 075-100.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi–English sequential bilinguals: an eye tracking visual world study. *Reading and Writing, 27,* 129-151. doi: 10.1007/s11145-013-9436-5

Miwa, K., Dijkstra, T., Bolger, P., & Baayen, R. H. (2014). Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition*, *17*(3), 445-463.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, *111*(3), 721-756

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese– English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes, 27*, 1563–1583.

New,B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 3,* 516-524.

Potter, M.C., So, K-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23-38. doi: 10.1016/S00225371(84)90489-4

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology, 13*, 199–203.

Rao, C., Vaid, J., Srinivasan, N., & Chen, H-C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: evidence from Hindi/Urdu biliterates. *Reading and Writing, 24,* 679-695. Doi:10.1007/s11145-010-9256-9

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197–212

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education, 21(4),* 360-407.

Sunderman, G. & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28,* 387– 422.

Swender, E., Conrad, D. J., & Vicars, R. (2012). ACTFL proficiency guidelines 39 2012. Retrieved August, 28, 2017.

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1412.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 616–627

Weinreich, U. (1968). *Languages in Contact: Findings and Problems.* The Hague, Netherlands: Mouton & Co. N. V. Publishers

Westbury, C., Buchanan, L., Sanderson, S., & Rhemtulla, M. and Phillips, L.(2003). Using genetic programming to discover non-linear variable interactions. *Behavior Research Methods, Instruments, and Computers*, *35*(2), 202-216.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratfcliff, R. (2011). Individual difference in visual word recognition: Insights from the English lexicon project. *Journal of Experimental Psychology and Human Perception Performance, 38(1),* 53-79. doi: 10.1037/a0024177

APPENDICES

APPENDIX A

HFU-LFE

آےا	Nanny
اردو	Urdu
آلو	Potato
بازار	Marketplace
مولوی	Imam
ادرک	Ginger
داغ	Stain
سالن	Curry
نوجوان	Teens
مسالہ	Spice
مسجد	Mosque
بنے	Woven
ابل	Boil
اشتہار	Advertisement
مسلم	Muslim
پاکستان	Pakistan
ہلدی	Turmeric
ڈراموں	Dramas
خشکی	Dandruff
کباب	Kebab
دواوں	Blessings
ذائقہ	flavor
ذرّات	Particles
قبض	Constipation
روپے	rupees
زلزلے	Earthquakes
شاہ	Shah
ضرورےات	necessities
غےر	outsider
ابتدائی	beginner

LFU-HFE

بٹھائے	Sit
بھاو	Price
سرگذشت	History
بےجان	dead
پروا	Care
تھتا	Board
جانو	Dear
جانوں	Lives
جذبے	Spirit
چارپائی	bed
اطلاعائی	Information
ہق	right
بآپ	Father
ادمی	Man
بھوری	Brown
تماشے	Show
جہگوں	Places
چٹی	White
حالوں	Cases
حکومتی	Government
خت	letter
خداوند	Lord
خوراک	Food
خوشنما	Beautiful
راجکمار	Prince
رونگٹے	Hair
زنانی	Woman
بےانی	Statement
پرزوں	Parts
ماّہ	Month

LFU-LFE

السی	Linseed
رشوت	Bribery
شکرقند	Yam
امتحانوں	Exams
اچار	Pickles
جووں	Lice
بھڑ	Wasp
بھکارےوں	Beggars
بہکنے	Stray
پادری	Pastor
پجارن	Priestess
پروس	neighbor
پستا	Pistachio
پسلی	Rib
پلندوں	Scrolls
پلوں	Bridges
پلک	Eyelash
محلات	Palaces
پنجہ	claw
پنکھ	Feather
پوتر	Grandson
پٹاخے	Fireworks
پھسلن	Lubrication
پہلوان	wrestler
رمالی	Kerchief
رکشا	rickshaw
زلفوں	Locks
سفرائ	Ambassadors
توتا	Parrot
توند	belly

HFU-HFE

مچھلی	Fish
اوپر	Top
منھ	Mouth
بچّے	Children
موضوع	Subject
بادشاہ	King
سر	Head
سرخ	Red
ماں	Mother
لکڑی	Wood
امر	age
سردی	Cold
انسان	human
ہاتھوں	Hands
باغ	garden
بےمار	ill
بےٹا	son
پہاڑی	Hill
تارےخ	Date
سمندر	sea
کہانی	Story
حقےقت	truth
کتاب	Book
خدمات	Services
خط	Letter
درخت	Trees
درےا	river
دن	Day
رات	Night
رقم	amount

English NW		**English NW**	
nanpy	Siy	binseed	zish
urde	Prihe	qribery	tov
sotato	historp	oam	louth
markeuplace	deaj	pxams	phildren
ipam	xare	zickles	lubject
oinger	noard	gice	ging
stiin	deas	masp	heak
curjy	qives	jeggars	rer
teenk	snirit	strab	kother
opice	yed	sastor	jood
qosque	ingormation	yriestess	yge
soven	rigkt	jeighbor	dold
xoil	qather	nistachio	zuman
Aptertisement	uan	zib	hando
ruslim	hrown	ecrolls	zarden
payistan	mhow	yridges	yll
turkeric	qlaces	lyelash	kon
lramas	phite	zalaces	hiyl
dandpruff	xases	olaw	xate
kekab	gozernment	zeather	kea
olessings	lettet	hrandson	storb
flaqor	lorw	oireworks	hruth
partijles	fovd	pubrication	uook
constikation	beajtiful	drestler	sersices
yupees	pyince	berchief	letier
eartheuake	hatr	qickshaw	wrees
shaj	wopan	zocks	piver
pecessities	stateqent	ambassazors	qay
oupsider	varts	varrot	jight
jeginner	yonth	lelly	ymount

APPENDIX B

Northwestern Bilingualism & Psycholinguistics Research Laboratory

Please cite Marian, Blumenfeld, & Kaushanskaya (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research, 50* (4), 940-967.

