Speech Assessment in Bilingual Children: Relationship between Perceptual Judgments of Accent/Comprehensibility and Formal Test Measures

by

Stephanie Limacher

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Communication Sciences and Disorders

University of Alberta

Abstract

There is a paucity of research examining how the presence of an accent may affect speech-language pathologists' (SLP) assessments of bilingual children. Generally, assessment tools normed on monolingual children cannot be used to make diagnostic decisions about whether a bilingual child has a speech delay or disorder. The questions that framed this study were: 1) how do bilingual children's scores compare to their monolingual peers on standardized tests of articulation (*Goldman-Fristoe Test of Articulation, 3rd Edition (GFTA-3)*) and nonword repetition (*NWR*), and SLP's perceptual ratings of accentedness and comprehensibility? and 2) do perceptual ratings of accentedness and comprehensibility correlate with scores from the *NWR* and *GFTA-3*?

The present study assessed 34 children, 17 bilingual and 17 monolingual, using parent report, the *GFTA-3*, the *NWR* subtest from the *CTOPP-2*, and a sentence imitation task. Ten SLPs were presented with audio recordings of sentences produced by the children and asked to rate them for accentedness and comprehensibility. The bilingual group's *GFTA-3* scores were lower than the monolingual group's scores but there was no significant difference between the groups on *NWR* scores. The SLPs' perceptual ratings of accentedness were higher for the bilingual group than for the monolingual group, but there were no significant group differences for the comprehensibility ratings. The comprehensibility ratings were positively correlated with the accentedness ratings, indicating that speech with more of an accent was perceived as harder to understand. The comprehensibility ratings were also negatively correlated with the *NWR* subtest, indicating that the harder the children were to understand overall, the more likely they were to score lower on the *NWR* task. This study did not find that the SLPs' comprehensibility ratings were correlated with the lower GFTA-3 scores. Though the accentedness ratings were indeed higher for the bilingual group as one would expect, the monolingual group also received accentedness ratings, which one would not expect from a group of children raised with only English exposure. Furthermore, the overall correlation between accentedness and comprehensibility ratings suggests that accent may be playing a role in the SLPs' perception of how easy the child is to understand.

Preface

This thesis is an original work by Stephanie Limacher. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Accent and Comprehensibility in Bilingual Children," No. Pro00074246, July 20, 2017.

Acknowledgments

I would like to thank my supervisor, Dr. Karen Pollock, who has had an incredible amount of patience guiding me through this project, keeping me on track through the whole thing despite my constant feelings of uncertainty. Your excitement for this project has kept me moving forward when I began to lose momentum. I must also thank Dr. Elena Nicoladis for filling in for Dr. Pollock when she was on leave. Your frank feedback was much appreciated!

I extend my heartfelt thanks to the parents of the 37 children who agreed to let me test their children for this project and the ten SLPs who agreed to listen to and rate 204 audio clips. Needless to say, none of this would be possible without all of you. Thank you as well to Angela Cullum, who helped me create my experiment in E-Prime - without your assistance, I might still be trying to set-up that experiment! And Alesha Reed for helping me run my last-minute correlations from off-site.

Thank you to my parents; my mother, Gena, who is my best cheerleader, and my father, Dean, for going before me and walking me through what to expect on this journey. Special thanks to my husband, Matt, for your unwavering support through this entire degree, and especially in the moments when I was confident I had bit off more than I could chew. I would not have had the courage to even apply to this program, let alone design and execute this project, without you by my side.

Table of Contents

List of Tables

List of Figures

The Impact of Accentedness

In Arizona in 2010, an unfortunate confusion about of the distinction between accent and intelligibility became a governmental issue when the Arizona State Department of Education decided that any teacher in front of a student should be a non-accented speaker, therefore disqualifying any teacher who had yet to master standard English articulation (Tomic, 2013; Derwing et al., 2014). The rationale for this assumption was the assumption that accented speech would be too confusing for students and would hamper learning. This use of accent as a measure of language competence was swiftly decried by the academic community, prompting the University of Arizona's Department of Linguistics (2010) to release a statement, declaring:

> It is our position, based on decades of scientific investigation into the nature of language and of language acquisition and learning, that such a policy undermines the effectiveness of the teaching and learning of English by non-native speakers and may lead to additional harmful socioeconomic effects. (p. 1)

One of the many reasons that the nineteen professors signed this statement was that accent does not guarantee "unintelligibility" (i.e., impossible to understand utterance) and that dialects (i.e., variations of a language that belong to specific regional and/or social groups) can also be difficult for a listener to understand, but this policy did not make that distinction.

There is not one way of speaking English; there are numerous dialectical varieties before one even considers foreign-accented speech. "Accent" refers to the extent to which speech output does not match the articulatory targets required for a particular language's sound inventory (Munro & Derwing, 1999; Trofimovich & Isaacs, 2012). Essentially, the short sightedness of the Arizona Department of Education's policy was that it focused on the 'otherness' that a foreign accent implies, and was used as a discriminatory measure under the guise of educational best practice for the sake of the students. Though

the Arizona Department of Education focused on accented speech from foreign speakers, this indication of otherness can be wielded against native English speakers who speak a non-standard variety of English as well. A population of Sub-Saharan Africans in Vancouver who speak English as their mother tongue and were educated with the British curriculum were surveyed, and they reported feeling that their accent was the cause of discrimination they had faced (Creese, 2010; Tomic, 2013).

The discrimination described by this African-Canadian population is unfortunately not an uncommon phenomenon, and the effects of accentedness are far-reaching. A study of discrimination towards foreign East and South Asian post-secondary students in Canada (Houshmand, Spanierman, & Tafarodi, 2014) found that ridicule or teasing of accent was one of the racially centred microaggressions that the students experienced with regularity. As a result, students felt maligned or unfairly judged in the school community, both at the hands of the professors and their fellow students. Beiser and Hou (2016) surveyed immigrants and refugees to Canada to examine the mental health effects of pre-migration and post-migration trauma on the individual; one of the measures of post-migration trauma was whether the individual had ever felt discriminated upon based on racial distinctions, including accent. The results were that 38% of the immigrants and 56% of the refugees surveyed did experience some form of post-migration trauma in the form of discrimination, including discrimination because of their accent. Unfortunately, the study only mentions that an accent is one of the many reasons that one may be discriminated against but does not look at each possibility individually, so the specific extent to which accent played a role in this discrimination is not known. Similarly interested in the effects of accent, George and Chaze (2014) wanted to determine how the experience of foreign-trained professionals differed from locally-trained professionals. They surveyed both of these populations within the Engineering field and accent was reported to be a factor in discrimination as it is often used as an indication of otherness. Soji Akomolafe MSC (2013) was more precise in defining the limitations that

accent placed on individuals by studying the “Glass Ceiling Syndrome” that bars a foreign-accented speaker from high-level positions in the workplace and, specifically, academia. They found that foreign-accented speakers are routinely hired for mid-level positions, but are disproportionately absent from higher-level positions.

What is barring foreign-accented speakers from positions of leadership within the workplace? Is it because they are considered unintelligible or incomprehensible? In an understandable effort to avoid this discrimination surrounding accentedness, a controversial sub-genre in the Accent Reduction realm has appeared in the English Language Learning (ELL) community (also called English as a Second Language (ESL)). Derwing, Fraser, Kang, and Thompson (2014) discuss the considerable debate about the ethics of accent reduction services and, though accent is not synonymous with intelligibility or comprehensibility, they become interchangeable in practice - usually to the detriment of anyone who does not have a Standard English accent. Derwing et. al. (2014) point out that many of the services are offered by unqualified teachers who make unsubstantiated claims about the effects of their services, treat accentedness as a negative character trait, or as an underlying pathological phenomenon (as may be the case if the services are provided in a clinical setting, or provided by a professional with clinical certification, such as a speech-language pathologist). This further reinforces that foreign accents are viewed negatively and are positively correlated with unintelligibility, a fact that may cause the language learner undue stress and could be a demotivating factor in their language learning journey.

Though unjustified, the misunderstanding in the Arizona Department of Education is therefore not surprising and embodies the assumptions that are made around the intersection between the concepts of accent and intelligibility, or comprehensibility. Contrary to these assumptions made by the Arizona Department of Education, and by many others, Derwing & Munro (2014) found that not all speech that was rated as heavily accented a 9-point scale (i.e., a rating of 7-9) was deemed to be unintelligible. So

the two are not mutually exclusive - one can have a heavy accent while also being intelligible or comprehensible. Despite this, there is an apparent contradiction stemming from two underlying principles that seem to be working against each other: the "nativeness principle," which seeks to have the speaker assume as native an accent as possible, and the "intelligibility principle," in which the speaker is encouraged to relay the message so that it is clear to the listener, whether an accent is present or not (Levis, 2005). Furthermore, one must take into consideration listener experience, as what can be considered intelligible is a highly subjective construct. To add to the complexity of the debate, there are numerous dialects of English, which make accent a moving target as well. What can be considered an accent to one region or strata of speakers is completely different to someone who hails from a different region or strata (Levis, 2005). Even within Canada, the accent of Eastern Canadian English and Western Canadian English is distinct enough that it must be taken into consideration when assessing a client's speech and language production.

Rating Accentedness, Intelligibility, and Comprehensibility

The aforementioned studies all underscore how complicated and often misunderstood attitudes surrounding accentedness can be. This is because the very concepts of accent, intelligibility, and comprehensibility - and the extent to which the three are used interchangeably - are complicated and subjective, and demand careful consideration when requiring a listener to rate any of the three. Accented speech production may contain either segmental features (word duration and stress, vowel duration and stress, voice onset time, consonant substitution) and/or suprasegmental features (articulation rate, speech rate, pitch, tonal differences, fluency, pauses) that differentiate it from native speech production (see Sereno, Lammers, & Jongman (2010) for a comprehensive summary of how segmental and suprasegmental features can affect speech production and perception between languages). As much as a speaker's production of speech affects the perception of accentedness, so to does the listener's

experience with accented speech affect perception and ratings of accented speech (Fledge, 1984; Derwing & Munro, 1997; Munro & Derwing, 1999; Derwing, Rossiter, & Munro, 2002).

Rating accentedness has taken many forms in the research studies. Jesney (2004) reviewed various accentedness rating methods and found that scales are most often used, though these are merely umbrella terms for a wide range of specific types of rating scales that have been employed. Likert scales are often used, and can range from three to ten points, though many researchers chose a nine-point scale. Sliding scales require the listener to move the point on the scale to precisely where they would like it to go, resulting in ratio data for the benefit of the researcher, and more precision on the part of the rater. Direct magnitude scales have the listener rate the first sample, and then multiply ratings of subsequent samples comparative to the first sample. To identify if there is an accent, or no accent, speech samples can range anywhere from 30 milliseconds in length to a full 2 minute clip of the speaker. Jesney (2004) notes that while Fledge (1984) deemed 30ms to be an adequate length of time for identifying accented speech, accuracy in ratings increased across various studies from 69% to 89% when listeners heard a full phrase.