Language Experience and Proficiency Questionnaire (LEAP-Q)

Last Name		First Name		Today's Date	
Age		Date of Birth		Male ☐	Female ☐

(1) Please list all the languages you know **in order of dominance**:

1	**2**	**3**	**4**	**5**

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	**2**	**3**	**4**	**5**

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language.

(*Your percentages should add up to 100%*):

List language here:					
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.

(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time.

(*Your percentages should add up to 100%*):

List language here					
List percentage here:					

(6) Please name the cultures with which you identify. On a scale from zero to ten, please rate the extent to which you identify with each culture. (Examples of possible cultures include US-American, Chinese, Jewish-Orthodox, etc):

List cultures here					
	(click here for scal	(click here for scal	(click here for scal	(click here for scal	(click here for scal

(7) How many years of formal education do you have? ______ _________________________________

Please check your highest education level (or the approximate US equivalent to a degree obtained in another country):

- ☐ Less than High School
- ☐ High School
- ☐ Professional Training
- ☐ Some College
- ☐ College
- ☐ Some Graduate School
- ☐ Masters
- ☐ Ph.D./M.D./J.D.
- ☐ Other:

(8) Date of immigration to the USA, if applicable ___ ____

If you have ever immigrated to another country, please provide name of country and date of immigration here. __________________ ____

(9) Have you ever had a vision problem ☐, hearing impairment ☐, language disability ☐, or learning disability ☐ ? (Check all applicable). If yes, please explain (including any corrections):

_____________________________________ ____

Language: Urdu

This is my (please select from pull-down menu) language.

All questions below refer to your knowledge of Urdu .

(1) Age when you…:

began acquiring **Urdu** :	*became fluent* in **Urdu** :	*began reading* in **Urdu** :	*became fluent reading* in **Urdu** :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where **Urdu** is spoken		
A family where **Urdu** is spoken		
A school and/or working environment where **Urdu** is spoken		

(3) On a scale from zero to ten, please select your *level of* ***proficiency*** in speaking, understanding, and reading **Urdu** from the scroll-down menus:

Speaking	(click here for scal	Understanding spoken language	(click here for scale)	Reading	(click here for scal

(4) On a scale from zero to ten, please select how much the following factors contributed to you

learning **Urdu** :

Interacting with friends	(click here for pull-down scal	Language tapes/self instructi	(click here for pull-down sca
Interacting with family	(click here for pull-down scal	Watching TV	(click here for pull-down sca
Reading	(click here for pull-down scal	Listening to the radio	(click here for pull-down sca

(5) Please rate to what extent you are currently exposed to **Urdu** in the following contexts:

Interacting with friends	(click here for pull-down scal	Listening to radio/music	(click here for pull-down sca
Interacting with family	(click here for pull-down scal	Reading	(click here for pull-down sca
Watching TV	(click here for pull-down scal	Language-lab/self-instruction	(click here for pull-down sca

(6) In your perception, how much of a foreign accent do you have in **Urdu** ?

(click here for pull-down scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your accent in **Urdu** :

(click here for pull-down scale)

Language: English

This is my (please select from pull-down menu) language.

All questions below refer to your knowledge of

(1) Age when you…:

began acquiring English :	*became fluent* in English :	*began reading* in English:	*became fluent reading* in English :

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where English is spoken		
A family where English is spoken		
A school and/or working environment where English is spoken		

(3) On a scale from zero to ten please select your *level of* ***proficiency*** in speaking, understanding, and reading English from the scroll-down menus:

Speaking	(click here for scale)	Understanding spoken language	(click here for scale)	Readin g	(click here for scal

(4) On a scale from zero to ten, please select how much the following factors contributed to you learning English:

Interacting with friends	(click here for pull-down scale	Language tapes/self instruction	(click here for pull-down sca
Interacting with family	(click here for pull-down scale	Watching TV	(click here for pull-down sca
Reading	(click here for pull-down scale	Listening to the radio	(click here for pull-down sca

(5) Please rate to what extent you are currently exposed to English in the following contexts:

Interacting with friends	(click here for pull-down scale	Listening to radio/music	(click here for pull-down sca
Interacting with family	(click here for pull-down scale	Reading	(click here for pull-down sca
Watching TV	(click here for pull-down scale	Language-lab/self-instruction	(click here for pull-down sca

(6) In your perception, how much of a foreign accent do you have in English ?

(click here for pull-down scale)

(7) Please rate how frequently others identify you as a non-native speaker based on your <u>accent</u> in English:

(click here for pull-down scale)

VITA AUCTORIS

NAME: Nawal Mustafa

PLACE OF BIRTH: Multan, Pakistan

YEAR OF BIRTH: 1993

EDUCATION: Vincent Massey Secondary School, Windsor, Ontario
2011 Highschool Diploma
University of Windsor, Windsor, Ontario
2011 – 2015 B. Sc. (Honors)
University of Windsor, Windsor, Ontario
2016 – 2018 M.A.