Where accent refers to the speaker's ability to meet articulatory targets, "intelligibility" refers to the extent to which the listener can identify word-for-word exactly what the speaker has said. Contrastively, "comprehensibility" refers to the extent to which the listener is able to understand, or garner the general concept of what is being said (Munro & Derwing, 1999). Trofimovich and Isaacs (2012) refer to Munro and Derwing's (1999) definition of intelligibility and suggest that, in practice, intelligibility is more of an umbrella term to denote the general understanding of an utterance (i.e. if one were to transcribe an utterance, how closely would it match the original spoken utterance). Comprehensibility is more appropriate for use with rating scales because it allows the listener to indicate how easily they felt they understood the utterance. This study will not be asking listeners to identify or

transcribe the words from the recordings, but to rate the extent to which they felt they were able to understand the utterances and will therefore use the term *comprehensibility* rather than *intelligibility*.

Children, Bilingualism, and Accentedness

Thus far, this discussion has been largely centred around accent and adult speakers. However, accentedness may also have adverse effects on children, who could be held back in schools or mistakenly identified as having a speech-language impairment (Hack, Marinova-Todd, & Bernhardt, 2012). Given the established possability for listeners to be biased against accented speech, it is especially important for Speech-Language Pathologists to take note of how accent may affect perceptions of comprehensibility as this is a tool commonly used in making clinical judgements about the presence and severity of speech sound delays or disorders. One of the reasons that it is so difficult to parse out the limits of where natural language development ends and a delay begins is because bilingual children develop language differently than monolingual children in terms of rate of acquisition and proficiency. Their development is affected by factors such as the nature, extent, and quality of the dominant language (L1) and additional language (L2) input and interaction between these languages - though exactly how this happens is still not well understood. Bilingual development is complicated precisely because of the complicated nature of the underlying development. For children, development must account for two or more language systems (Hambly, Wren, McLeod, & Roulstone, 2013). There is sufficient evidence of transfer between lexical and semantic networks, though a full discussion of this is outside the scope of this study (for review, see Kartushina, Frauenfelder, & Golestani, 2016). However, the interaction between L1 and L2, and how this interaction affects language output, is important for educators and speech-pathologists to understand when assessing bilingual children.

The lack of general knowledge about this interaction between L1 and L2 has led to some assumptions about learning an additional language during the early child language developmental years.

In a systematic review of bilingualism and speech production, Hambly et al. (2013) specifically examined bilingual children's language acquisition and addressed the idea that acquiring two or more languages causes a delay. They also addressed the idea of accentedness in children in terms of phonological transfer between the child's first or dominant language (L1) and their second or non-dominant language (L2). The conclusion of this systematic review of 66 studies was that there is evidence that the child's L1 influences his or her L2, as well as evidence of transfer from the child's L2 towards his or her L1. The nature of this transfer is complicated by factors such as age of acquisition (AoA), the amount of time the child has been in the L2 environment, and the amount of exposure he or she receives in each language (Paradis, Genesee, & Crago, 2011; Hopp & Schmid, 2013; Verdon, McLeod, & Wong, 2015).

AoA is an important factor to consider when qualifying bilingualism because whether one is a simultaneous bilingual or sequential bilingual will change the trajectory of development for the child and will impact the nature of transfer each language has on the other; specifically, in terms of phonological transfer, the younger the child is when he or she begins to acquire the L2, the more likely the L1 and L2 will develop simultaneously and have less impact on the other (Hopp & Schmid, 2013; Kartushina, et al., 2016). Simultaneous language development refers to children who receive input from two language systems starting from the time of birth, where sequential language development refers to children who receive input in their L1 at the time of birth but do not begin receiving input for the additional language system until later, though exactly when this additional language input begins requires further classification. A systematic review conducted by Kartushina et al. (2016) explored the nature of transfer in bilingual speakers, classifying AoA into the following categories: simultaneous bilinguals (who acquire the two languages at the same time, or simultaneously), very early bilingual (speakers who acquire their L2 before 3 years old), early bilingual (those who acquire their L2 before 7

years old), and later bilingual (those who acquire their L2 after 8 years old). They found that as the child ages, L1 informs and affects the phonological production of L2. Very early bilinguals are able to pronounce the speech sounds of both L1 and L2 in a way that is consistent with a monolingual speaker, but later bilinguals have a harder time meeting the articulatory targets of the speech sound inventory (Kartushina et al., 2016).

Speech-Language Pathology and the Assessment of Bilingual Children

Bilingual language development is complicated and unfortunately there are factors that may block bilingual children's access to the speech-language services they need, or end up erroneously putting them on a therapist's caseloads. These factors include a general lack of knowledge regarding the trajectory of bilingual development, parents or teachers either not recognizing signs of impairment or being overly cautious, or speech-language pathologists not feeling comfortable assessing bilingual children or relying on tools that are not evidence-based (for review, see Paradis, Schneider & Sorenson Duncan, 2013). Young and Westernoff (1999) identified four barriers to service for bi- or multilingual clients: 1) cultural differences, which can affect assessment on the SLPs part, and treatment decisions on the family's part; 2) language, which can act as a barrier to service, both in the client not knowing about services available and a lack of SLPs or interpreters to carry out a full assessment; 3) the adequacy of SLPs` training in how to work with bilingual clients; and 4) professional matters, or more specifically the lack of resources available to speech-language pathologists and audiologists.

More recently, Williams and McLeod (2012) and McLeod and Baker (2014) reported that very few SLPs say they use standardized tests with bilingual children, but that they use informal measures, and rely most heavily on parents to translate for them, or otherwise use an interpreter. In Canada, D'Souza, Kay-Raining Bird, and Deacon (2012) conducted a survey of speech-language pathologists regarding their assessment and treatment of bilingual and multilingual clients. Similar to Young and

Westernoff (1999), Williams and McLeod (2012), and McLeod and Baker (2014), found that SLPs feel inadequate and ill-prepared to serve their bi- or multilingual clients, and generally for the same reasons. Despite the reported lack of confidence on the part of SLPs, bilingual speakers comprise a significant portion of the population in Canada and require the same level of service that a monolingual speaker would. According to the 2016 Canadian census data, 7.3 million people had a mother-tongue that was neither French nor English (the two official languages of Canada). This accounts for 21.1% of the population. Therefore, since the majority of these people learn English and/or French, as of 2016, at least 20% of the Canadian population is sequentially bilingual. The "2016 Annual Report to Parliament on Immigration," given by John McCallum, PC, MP and Minister of Immigration, Refugees and Citizenship, reports there were 271,845 immigrants, refugees, and Temporary Foreign Workers admitted to Canada in 2015.

With the numbers reported in 2016 census, and the 2016 report on immigration, it is safe to conclude that bilingual clients are a rapidly increasing population who require the same access to services that monolingual clients would. Within the realm of child speech assessment and treatment, standardized tools that SLPs regularly use for assessment of the monolingual children on their caseloads are normed on monolingual populations and are therefore not appropriate for use on bilingual and multilingual populations, since these children may receive artificially low scores due to differences in articulation (Hack et al., 2012). This leaves SLPs with little choice but to use their perceptual judgment, the limited normed tools that exist (depending on the child's languages), and parent report to assess whether or not a delay or disorder is present. While this is not meant to take a critical look at the job being done in the field, perception is a subjective measure and changes based on amount of experience one has with accented speech; best practice requires objective measures in addition to subjective measures for the purposes of diagnosis.

In addition to not always having the appropriately translated and normed tools necessary to complete the job, and not always having adequate training or confidence in working with bilingual or multilingual clients, SLP's are encountering these clients at increasing rates. The research in bilingual language development and processing has increased proportionate to the increased prevalence of bilingual students in schools and on SLP caseloads (Paradis, 2010), but the study of how accent may affect assessment and treatment has not. This is worrying when one considers the increasing number of bilingual children in schools that do not have adequate resources to accommodate them.

The challenge also arises, as Paradis (2010) points out, that what we are looking for in a child with a language delay or disorder may look very similar to what we see in typically developing children who are still acquiring language skills. This overlap between characteristics can be a confounding factor in both over and under-identification of bilingual children requiring speech services. What we do not know exactly, is the role that accent plays in confusing the boundaries between language acquisition and language delay or disorder. They are certainly all contributing factors, but it is interesting to consider whether one may be more so than the others. The implication may be that if a typically developing child's speech is heavily accented, and they make different lexical and syntactic choices and errors, they may be misdiagnosed and added to an SLP's caseload.

Despite a seeming dearth of useful assessment tools for assessing bilingual populations, recent research is finding some assessment options that SLPs may or may not be aware of. McLeod and Verdon (2017) created a tutorial for assessing multilingual children. They warn against using assessments in the dominant language due to their incFapacity to fully account for the multilingual children's speech production. Among their recommendations for assessment is the NWR task. The experience a child brings to any assessment factors into the results of that assessment; this is true for both bilingual children and children of a low socioeconomic status (SES). The Nonword Repetition

(*NWR*) task has been described as a fairer assessment for both of these populations due to its ability to remove the linguistic processing knowledge needed to complete other language assessments (Paradis, 2013; Ebert, & Kohnert, 2016; Chiat & Polišenská, 2016). However, there is still some question as to how effective *NWR* tests are for bilingual children, and studies have not found that bilingual children score similarly to their monolingual peers in every instance (see Brandeker & Thordardottir (2015) for a review of NWR in bilingual populations). Instead, using a linear mixed regression, Sorenson Duncan and Paradis (2016) found that age, English vocabulary size, amount of English exposure, and L1 were significant predictors of children's *NWR* performance. This study found that 29% of the bilingual children scored lower than their monolingual peers on *NWR* measures. However, the researchers did not recommend discontinuing use of the *NWR* task with bilingual populations altogether, noting that differing results between studies may be due to differences in populations, and rather suggested further research and development of an ELL specific *NWR* task for use with bilingual populations. Additionally, Brandeker and Thordardottir (2015) found that *NWR* tasks have good diagnostic accuracy for language impairment in bilingual populations and also recommend this task for speech assessment purposes.

In addition to the *NWR* task, parent questionnaires can be a useful tool in helping to identify a possible language delay or disorder. Paradis and colleagues at the University of Alberta have produced two such parent questionnaires to use with bilingual clients. *The Alberta Language Development Questionnaire* (*ALDeQ*) (Paradis, Emmerzael, & Sorenson Duncan, 2010) and the *Alberta Language Environment Questions* (*ALEQ*) (Paradis, 2011) are useful in helping SLPs gain sense of the child's language abilities in their first language. These tools also examine the child's developmental history (i.e., insofar as delays in other areas may indicate increased risk of the child also having a language disorder), whether there is any family history that may also be an indicating factor of delay or disorder, and how much exposure the child gets in both L1 and L2.

SLP Perceptual Ratings of Accent and Intelligibility

In 2012, Hack, Marinova-Todd, and Bernhardt were interested in this very problem of children being misdiagnosed and erroneously added to an SLP's caseload. They designed an exploratory study to look at three questions: whether there is divergent development of consonants in children's L1 (Cantonese or Mandarin) and their L2 (English); whether bilingual children had lower scores on a standardized articulation test (*Goldman-Fristoe Test of Articulation, 2nd Ed. (GFTA-2)*) than their monolingual peers; and whether an SLP's perceptual rating of accent and developmental level would correlate with the child's *GFTA-2* score. They recruited 29 bilingual children (16 Cantonese-English speaking children; 13 Mandarin-English speaking children) who spoke only their L1 at home (per parent report). The study's control group consisted of 25 monolingual children who only spoke English at home (per parent report). Both groups were administered the *GFTA-2*, as well as a sentence repetition task consisting of 3 sentences.

Ten monolingual SLP raters, blinded to the results of the child's *GFTA-2*, were presented with the randomized sentences and were asked to first determine whether the speech was accented or not. If the speech was judged accented, the SLPs were asked to rate the sentence on a 9-point Likert scale for accentedness. If the speech was not considered accented, the SLPs were asked to rate the utterance on a 9-point Likert scale for developmental level. The results found that their first hypothesis, that the Cantonese and Mandarin speaking children had a similar trajectory of consonant development as their monolingual peers, held true but with a few exceptions (e.g., the Cantonese and Mandarin children had more consonantal mismatches overall on the GFTA-2, including some that were not expected for their age, than their monolingual peers, and they had additional uncommon phonological patterns). In regards to their second hypothesis, that their bilingual subjects would score lower on the articulation test compared to their age matched peers, they found that this was indeed the case. Lastly, in terms of their

final hypothesis that the SLPs perceptual rating of accentedness would correlate with the child's *GFTA-2*, they did not find that such a correlation was present.

One of the questions in Hack et al.'s (2012) study is built on the assumption that if an accented bilingual child were to take an articulation test, he or she would score lower due to his or her accent, thereby leaving the interpretation of the results open to a possibly erroneous judgement about why the score is falling below normal limits. Indeed, upon testing this hypothesis, the researchers found that the bilingual sample received lower scores than their monolingual counterparts, but the study did not find that the difference was statistically significant. Nonetheless, if the SLP were to score the test as it requires (which is not encouraged because the norms cannot be used for scoring as bilingual children are not included in the population that the test was normed on), the child may score within a standard deviation below the mean, and perhaps even more children than normal may fall within the "at-risk" zone.

Given the fact that articulation tests may not produce trustworthy scores for a bilingual child, SLPs may need to parse where accent ends and disordered speech begins. The task to decide where accent ends and disorder begins is complicated because the speaker may additionally have prosodic, syntactic, or semantic errors, and may code-switch (inserting words or phrases, or even grammatical structures or rules at the discourse level from the speaker`s alternate language into the current language being used during conversation) (Martin, Krishnamurthy, Bhardwaj, & Charles 2003). All of these add complexity to the job of the SLP, who must make a problem/no-problem decision. So, while accentedness is just one piece of the puzzle, it is an under-studied piece that impacts the assessment and treatment of bilingual and multilingual clients, and the relationship between accentedness and comprehensibility deserves attention because the tools used to assess and diagnose bilingual clients for

LI are not as varied as the tools available for monolingual clients, and SLPs' themselves admit that they do not feel confident in the assessment and treatment of bilingual clients.

Purpose and Research Questions

In an effort to contribute to the growing evidence base regarding speech assessment with bilingual children, especially in light of the significant percentage of the population whose first language is neither English nor French in Canada, this particular study was designed to try and bridge the gap between Hack et al.'s findings regarding the impact of accentedness on speech assessment and some of the conflicting but overall positive evidence (Sorenson Duncan & Paradis, 2016) regarding the use of nonword repetition tasks with bilingual children. Furthermore, given the previously identified lack of confidence within SLPs who work with bilingual children, this study hoped to find a change in current confidence levels and to examine whether these findings might somehow strengthen the confidence of SLPs working with bilingual children.

Taking into consideration the evidence collected from previous studies, the purpose of this study was to further examine the extent to which there is a correlation between SLPs accentedness ratings and scores obtained via a standardized articulation test. This study sought to partially replicate the 2012 Hack, Marinova-Todd, and Bernhardt study to confirm the findings that the bilingual children had lower scores on the *GFTA-2* in comparison to their monolingual peers, as well as the finding that SLPs perceptual ratings of accentedness did not correlate with the *GFTA-2* scores. The present study also added SLPs' perceptual ratings of comprehensibility rather than developmental level and the use of a *NWR* task.

This current study was interested in answering the following questions:

1. How do bilingual children's scores compare to their monolingual peers on the *GFTA-3*, *NWR*, and SLP's perceptual ratings of accentedness and comprehensibility?

2. Do SLP perceptual ratings of accentedness and comprehensibility correlate with scores from a nonword repetition (*NWR*) task and/or an articulation test (*GFTA-3*)?

Our predictions regarding these questions were: 1) the bilingual children will score lower on the *GFTA-3*, and similarly in *NWR* than their monolingual peers and the monolingual children will receive low accentedness ratings, and high comprehensibility ratings while the bilingual children will receive less extreme ratings of accentedness and comprehensibility than their monolingual peers; 2) the SLP's perceptual ratings will correlate with the *NWR* test.

Methods

Child Participants

A total of thirty-seven child participants were tested for this study. The inclusionary criteria specified that the children were five years of age (+/- 1 year), were in daycare or kindergarten, and had no history of hearing impairment or speech-language intervention services. The parents were asked if they had any concerns about their child's speech or language development in their L1. In addition, the *Intelligibility in Context Scale* (McLeod, Harrison, & McCormack (2012), a parent rating scale that measures the child's intelligibility to listeners with varying degrees of familiarity, ranging from very familiar (e.g., parents) to unfamiliar (e.g., strangers), was used as a control for ensuring the children's speech and language development was not impaired in their L1. The combination of the parent report and the examiner's judgment were used to qualify children for the study.

There was only one concern noted by a parent, which had to do with her child's hoarse vocal quality. The parent revealed that the child had been assessed by an SLP and diagnosed with vocal nodules resulting in a hoarse vocal quality. The child was not excluded from the study as the parent was not worried about speech or language development. A total of three children were excluded from the study. One bilingual child was excluded from the study due the presence of an undiagnosed speech

impairment. This child was referred to Corbett Clinic at the University of Alberta for assessment and intervention. Two monolingual children were excluded from this study because they were older than any of the bilingual children and could not be matched to ensure equality of variance between the two groups.

The children included in the study comprised two participant groups consisting of 17 bilingual children and 17 monolingual children. The mean age of the bilingual group was 55.71 months and the mean age of the monolingual group was 59.06 months. A independent *t*-test showed that the difference between groups' (bilingual and monolingual) mean age was not statistically significant (t = -0.983, df = 32, p = 0.333, two-tailed). See Appendix A for the complete bilingual participant data and Appendix B for the complete monolingual participant data. The bilingual children spoke their L1 primarily in their home and received their L2 (English) exposure primarily through daycare or school. The mean age at which the bilingual children started to receive consistent exposure to English was 20.82 months (about 1;8), though the range was from birth (0 months) to only two months of consistent exposure. The monolingual children's L1 was English and they spoke English exclusively at home (per parent report) regardless of any other language exposure (e.g., two of the children had started French immersion at school 6 months prior to testing).

In a departure from the methodology of the Hack et al. (2012) study, this study did not control for the bilingual children's L1 in an effort to simulate an SLP's real-world caseload, as well as avoiding restrictions on the number of eligible children. Five of the children's L1 was Portuguese. The remainder of the children spoke different languages, including: Hungarian, Nepalese, Swedish, French, Russian, Persian, Hindi, Chinese, and Spanish. Three children spoke more than one language at home; these L1 combinations included French/Arabic, Swedish/Romanian, and Hindi/Marathi.

Recruitment. All of the child participants were recruited in one of two ways: a public facebook post requesting that any interested parents whose child or children qualified for the study contact the examiner, or through letters that were distributed to families whose child or children fit the research criteria at daycares in and around the University of Alberta area (within a five-kilometer radius of the university campus). These letters explained the nature of the research and detailed what the children would do as a part of the study. Free hearing screenings were offered as compensation for participation in the study. A total of 22 child participants were recruited via the open facebook post, and 15 child participants were recruited from the letters sent home by daycares. A total of 11 hearing screenings were completed as compensation for participation.

Instruments

The following assessments were used to test the child participants.

Alberta Language Environment Questionnaire (ALEQ). The parents of the bilingual participants completed the *ALEQ* (Paradis, 2011) and were given a score for language use in the home, L1 environmental richness, and English environmental Richness. Language use in the home scores range from 0.00-1.00. A lower score indicates maintenance of the child's L1 in the home, while a higher score indicates a shift towards primarily English use in the home. The English and L1 richness scores were also calculated with a range from 0.00 to 1.00. For both the English or L1 richness scores, a lower score indicates less exposure, while a higher score indicates more exposure to the language. Lastly, the *ALEQ* allowed for examination of where the child was born, what language the mother and/or father, siblings, and other caretakers primarily speak to the child, and how the child primarily responds to each individual.

Intelligibility in Context Scale (ICS)*. The Intelligibility in Context Scale (ICS)* (McLeod, Harrison, & McCormack, 2012), is a parent rating scale that measures the child's intelligibility across

varying communication partners, ranging from very familiar (e.g., parents) to completely unfamiliar (e.g., strangers). Parents were asked to complete the *ICS* in relation to their child's L1. Each category is rated on a scale of 1 (Never) to 5 (Always) and the scores are totalled for an overall average out of 5.

Goldman Fristoe Test of Articulation, 3rd Ed. (GFTA-3). As a part of this study was to replicate the Hack et al. (2012) study, we used the *Goldman Fristoe Test of Articulation, 3rd Edition (GFTA-3)* (Goldman, Fristoe, 2015), which is a newer edition of the same test used in the previous study. This test was used to collect information on the child's production of English consonants in different word positions. In the administration of the test, the children were asked the question prompt given by the book (e.g., "what is this?") to elicit a spontaneous response. If the child could not answer or did not provide the correct answer, delayed imitation was used (e.g., "This is a lion. It has a big mane. What is it called?). The child's responses were recorded phonetically with broad transcription.

Scoring was completed following the test manual instructions, which breaks down speech-sound errors according to word placement (eg., word-initial, medial, or final). Speech sound distortions (e.g., a lateralized /s/) and incomplete and/or absent productions (e.g., the child leaves the sound out or does not attempt the word) are considered errors in scoring. If the child produces an additional sound, these are recorded but are not considered errors in scoring; similarly vowel distortions do not count as errors. All errors across each word position are tallied to produce a final raw score. Raw scores are converted to standard scores, where the mean is 100 and standard deviation is 15. Interjudge reliability was measured by having a second listener independently transcribe the audio-recorded responses from 5 randomly selected children (3 bilingual and 2 monolingual). Point-by-point agreement on broad transcription averaged 95%.

Nonword Repetition Subtest. Sorenson Duncan and Paradis (2016) used the *Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2)* (Wagner, Torgesen, Rashotte, & Pearson,

2013) for their study as the *Nonword Repetition (NWR)* subtest was developed for Standard American English and is commonly used clinically in Canada. For these reasons, this study also used the *CTOPP-2 NWR* subtest. The non-word stimuli were presented using a Samsung tablet. The child was instructed to listen carefully, as the stimuli could not be presented more than one time. Responses were recorded and scored according to the test manual instructions, which stipulates that the word in its entirety must be produced as written on the test form and any speech-sound error results in a score of 0. Words produced identical to the test form receive a score of 1. Correct responses are tallied to produce a final raw score. The results are reported as scaled scores, where the mean is 10 and the standard deviation is 3. As with the GFTA-3, a second listener independently scored the audio-recorded responses from 5 children (3 bilingual and 2 monolingual). Point-by-point agreement on correct/incorrect responses averaged 95%.

Sentence Imitation Task

The children were presented with picture support using a tablet and were asked to repeat three sentences via direct imitation. The younger children (from about 4;0-4;5) sometimes needed the sentence to be given to them in chunks rather than in its entirety as it proved too much to repeat all at once. The sentences are as follows:

- The elephant ate a banana plant
  [ði'ɛləfənt eɪɾ ə bən'æ̃nə plæ̃nt]
- Two big mice chase one little black cat
  [t^{h}u bɪg mʌɪs ʧeɪs wʌ̃n 'lɪtl̩ blæk khæt]
- Five sheep get on a long train
  [faiv ʃip gɛɾ ɑn ə lɑ̃ŋ tɹẽɪñ]

These sentences were developed by Hack and colleagues (Hack et al., 2012) for their sound inventory, and contain later developing sounds that would present a challenge to the monolingual control group in addition to their bilingual sample.

Testing Procedure

The children were tested at Corbett Hall, at their daycare centre, or in their homes. Each assessment was recorded on an Olympus Digital Voice Recorder VN-5200PC. Parents filled out the *ALEQ* (bilingual children only) and *ICS*, and the children were administered the *GFTA-3* and *NWR* subtest from the *CTOPP-2.* Following the formal testing, the children were asked to repeat three sentences, which were presented to the child by the examiner while looking at picture prompts on an iPad.

SLP Participants (Listeners)

Ten speech-language pathologists (SLPs) were contacted by email requesting participation or were referred to the study by other SLP participants. The SLPs were either affiliated with the University of Alberta (as clinical educators or volunteers) or worked on the Inclusive Learning team with the Edmonton Public School Board. SLPs were given a $5 Starbucks card as compensation for their participation.

Nine of the SLPs reported that they were functionally monolingual in English, and one reported being able to speak "some Spanish." The SLPs all reported having bilingual clients on their caseload, with the average ranging from 2 bilingual clients at a time to 60% of their caseloads being bilingual clients. Before the SLPs listened to the stimuli, they were asked to complete a survey that asked the following questions, which were rated on a 8-point Likert scale ranging from 1 (not comfortable) to 8 (completely comfortable):

- On average, how many bilingual clients do you usually have on your caseload?
- How comfortable would you rate yourself in assessing bilingual clients?
- How comfortable would you rate yourself in separating accent from disordered speech?
- How comfortable would you rate yourself in rating comprehensibility of bilingual clients?

Following completion of the testing, the SLP raters were asked to rate the following questions, which were rated on an 8-point Likert scale ranging from 1 (very easy) to 8 (very difficult):

- How easy / difficult did you find it to rate accentedness based on the sound files you heard?
- How easy / difficult did you find it to rate comprehensibility based on the sound files you heard?

Additionally, the SLP raters were asked to describe how they made a decision about rating accentedness for each sound file, and how they made a decision about rating comprehensibility for each sound file. Appendix C shows the SLP raters' numerical questionnaire responses.

Listener Ratings

Likert Rating Scales. Similar to the rating method used by Hack et al. (2012), 9-point Likert scales for Accentedness and Comprehensibility were used to judge each sentence in the sentence repetition task.

- ***Accentedness.*** A rating of 1 indicated that the speaker was completely native-like with no non-native accent, while 9 indicated that the speaker had a strong foreign-accent.
- ***Comprehensibility.*** A rating of 1 indicated that the speaker was completely comprehensible and a rating of 9 indicated that the speaker was completely incomprehensible.

Rating Procedure

The SLPs were presented with recordings of the children's sentence repetitions using E-Prime 3.0 Software (Psychology Software Tools, Pittsburgh, PA), and were asked to rate each sentence. To familiarize themselves with the procedure, the SLPs rated two practice sentences and given the opportunity to ask questions before being presented with the actual stimuli. The SLPs listened to the full set of 102 sentences twice and rated them first for accentedness. After a short break the SLPs listened to the same set of sentences and rated them for comprehensibility. The order of presentation of sentences

within each condition was randomized by E-Prime. The SLP ratings were averaged together for each sentence, and the three sentences were averaged for each child.

Results

Descriptive Participant Data

Alberta Language Environment Questionnaire

The L1 Richness scores for the bilingual participants ranged from 0.1 to 0.65 (M = 0.36; SD = 0.18) and the English Richness scores for the bilingual participants ranged from 0 to 0.81 (M = 0.6; SD = 0.2). The scores for the bilingual group are reported in Appendix A. In addition to calculating language use in the home and English/L1 environmental richness, parents were asked whether the child was born in Canada. 64.7% of the bilingual children were born in Canada and 35.3% of children came to Canada with their parents. Few of the parents filling out the questionnaire added any qualitative information, though one parent did qualify that the child is only allowed to speak Russian at home, while another parent reported that the child prefers to speak English at home but they try to encourage her to speak Nepalese. Another parent reported that the child switches between languages (French and English) depending on who is speaking to him.

The *ALEQ* also asks what language the mother and father individually speak to their child, and in what language the child responds. 58.8% of the mothers and/or fathers reported that they usually or only speak the L1 to the child, while 32.3% speak half-and-half. Only 9% of the mothers and/or fathers reported usually or only speaking English at home. Conversely, 47% of the children were reported to usually or only speak the L1 to the parents, while 26.5% of the children reportedly speak half-and-half to their mothers and/or fathers and 26.5% of the children usually or only respond in English.

Intelligibility in Context Scales

The intelligibility scores for the participants across all listeners (e.g., familiar to unfamiliar) ranged from 3.85 to 5 (M = 4.42; SD = 0.36). An independent *t*-test showed that the difference between group (bilingual and monolingual) *ICS* scores was not statistically significant (t = -0.085, df = 32, p = 0.933, two-tailed). The scores for the bilingual group are reported in Appendix A and the scores for the monolingual group are reported in Appendix B.

Research Questions

The first research question that frames this study asked how bilingual children's scores compared to the their monolingual peers on the GFTA-3, NWR, and SLP's perceptual ratings. The second question asked whether SLP perceptual ratings of accentedness and comprehensibility correlate with scores from a nonword repetition (NWR) task and/or an articulation test (GFTA-3). The following results obtained from the collected data address these questions.

Bilingual versus Monolingual Group Comparison

GFTA-3 and NWR scores. The bilingual and monolingual children's scores from GFTA-3 and NWR are summarized in Table 1. Additional details are provided in Appendix D. The bilingual children's mean performance on the GFTA-3 (M = 90.29) was lower than that of the monolingual children (M = 98.24). Four of the bilingual children had scores below average (i.e. > 1.25 SD below the mean) when compared to the test norms, but all of the monolingual children's scores were within the normal range. An independent *t*-test showed that the difference between groups (bilingual and monolingual) was statistically significant (t = -2.037, df = 32, p = 0.050, two-tailed). The magnitude of differences in the means (mean difference = -7.94, 95% CI: -15.88 to -.001) was moderately large (d = -0.71).

The bilingual children's performance on the *NWR* subtest (M = 7.18) was marginally better than the monolingual children's (mean = 7.06) (see Table 1). Three bilingual children had NWR scores that

were below average (i.e. more than 1.25 SD below the mean for the normative sample). Surprisingly, five of the monolingual children also had below average *NWR* scores that were below average, despite having *GFTA-3* scores within the normal range. An independent *t*-test showed that the difference between the two groups (bilingual and monolingual) was not statistically significant (t = .129, df = 32, p = 0.898, two-tailed).

SLP ratings of accentedness and comprehensibility. The SLPs' ratings of accentedness and comprehensibility are also summarized in Table 1, with additional detail provided in Appendix E.

The SLPs' perceptual ratings of accentedness for the bilingual children (mean = 3.38) were higher than those for the monolingual children (mean = 2.2). In the independent *t*-test, Levene's test for equality of variance was not met in this case with a significance of 0.015; the results of the *t*-test showed that the difference between groups (bilingual and monolingual) was statistically significant (t = 4.287, df = 24.374, p = 0.000, two-tailed). The magnitude of differences in the means (mean difference = 1.18, 95% CI: 0.62 to 1.737) was large (d = 1.54). Cronbach's alpha for the ten SLPs' accentedness ratings was found to be highly reliable (10 items; α = .853).Cronbach's alpha for the accentedness ratings per child were found to be highly reliable (34 items; α = .938). Cronbach's alpha for the accentedness ratings per sentence were found to be reliable (3 items; α = .719).

The SLPs' perceptual ratings of comprehensibility in the bilingual children (mean = 3.38) were higher (where higher ratings indicate more difficulty understanding the utterance) than the ratings for the monolingual children (mean = 2.71). However, an independent *t*-test showed that the difference between groups (bilingual and monolingual) was not statistically significant (t = 1.202, df = 32, p = 0.238, two-tailed). Cronbach's alpha for the ten SLPs' comprehensibility ratings was found to be highly reliable (10 items; α = .951). Cronbach's alpha for the comprehensibility ratings per child were found to be highly reliable (34 items; α = .897). Cronbach's alpha for the comprehensibility ratings per sentence were found

to be reliable (3 items; α = .818).

Table 1. Means (M) and standard deviations (SD) for the *Goldman Fristoe Test of Articulation - Third edition* (GFTA-3) standard scores (SS), *Nonword Repetition* (NWR) subtest standard scores, and SLP ratings of accentedness and comprehensibility for the bilingual and monolingual children.

Group	GFTA-3 M (SD)	NWR M (SD)	Accentedness M (SD)	Comprehensibility M (SD)
Bilingual	90.29 (13.11)	7.18 (2.92)	3.38 (1.0)	3.38 (1.64)
Monolingual	98.24 (9.30)	7.06 (2.36)	2.2 (0.53)	2.71 (1.61)
t-test	*-2.04	ns	***4.29	1.20

*p < .05. ***p < .001

Note: GFTA-3 scores represent standard scores (Mean = 100, SD = 15); NWR subtest scores are scaled scores (M = 10, SD = 3). Accentedness and Comprehensibility ratings are based on a 9-pt Likert scale, with 1 indicating no foreign accent and easy to understand, and 9 indicating heavy accent and difficult to understand.

Correlations Between Assessment Measures

A correlation matrix illustrating the relationship between scores from the *GFTA-3*, *NWR* task, and SLPs' ratings of accentedness and comprehensibility is shown in Table 2 (additional details in Appendix F).

There was a significant positive correlation between the perceptual ratings of accentedness and comprehensibility for both bilingual and monolingual children (r_s = .664, N = 34, p < .0005, two-tailed).

This positive correlation became stronger when the monolingual group was removed from the analysis (see Appendix G) (r_s = .701, N = 17, p = .002, two-tailed). There was also a positive correlation between the SLPs' perceptual ratings of comprehensibility and the *ALEQ* rating of the bilingual children's L1 richness scores (r_s = .550, N = 14, p = .042, two-tailed).

There was a negative correlation between the *NWR* subtest scores and the SLPs' perceptual rating of comprehensibility in both bilingual and monolingual children (r_s = -.489, N = 34, p = .003, two-tailed). This negative correlation was still present when the monolingual group was removed from the analysis (see Appendix G) (r_s = -.582, N = 17, p = .014, two-tailed).

There were no other significant correlations between measures or with participant age. See Appendices F and G for full details.

Table 2. Correlation coefficients for the *Goldman Fristoe Test of Articulation - Third edition* (GFTA-3) standard scores (SS), *Nonword Repetition* (NWR) subtest standard scores, and SLP ratings of accentedness and comprehensibility for both the bilingual and monolingual children (N=34).

Group	GFTA-3	NWR	Accent.	Comp.	Age
GFTA-3	-	0.124	-0.103	-0.092	-0.120
NWR		-	-0.254	**-0.489	.229
Accent.			-	**0.664	-0.206
Comp.				-	-0.175

**p < 0.01 level (2-tailed).

Effect of age of onset, length of exposure, and age on test scores and SLP perceptual ratings of bilingual children. The correlation between bilingual participants' age of consistent exposure to English, length of exposure to English, age at time of testing, both standardized test scores, and SLP perceptual ratings is shown in Table 3 (see Appendix G for full details). There was a significant negative correlation between age and the accentedness ratings (r_s = -.675, N = 17, p = .003, two-tailed) and the comprehensibility ratings (r_s = -.682, N = 17, p = .002, two-tailed). There was also a significant negative correlation between the age of consistent exposure to English (in months) and the length of consistent exposure to English (in months) (r_s = -.892, N = 17, p = .000, two-tailed). There was no significant correlation between age and the GFTA-3 scores (r_s = -.318, N = 17, p = .214, two-tailed) or the NWR scores (r_s =.107, N = 17, p = .683, two-tailed).

Table 3. Correlation coefficients (two-tailed) for the *Goldman Fristoe Test of Articulation - Third edition* (GFTA-3) standard scores (SS), *Nonword Repetition* (NWR) subtest scaled scores, SLP ratings of accentedness, SLP ratings of comprehensibility, age (months), age of consistent exposure to English (months), and length of consistent exposure to English at time of testing (months) for the bilingual children only (N=17).

Group	GFTA-3	NWA	Accent.	Comp.	Age	Age Exp.	Length Exp.
GFTA-3	-	0.380	0.092	-0.204	-0.318	-0.255	0.102
NWR		-	-0.244	***-0.582**	0.107	0.032	0.057
Accent.			-	****0.701**	****-0.675**	0.264	-0.450
Comp.				-	****-0.682**	-0.114	-0.167

Age	-	0.011	0.319
Age Exp.		-	**-0.892

*p <0.05
** p < 0.01.

SLP Questionnaires

Before the SLPs completed the rating task, they filled in a short questionnaire about their comfort level assessing bilingual clients, separating accented speech from disordered speech, and rating comprehensibility in bilingual clients. See Table 4 for the range of ratings, mean, and mode for the pre- and post-rating questionnaire, and see Appendix C for more detailed information. The likert scales ranged from 1 (completely comfortable) to 8 (not comfortable). Regarding how comfortable they were in assessing bilingual clients in general, the SLP responses ranged from a 3 to a 6 (mean = 4.35; mode = 3, though responses were polarized with half responding with 3 or 3.5, and half responding with 5 or 6). The range for comfort level in separating accent from disordered speech ranged was 2 to 6 (mean = 4.3; mode = 6). Overall the SLPs were more comfortable rating comprehensibility with responses that ranged from a 2 to a 5 (mean = 3.1; mode = 2).

Following the listening task, SLPs were asked to rate how difficult or easy they found rating accentedness or comprehensibility based on the sound files they heard. The range of responses on both scales was wide (1 to 8 for accentedness, 1 to 7 for comprehensibility), but overall the SLPs found it more difficult to rate accentedness (mean = 4.9; mode = 6) than comprehensibility (mean = 2.9; mode = 2).

Table 4. Highest and Lowest Ratings, Means, and Modes of SLP's Pre- and Post-rating questionnaires.

Question	Lowest Rating	Highest Rating	Mean	Mode
Pre: Comfort assessing bilingual clients	3	6	4.35	3
Pre: Comfort separating accent from disordered speech	2	6	4.3	6
Pre: Comfort rating comprehensibility of bilingual clients	2	5	3.1	2
Post: Difficulty rating accentedness	1	8	4.9	6
Post: Difficulty rating comprehensibility	1	7	2.9	2

Note: Ratings are based on an 8-pt Likert scale, with 1 indicating completely comfortable/very easy, and 8 indicating not comfortable/very difficult.

Finally, the SLP listeners were asked how they made decisions about how to rate accentedness and comprehensibility. Their responses often included more than one factor and there was some overlap in their responses. To rate accentedness, six SLPs said they used vowels in their assessment while only one mentioned paying attention to the consonants. Four SLPs indicated that they used developmental patterns in their considerations, and two listened for articulatory precision or errors. To rate comprehensibility, the responses were more diverse, though six of the respondents indicated that they considered how well they could understand the words and two tried to assess how well another listener might understand the sentences. Four SLPs reported that they considered how much effort it took for them to understand the sentences, and four SLPs reported that they thought the fact that they already knew the sentences ahead of time might have made it easier for them to understand the child's speech.

Other factors that were taken into consideration included the child's loudness, vocal quality, rate, resonance, grammar, fluency, and speech-sound productions.

Discussion

Bilingual versus Monolingual Group Comparison

GFTA-3 and NWR standard scores and SLP perceptual ratings. The first question asked how bilingual children's scores compared to the their monolingual peers in the *GFTA-3*, *NWR*, and SLP's perceptual ratings. As expected, the bilingual children scored lower on the *GFTA-3*, but not the *NWR*, than their monolingual peers, and the monolingual children rated lower on accentedness than the bilingual children.

As hypothesized, the bilingual children's performance on the *GFTA-3* was lower than the monolingual children's and a *t*-test comparison of the means found the difference statistically significant with a moderately large effect size (d = -0.71). Interestingly, while the Hack et. al. study also found that the bilingual children scored lower than their monolingual peers, their results did not find the difference to be statistically significant (p=.19). There were some fundamental differences between the participants in the Hack et. al. study and this present study. For example, the sample size of the Hack et. al. study was roughly double the number of participants in this present study and their participants all spoke either Cantonese or Mandarin, which allowed them to easily do an analysis of the structural differences between English and Cantonese/Mandarin. Furthermore, the average age of their participants was 7;4, while the average age of the participants in this study was 4;9. Additionally, the average age of consistent exposure to English in the Hack. et. al. group was 5;2 while the average for the participants in this study was just over 1;8, though there was a large range in the length of exposure (e.g., from consistent exposure to English since birth to just three months of consistent exposure to English). Given the large differences between the Hack et. al. participant group and the children in this study, it is hard to

pinpoint why Hack et. al. did not find significance and this study did. One possibility is that the difference of age may have impacted the test results as the bilingual children in Hack et. al.'s study may have had more exposure to academic English (e.g., longer exposure through school) and the monolingual children may have had fewer developmental errors that could have been confused for accentedness. This seems to suggest that test measures may result in less of a difference between bilingual and monolingual populations as they age and gain more English exposure.

Though our sample size was smaller than the Hack et. al. study, one consistency between both studies was that many of the bilingual children were born in Canada. Despite being born in Canada, these children still scored lower than the monolingual children, which may indicate that the predominance of the L1 in their speaking environment does have an impact on their speech sound production. The implication, therefore, is that articulation tests normed on a monolingual English population are not appropriate for bilingual children, even those who were born in Canada. The *GFTA-3* does not take vowels into consideration, though this is what many of the SLPs reported listening for when they were making the decision regarding accentedness (60%). This many also explain the absence of a significant correlation between the *GFTA-3* scores and the accentedness ratings.

As hypothesized, the bilingual children's scores on the *NWR* Test were not significantly different than the monolingual children's scores. However, several children in the monolingual group had surprisingly low scores on the NWR task, even though their *GFTA-3* scores were within normal limits. Initially, there was concern that a difference in testing environments may have affected the results, as there were six children who were tested in a home with more background noise. However, when these children were removed from the analysis, there was no change in significance. These results support the Sorenson, Duncan, and Paradis (2016) findings that suggest that the *NWR* may be a fairer assessment for bilingual children and suggests that it would be a better choice for diagnostic purposes than the *GFTA-3*.

SLP ratings of accentedness and comprehensibility. The SLPs' mean perceptual ratings of accentedness in the bilingual children were significantly higher than the ratings in monolingual children. Therefore, SLPs were able to identify the presence of a foreign accent in the bilingual group. The presence of a perceived foreign accent in young children has not been studied widely. However, Flege, Birdsong, Bialystok, Mack, Sung, & Tsukada (2006) found that some 9-17 yr old Korean children who had been exposed to English for 2-4 years were perceived by native listeners to have an accent, but overall their accents were not as strong as Koreans who learned English as adults. Hack et al. (2012) also found that listeners were able to detect foreign accents in 5-9 yr old bilingual Cantonese-English and Mandarin-English children. Using the same 9 point Likert scale, their mean accentedness ratings were slightly higher (M = 3.8) than those in the present study (M = 3.38), suggesting that accentedness is more prominent in older children.

The SLPs mean perceptual ratings for comprehensibility were also higher in the bilingual group (where higher ratings indicate more difficulty understanding the utterance), but the difference was not significant. However, there was also a significant positive correlation between the overall accentedness ratings and comprehensibility ratings (see Figure 1) indicating that when the child was thought to have an accent, they were also more difficult to understand. This suggests that the presence of an accent may increase the effort on the part of the listener to comprehend speech, even in young children. In contrast, Derwing and Munro (1997) found a relatively weak correlation between accentendess and comprehensibility ratings of non-native adult speakers. They found that accentedness ratings were distributed across the scale, with the majority falling between 6 and 8, whereas comprehensibility ratings were also spread across the full scale but relatively less harsh, with most receiving a rating of 5 or less. In comparison, the accentedness and comprehensibility ratings for the bilingual children in the present

study were much lower, with the majority of accentedness ratings < 6 and the majority of comprehensibility ratings <4 (see histograms in Appendix H).

That the SLPs may be confounding accent and comprehensibility supports findings of the Hack et. al. study, which had their SLPs rate their perception of the child's developmental level if they did not feel their speech was accented. They reported that when the SLPs were asked after completing the rating task, they often referred to using "intelligibility" to make their decision about the child's developmental level. The researchers reported that using intelligibility during perceptual ratings was in line with previous research, but they felt that it is a potentially problematic way to approach rating developmental level because intelligibility can also be used to rate accentedness.

Figure 1. Scatterplot showing the positive correlation between SLP's accentedness and comprehensibility ratings in the bilingual group.

Correlations Between Assessment Measures

The second question this study addressed was whether SLP perceptual ratings of accentedness and comprehensibility correlated with scores from standardized tests of *nonword repetition* (*NWR*) and articulation (*GFTA-3*). The results show that the SLP perceptual ratings of accentedness correlated with neither the *NWR* or *GFTA-3.* On the other hand, there was a negative correlation between the SLP perceptual ratings of comprehensibility and the *NWR*, but not the *GFTA-3*. This is partially consistent with our hypothesis leading into the study that the SLPs' perceptual ratings would correlate with the NWR test based on the evidence that the *NWR* task is a more appropriate way to capture a bilingual child's language capacity. However, only the perceptual ratings of comprehensibility correlate; there was no correlation between the perceptual ratings of accentedness and the *NWR* task.

There was a negative correlation between the SLPs' perceptual rating of comprehensibility and the *NWR* standard score, which indicates that when the *NWR* scores were higher, the SLP's comprehensibility ratings were lower and therefore the child was easier to understand. These results suggest that when the child is better able to match the phonological production of nonsense words, they are generally easier to understand as a whole. On the other hand, there was not a significant correlation in either direction between accentedness and the *NWR* task. This is consistent with the fact that the bilingual children's scores were not statistically different from those of the monolingual group. Therefore, despite the possible presence of an accent, the child's *NWR* task score did not appear to be affected - further supporting evidence for the assertion that the *NWR* is a fairer task for assessing bilingual children.

Effect of age of onset, length of exposure, and age on test scores and SLP perceptual ratings of bilingual children. There was a negative correlation between the child's age and both the SLP accentedness ratings and SLP comprehensibility ratings, but there was no correlation between the child's

age and their *GFTA-3* or *NWR* scores. This is to be expected, as the SLP's accentedness and comprehensibility ratings are perceptual measures whereas the *GFTA-3* and *NWR* scores are standardized scores based on age. Perceptual ratings, on the other hand, are subject to factors such as developmental errors. There was also a significant negative correlation between the child's age of consistent exposure to English and the length of exposure. This means that the older the child was when they were when they started receiving consistent exposure to English, the less consistent exposure to English they had.

There was not, however, a significant correlation between the SLPs' accentedness ratings or comprehensibility ratings and the age of consistent exposure to English or the length of consistent exposure to English. This indicates that the SLPs did not necessarily rate the children who were older when they began getting more consistent exposure to English, or had less exposure to English (measured in months), differently than those who were younger when they started getting consistent exposure to English or had been getting consistent exposure to English for a longer time.

SLP Questionnaires

Though not one of the questions this study initially set out to answer, the SLP questionnaires garnered some interesting information regarding confidence in assessing bilingual clients and how they performed the rating task. Though the mean response was 4.35, falling at about the halfway point on the scale, this was an average of a polarized response. Looking at the average number of bilingual clients that the SLPs have on their caseloads, we see that those who report a higher number of clients feel more comfortable with assessments than those who report a lower number of clients, which is to be expected.

On the other hand, what we do not necessarily see is more confidence from those SLPs who report higher caseloads of bilingual clients in being able to separate accented from disordered speech, with some of the SLPs reporting a caseload of 30% bilingual clients rating their ability to separate

accented from disordered speech at a 6 on the 8-point scale (where 8 indicates “not comfortable”). It is possible that the variation in confidence level may have to do with the length of practice, though the questionnaire did not ask for SLPs to report how long they had been practicing.

There is much less variability in the SLPs reported comfort level rating the comprehensibility of bilingual clients, with the mean and mode being only one point off from each other. This indicates that SLPs overall feel much better assessing the comprehensibility of bilingual clients; however, there is some suggestion that those comprehensibility ratings are impacted by the presence of an accent. This is important because accent is not disorder. Therefore, if SLPs cannot separate accented speech from comprehensibility ratings, these comprehensibility ratings are not necessarily an effective tool to identify the presence or absence of a speech disorder.

Overall, the SLPs found the task of rating accentedness difficult with the mean on the post-rating questionnaire of 4.9 and the mode of 6. This is consistent with the lack of general confidence that they reported regarding separating accented speech from disordered speech. The SLPs found the task of rating comprehensibility much easier, with the mean response of 2.9 and the mode of 2. This is again consistent with their overall confidence rating comprehensibility, though the correlational analysis does raise some question about whether accent is playing a role in these comprehensibility ratings.

Limitations

This study did not specify a preferred testing environment and the testing environments ranged from ideal testing conditions in a quiet room at Corbett Hall to a busy home or daycare environment. The majority of the bilingual children were tested in quiet rooms at Corbett Hall or in their daycare, with a few being tested at home. However, the majority of the monolingual children were tested at home. Though the home environments were often quiet enough for testing purposes, there were a few exceptions. There was some question that the results of the *NWR* task may have been impacted due to

background noise masking the stimulus presentation. When these participants were removed from the statistical analysis,the mean NWR score increased, but there was not a statistically significant difference between the monolingual and bilingual groups.

Another limitation to generalizability of the results was the relatively small sample size and the heterogeneity of the participant's L1. The Hack et. al. model study had roughly double the number of participants and they all spoke either Mandarin or Cantonese as their L1. Though the heterogeneity of L1s in the present study may be more reflective of an SLP's real-world caseload, it makes it difficult to pinpoint the features that may be affecting perceptions of accentedness. Despite these factors, the findings are still in line with the research previously done in this area. A bigger sample size would be useful to offer more confidence in the results, but the aggregate of research shows a consistent picture regarding how accented speech affects perceptual ratings of accentedness and comprehensibility. Similarly, having more SLP raters and asking more extensive questions about their current practice assessing bilingual children would be a good direction for future research. There is obviously some variability in how SLPs are approaching bilingual speech assessment. As bilingual children are often a significant portion of an SLP's caseload, understanding how SLPs are currently assessing this population would be useful.

Being able to compare SLP ratings to phonetic transcriptions or acoustic analyses would also have given the present study a better sense of what features may contribute to the perception of accented speech and developmental speech errors. The perceptual ratings and the pre- and post-task questionnaires highlighted some of the variability in SLPs' confidence levels and the features they are paying attention to, but more follow-up questions would have been useful to help target how the SLPs approached the task, and whether it mirrors what they do in a real-world setting. A direct comparison of

the phonetic and acoustic features associated with the perception of accentedness would be useful to shed light on the nature and degree of overlap with developmental errors.

Clinical Implications

The present study by-and-large contributes to the growing body of evidence of previous studies. The Hack et. al. study found that bilingual children scored lower than monolingual children on the *GFTA-2*, though they did not find significance in the difference between group means as the present study did. Hack et. al. also found that the SLP perceptual ratings did not correlate with the *GFTA* scores, which is consistent with our results. In alignment with the Sorenson Duncan and Paradis (2016) study examining *NWR* scores, we found that the bilingual children's *NWR* scores were not statistically different than those of their monolingual peers. These supportive findings are promising for clinical use because it suggests that SLPs may be able to use a *NWR* task with more confidence to assist with the diagnosis of speech sound disorders in bilingual children. That being said, however, the nature of the scoring in the *CTOPP-2*'s *NWR* task may be troublesome as it does not reflect the number or nature of any errors that a child makes in the production of the speech sounds when it uses an all-or-nothing rating system. For example, though the bilingual and monolingual children did not score statistically different on the NWR task, the test scoring does not break down how many speech-sound errors per nonword that the bilingual or monolingual children are making. This is particularly problematic because though the monolingual children appear to be getting similar raw scores to their bilingual peers, they may be make fewer error overall than their bilingual peers, but this is not reflected in the scoring.

An additional finding that supports previous research was that SLPs may not be separating accentedness from their overall comprehensibility ratings. Though it seems to be a fine line, it is worth mentioning the importance of such a distinction because the scope of practice of an SLP who is concerned with a child's possible delayed or disordered speech is not concerned with accent

remediation. Given the already negative attitude towards accent that was discussed at the beginning of this study, it is not appropriate for SLPs to pathologize accented speech. SLPs need to be able to separate accent from delayed or disordered speech when assessing bilingual children. There is a risk that if SLPs are finding accented speech more difficult to understand and are not able to confidently differentiate between accentedness and speech errors, children will continue to erroneously end up on their caseloads. Conversely, if an SLP over-attributes differences in speech-errors to accent, there exists the possibility that a child with and accent *and* a speech-sound delay or disorder may not be identified properly. The possibility of both false negatives and false positives highlight the importance for SLPs to be aware of tools available to help them understand the speech-sound profile for the child's L1, and which will help them sort through where accented speech begins and delayed or disordered speech may begin.

Lastly, the variability in SLPs' confidence levels assessing bilingual children, as well as the features that they use to analyze accentedness suggests that better training is required for SLPs. These results support the findings of Young and Westernoff (1999), namely that SLPs feel inadequately prepared to work with bilingual clients. The variability in how SLPs approached rating the accentedness in speech seems to suggest an overall lack of training. Several SLPs noted that vowels influenced their perception of foreign accent. This makes sense given that vowels are typically mastered by children by 3 years of age, and therefore the presence of vowel differences in a child's speech would be unexpected. Given that the bilingual participants' *GFTA-3* scores were lower than their monolingual peers, despite the fact that the test does not consider the vowels in scoring, suggests that there are enough consonantal differences between the bilingual and monolingual children's speech to have caused a difference in the scores. However, only one SLP reported listening for consonantal differences in the questionnaire. Training courses for SLPs may be a solution worth considering and there is precedent for accent training

to make a difference in both the confidence and the accuracy of a native speaker's ability to discern accented speech.

Derwing, Rossiter, and Munro (2002) conducted a study wherein native English speakers were split into three groups - a control group, a group who received cultural instruction alone, and a group who received specific accent training for features of the Vietnamese accent. In a pre- and post-training questionnaire and assessment task, each participant was asked to rate their confidence listening to accented speech and were asked to transcribe a segment spoken by Vietnamese L1 speakers with a discernible accent. All groups were more accurate in the post-training task, which means that the specific accent-training might not have given them an advantage in accuracy but results do show that those who received accent training had greater confidence in speaking to those with an accent. Since the SLPs in this study had varying degrees of confidence discerning accented speech, a training program for SLPs targeting accented speech may be of benefit in at least increasing their confidence of assessing bilingual clients. However, specific training in how to separate accented speech from disordered speech would be more useful to SLPs since it would help them tease apart what exactly they are hearing during their assessments.

Further Research

Though this study has generally supported previous research in the area of speech assessments for bilingual children, more research needs to be done to more fully understand how accented speech affects bilingual speech assessment and how SLPs could be better prepared to serve bilingual clients on their caseload. Specifically, the variation in what the SLPs were paying attention to when rating accented speech suggests that a features analysis and comparison of the bilingual and monolingual children's productions might be useful in highlighting practical things SLPs can pay attention to when doing assessments with bilingual children. The Hack et. al. (2012) study did examine the differences

between the monolingual and bilingual children's speech, but their participants were either Cantonese or Mandarin speakers. The participants in this study were of varying language backgrounds, so it would be beneficial to do an overarching analysis to see if there are any factors that are common to all children with accented speech. While it is important for SLPs to understand the differences between the child's specific L1 sound profile and English, having a general guide for differences (if any) that all bilingual speakers share would be a good starting place for an assessment.

Another addition to the current study that may be worthwhile pursuing would be to add monolingual children with speech delays or disorders to the mix to test whether SLPs can tell the difference between accented speech and disordered speech. The SLPs in this present study knew that the parents did not have concerns about the child's speech or language development. Knowing that some of the children have a speech delay or disorder may change how the SLPs approach the rating task for both accent and comprehensibility. This would also be a chance to collect more qualitative data about what features SLPs are using to make decisions about the distinction between accent and a delay or disorder. This present study only looks at what features they are aware of using to make decision regarding accent, so it would be worthwhile to see if there is crossover with the features they cite and those that are confirmed in transcription or acoustic analyses.

Lastly, in addition to delving further into how the actual productions of bilingual and monolingual children differ, regardless of the presence or absence of a speech or language delay or disorder, having SLPs complete a similar task with a longer, unscripted conversation sample to see how that might change their perceptual ratings of accent and comprehensibility would be a better match for what SLPs actually encounter in their day-to-day practice. In addition to having longer speech samples to work from, conversation samples may elicit a more accurate production from the child as the nature of a sentence repetition task might mean that some of the children's sentences are not as accented as the

child's general speech might be. Many of the SLPs commented that they found the task difficult with such a short clip and wanted to hear more. Furthermore, the comprehensibility-rating task would be potentially very different as the SLPs would not know what the child was saying before making their rating. If this is the case, the relationship between the accent ratings and comprehensibility ratings may increase.

Conclusion

Whether conscious or not, accentedness as an effect on the judgements people make about each other. This is well documented in the adult world, especially in terms of perceived competence in the corporate or academic world. There is less research about how accent may affect children in their academic setting - and specifically about how the presence of an accent may affect SLPs assessments on children. Most standardized speech-sound assessment instrument are not normed on bilingual children, making it so SLPs cannot use standardized scores to make diagnostic decisions the presence or absence of a speech delay or disorder.

Overall, the findings of this study support the growing body of evidence examining the difference between bilingual children and monolingual children's test scores on the *GFTA-3* and the lack of any significant difference on *NWR* tasks. This suggests that *NWR* tasks may be a fairer assessment to use with bilingual populations as they are able to score similarly to their monolingual peers, though a more detailed look at the scoring of *NWR* tasks is warranted due to the lack of specificity test provides regarding both the quantity and quality of the speech-sound errors produced by the child. Furthermore, the findings of this study highlighted how SLPs understand the nature of accented speech. Despite that over half of the SLPs reported using vowels to make decisions about the presence of an accent, the *GFTA-3* Sounds-in-Words subtest does not take vowels into consideration; this suggests that there are

factors contributing to accented speech of which SLPs are either not aware or are not consistently taking into consideration.

Future studies may consider more directly examining the difference between and SLP's perceptual judgement of accent and comprehensibility with conversational speech samples where they do not already know what the child is saying or with a participant group that contains children with speech delays or disorders, which may give a more accurate picture of how well SLPs can distinguish accented speech from delayed or disordered speech. Additionally, there is an important need for training programs designed specifically for SLPs with the goal of increasing both confidence and competence when approaching bilingual children's speech-sound assessments.

References:

Akomolafe, Soji. MSC, P. H. D. (2013). The invisible minority: Revisiting the debate on foreign-accented speakers and upward mobility in the workplace. *Journal of cultural diversity*, *20*(1), 7.

American Speech-Language-Hearing Association Spoken Language Disorders: Assessment. (n.d.). Spoken Language Disorders: Assessment. Retrieved June 27, 2017, from http://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935327§ion=Assessment

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics,39*(1), 1-17. doi:10.1016/j.wocn.2010.10.006

Beiser, M., & Hou, F. (2016). Mental Health Effects of Premigration Trauma and Postmigration Discrimination on Refugee Youth in Canada. *The Journal of Nervous and Mental Disease,204*(6), 464-470. doi:10.1097/nmd.0000000000000516

Bird, E. K. R., & Deacon, H. (2012). Survey of Canadian Speech-Language Pathology Service Delivery to Linguistically Diverse Clients. *Canadian Journal of Speech-Language Pathology & Audiology*, *36*(1).

Brandeker, M., & Thordardottir, E. (2015). Language Exposure in Bilingual Toddlers: Performance on Nonword Repetition and Lexical Tasks. *American Journal of Speech-Language Pathology,* 24(2), 126-138. doi:10.1044/2015_AJSLP-13-0106

Chiat, S., & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1179-1189

Creese, G. (2010). Erasing English Language Competency: African Migrants in Vancouver, Canada. *Journal of International Migration and Integration / Revue de l'integration et de la migration internationale, 11*(3), 295-313. doi:10.1007/s12134-010-0139-3

D'Souza, C., Kay-Raining Bird, E., & Deacon, H. (2012). Survey of Canadian speech-language pathology services to linguistically diverse clients. Canadian Journal of Speech-language Pathology and Audiology, 36, 18 – 39.

Derwing, T. M., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 Accent and Ethics: Issues that Merit Attention. *Englishes in Multilingual Contexts Multilingual Education,* 63-80. doi:10.1007/978-94-017-8869-4_5

Derwing, T.M. & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19* (1), 1–16.

Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching Native Speakers to Listen to Foreign-Accented Speech. *Journal Of Multilingual And Multicultural Development*, 23(4), 245-59. doi:10.1080/01434630208666468

Derwing, T. M., & Murray, J. (2014). Accent and intelligibility: cracking the conundrum. *SPEAK OUT!*, *50*, 12-17.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the Window on Comprehensible Pronunciation After 19 Years: A Workplace Training Study. *Language Learning, 64*(3), 526-548. doi:10.1111/lang.12053

Ebert, K. D., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language Teaching*, *49*(03), 301-338.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America,84*(1), 70-79. doi:10.1121/1.396876

Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, G., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics, 34*, 153-175.

George, U., & Chaze, F. (2014). Discrimination at Work: Comparing the Experiences of Foreign-trained and Locally-trained Engineers in Canada. *Canadian Ethnic Studies, 46*(1), 1-21. doi:10.1353/ces.2014.0002

Goldman, R. & Fristoe, M. (2015) *Goldman-Fristoe Test of Articulation – Third Edition (GFTA-3)*. Bloomington, MN: NCS Pearson, Inc..

Government of Canada, Immigration, Refugees and Citizenship Canada, Communications Branch. (2016, October 31). 2016 Annual Report to Parliament on Immigration. Retrieved May 04, 2017, from http://www.cic.gc.ca/english/resources/publications/annual-report-2016/index.asp

Hack, J., Marinova-Todd, S. H., & May Bernhardt, B. (2012). Speech assessment of Chinese–English bilingual children: Accent versus developmental level. *International journal of speech-language pathology*, *14*(6), 509-519.

Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Language & Communication Disorders, 48*(1), 1-24. doi:10.1111/j.1460-6984.2012.00178.x

Hopp, H., & Schmid, M. S. (2011). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics, 34*(02), 361-394. doi:10.1017/s0142716411000737

Houshmand, S., Spanierman, L. B., & Tafarodi, R. W. (2014). Excluded and avoided: Racial microaggressions targeting Asian international students in Canada. *Cultural Diversity and Ethnic Minority Psychology, 20*(3), 377-388. doi:10.1037/a0035404

Jesney, K. (2004). The use of global foreign accent rating in studies of L2 acquisition. A report prepared for the Language Research Centre, University of Calgary.

Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, *66*(S2), 155-186.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly, 39*(3), 369. doi:10.2307/3588485

Martin, D., Krishnamurthy, R., Bhardwaj, M., & Charles, R. (2003). Language change in young Panjabi/English children: implications for bilingual language assessment. *Child Language Teaching & Therapy*, *19*(3), 245-265. doi:10.1191/0265659003ct254oa

McLeod, S., Harrison, L. J., & McCormack, J. (2012). Intelligibility in Context Scale. Bathurst, NSW, Australia: Charles Sturt University. Retrieved from http://www.csu.edu.au/research/multilingual-speech/ics

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics, 28*(7-8), 508-531. doi:10.3109/02699206.2014.926994

McLeod, S., & Verdon, S. (2017). Tutorial: Speech Assessment for Multilingual Children Who Do Not Speak the Same Language(s) as the Speech-Language Pathologist. *American Journal of Speech-Language Pathology, 26*(3), 691. doi:10.1044/2017_ajslp-15-0161

Munro, M. J. (1993). Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings. *Language and Speech,36*(1), 39-66. doi:10.1177/002383099303600103

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *49*(s1), 285-310.

Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied Psycholinguistics* 31.2, 227−252.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic approaches to bilingualism*, *1*(3), 213-237.

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*, 474-497.

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual Language Development and Disorders. A handbook on bilingualism and second language learning (2nd edition).*Baltimore, MD: Paul H. Brookes.

Paradis, J., Nicoladis, E., & Genesee, F. (2000). Early emergence of structural constraints on code-mixing: Evidence from French–English bilingual children. *Bilingualism: language and cognition*, *3*(03), 245-261.

Paradis, J., Schneider, P., & Sorenson Duncan, T. (2013). Discriminating children with language impairment among English language learners from diverse first-language backgrounds. *Journal of Speech Language and Hearing Research, 56*(3), 971. doi:10.1044/1092-4388(2012/12-0050)

Psychology Software Tools, Inc. (2012). [E-Prime 2.0]. Retrieved from http://www.pstnet.com.

R. K. Wagner, J. K. Torgesen, C. A. Rashotte, & N. A. Pearson. (2013). *Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2)*. Austin, TX: PRO-ED.

Sereno, J. A., Lammers, L., & Jongman, A. (2010). Segmental versus suprasegmental information in the perception of foreign-accented speech. *PsycEXTRA Dataset*. doi:10.1037/e520592012-255

Sorenson Duncan, T., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech\ Language Hearing Research*, *59*(1), 39-48.

Statistics Canada Linguistic Characteristics of Canadians. (2015, December 22). Linguistic Characteristics of Canadians.Retrieved May 04, 2017, from http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.cfm

Tomic, P. (2013). The colour of language: Accent, devaluation and resistance in Latin American immigrant lives in Canada. *Canadian Ethnic Studies* 45(1), 1-21.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(04), 905-916.

Verdon, S., McLeod, S., & Wong, S. (2015). Reconceptualizing practice with multilingual children with speech sound disorders: people, practicalities and policy. *International Journal of Language & Communication Disorders*, *50*(1), 48-62.

Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal Of Speech-Language Pathology*, *14*(3), 292-305. doi:10.3109/17549507.2011.636071

Young, T., & Westernoff F.(1999). Reflections of speech-language pathologists and audiologists on practices in a multicultural, multilingual society. *Journal of Speech Language Pathology and Audiology*, *23*, 24-30.

University of Arizona, Department of Linguistics. (2010, May 26). *Teachers' English Fluency Initiative in Arizona* [Press release]. Retrieved May 13, 2017, from http://www.u.arizona.edu/~hammond/ling_statement_final.pdf

Appendix A

Bilingual Children Participants													
Participant Number	**Age (months)**	**Age of Consistent Exposure to English (months)**	**L1**	**ALEQ: English Richness Score**	**ALEQ: L1 Richness Score**	**ALEQ: Language Use Home**	**ICS**	**GFTA-3 Standard Score**	**NWR Scaled Score**	**SLP Accent Rating**	**SLP Comp Rating**	**Hearing Screening Completed**	**Test Location**
B1	72	46	Spanish	0.56	0.3	0.44	5	68	11	1.4	1.17	No	Home
B2	63	36	Portuguese	Not Given	Not Given	0.56	4.57	71	6	4.23	3.06	Yes	Corbett Hall
B3	48	21	Portuguese	Not Given	Not Given	0.59	4.43	92	6	4.63	5.03	Yes	Corbett Hall
B4	48	21	Portuguese	Not Given	Not Given	0.59	4.43	91	6	4.1	4.8	Yes	Corbett Hall
B5	78	0	Portuguese	0.81	0.3	0.56	4.29	102	6	1.63	1.27	Yes	Corbett Hall
B6	49	46	Swedish	0.6875	0.3	0.25	4.3	69	2	4.37	5.97	Yes	Corbett Hall
B7	50	48	Persian	0.6875	0.15	0.5	4.4	113	12	4.03	2.17	No	Home
B8	50	18	Russian	0.5	0.45	0	4.28	101	8	4.4	3.67	No	Daycare
B9	57	13	Hindi	0.75	0.25	0.625	4	97	11	2.7	1.37	No	Daycare
B10	51	0	French	0.5625	0.65	0.625	3.85	86	6	3.4	5.87	No	Home
B11	49	12	Portuguese	0.5	0.5	0.17	3.88	85	3	3.1	5.2	Yes	Corbett Hall
B12	71	18	Swedish/ Romanian	0.6875	0.1	0.71	4.7	78	9	1.93	1.73	No	Daycare
B13	67	12	Chinese	0	0.5	0.42	5	87	6	2.93	2.27	No	Daycare
B14	48	0	Hindi/ Marathi	0.8125	0.65	0.63	4.5	105	9	3.9	4	No	Daycare
B15	50	36	Nepalese	0.6875	0.25	0.25	4.71	95	3	3	2.4	yes	Daycare
B16	48	9	Hungarian	0.5625	0.5	0.06	4.85	92	9	4.07	4.63	No	Daycare
B17	48	18	French/ Arabic	0.625	0.15	0.625	4	103	9	3.57	2.8	Yes	Corbett Hall
Mean	**55.71**	**20.82**		**0.60**	**0.36**	**0.45**	**4.42**	**90.29**	**7.18**	**3.38**	**3.38**		
Standard Deviation	**10.4**	**16.10**		**0.20**	**0.18**	**0.22**	**0.36**	**13.11**	**2.92**	**1.0**	**1.64**		
Minimum	**48**	**0**		**0**	**0.1**	**0**	**3.85**	**68**	**2**	**1.4**	**1.17**		
Maximum	**78**	**48**		**0.8125**	**0.65**	**.71**	**5**	**113**	**12**	**4.63**	**5.97**		

Appendix B

Monolingual Children Participants								
Participant Number	**Age**	**ICS Score**	**GFTA-3 Standard Score**	**NWR Scaled Score**	**SLP Accent Rating**	**SLP Comp Rating**	**Hearing Screening Completed**	**Test location**
C1	70	5	109	10	1.43	2	Yes	Corbett Hall
C2	70	5	99	9	1.87	2.03	Yes	Corbett Hall
C3	67	4.57	92	9	2.23	1.37	No	Corbett Hall
C4	67	4.85	92	11	2.3	1.7	No	Home
C6	67	4.42	76	10	1.77	1.4	No	Home
C7	54	4.14	95	7	1.87	2.47	No	Home
C8	52	4	95	8	2.73	3.83	No	Home
C9	77	5	98	6	2.4	1.83	No	Home
C10	70	4.57	90	3	2.6	3.13	No	Home
C11	50	3.57	95	5	1.3	1.33	No	Home
C13	52	4.71	108	5	1.27	1.57	No	Home
C14	47	3.43	96	6	2.73	6	No	Home
C15	55	4	96	4	2.2	3.27	No	Home
C16	51	4.71	111	4	2.6	6.9	No	Home
C17	54	4.42	115	8	3	1.43	Yes	Corbett Hall
C18	51	4.3	98	8	2.37	3.3	No	Daycare
C19	50	4.71	105	7	2.7	2.43	No	Corbett Hall
Mean	**59.06**	**4.44**	**98.24**	**7.06**	**2.2**	**2.71**		
Standard Deviation	**9.61**	**0.48**	**9.3**	**2.36**	**0.53**	**1.61**		
Minimum	**47**	**3.43**	**76**	**3**	**1.3**	**1.33**		
Maximum	**77**	**5**	**115**	**10**	**3**	**6.9**		

Appendix C

SLP Raters										
SLP Participant Number:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
On average, how many bilingual clients do you usually have on your caseload?	30%	30%	30%	30%	10	20%	5	2	1-2	60%
How comfortable would you rate yourself in assessing bilingual clients? 1 (completely comfortable) - 8 (not comfortable)	3	3.5	3	3	6	3	5	6	6	5
How comfortable would you rate yourself in separating accent from disordered speech in bilingual clients? 1 (completely comfortable) - 8 (not comfortable)	3	6	6	2	5	2	4	6	6	3
How comfortable would you rate yourself in rating comprehensibility of bilingual clients? 1 (completely comfortable) - 8 (not comfortable)	2	4	2	3	4	2	2	5	5	2
How easy / difficult did you find it to rate accentedness based on the sound files you heard? 1 (very easy) - 8 (very difficult)	1	4	2	6	7	6	6	8	8	5
How easy / difficult did you find it to rate comprehensibility based on the sound files you heard? 1 (very easy) - 8 (very difficult)	1	2	2	7	3	3	2	1	2	2

Appendix D

Comparison of means between bilingual and monolingual children's GFTA-3 and NWR test scores:

Bilingual vs. Monolingual Standard Scores on GFTA-3 and NWR Tests					
Test	**Condition**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
GFTA-3	Bilingual	17	90.29	13.11	3.18
	Monolingual	17	98.24	9.30	2.26
NWR	Bilingual	17	7.18	2.91	.71
	Monolingual	17	7.06	2.36	.57

Bilingual vs. Monolingual Standard Scores on GFTA-3 and NWR Tests										
		Levene's Test for Equality of Variances					**T-test for the Equality of Means**		**95% CI of the difference**	
Test		**F**	**Sig.**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
GFTA-3	Equal variances assumed	2.288	.140	-2.037	32	**.050**	-7.941	3.898	-15.881	-.001
NWR	Equal variances assumed	1.010	.322	.129	32	**.898**	.118	.910	-1.737	1.972

Appendix E

Comparison of means between bilingual and monolingual SLP perceptual ratings of accentedness and comprehensibility:

Bilingual vs. Monolingual SLP Perceptual Ratings of Accentedness and Comprehensibility					
Perceptual Rating	**Condition**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
SLP Perceptual Ratings: Accent	Bilingual	17	3.3759	1.000007	.24255
	Monolingual	17	2.1982	.53163	.12894
SLP Perceptual Ratings: Comp.	Bilingual	17	3.3771	1.64287	.39846
	Monolingual	17	2.7053	1.61484	.39166

Bilingual vs. Monolingual SLP Perceptual Ratings of Accentedness and Comprehensibility										
		Levene's Test for Equality of Variances					***t*-test for the Equality of Means**		**95% CI of the difference**	
Perceptual Rating		**F**	**Sig.**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
SLP Perceptual Rating: Accent	Equal variances assumed	6.673	**0.015**	4.287	32	**.000**	1.17765	.27469	.61811	1.73718
	Equal variances not assumed			4.287	24.374	.000	1.17765	.27469	.61117	1.74413
SLP Perceptual Rating: Comp.	Equal variances assumed	.560	.460	1.202	32	**.238**	.67176	.558	-.4663	1.80983

Appendix F

Correlations Between Assessment Measures

Spearman's Rho		GFTA-3 Standard Score	NWR Standard Score	SLP Accent Rating	SLP Comp. Ratings	ALEQ: English Richness Score	ALEQ: L1 Richness Score	Intelligibility in Context Scale
GFTA-3 Standard Score	Correlation Coefficient		.124	-.103	-.092	.432	-.111	.014
	Sig. (2-tailed)		.484	.560	.604	.123	.936	.936
	N		34	34	34	14	34	14
NWR Standard Score	Correlation Coefficient			-.254	**-.489****	.190	-.347	.254
	Sig. (2-tailed)			.148	**.003**	.514	.225	.147
	N			34	34	14	14	34
SLP Accent Rating	Correlation Coefficient				**.664****	-.114	.247	-.201
	Sig. (2-tailed)				**.000**	.699	.395	.255
	N				34	14	14	34
SLP Comp. Ratings	Correlation Coefficient					-.233	**.550***	-.280
	Sig. (2-tailed)					.444	**.042**	.108
	N					14	14	34
ALEQ: English Richness Score	Correlation Coefficient						-.315	-.083
	Sig. (2-tailed)						.273	.779
	N						14	14
ALEQ: L1 Richness Score	Correlation Coefficient							-.069
	Sig. (2-tailed)							.814
	N							14

Intelligibility in Context Scale	Correlation Coefficient							
	Sig. (2-tailed)							
	N							

* Correlation is significant at the 0.01 level (2-tailed).

** Correlation is significant at the 0.05 level (2-tailed).

Appendix G

Correlation Examining the Effect of Age on Testing Scores and SLP Perceptual Ratings of Bilingual Children

Spearman's Rho		GFTA-3 Standard Score	NWR Standard Score	SLP Accent Rating	SLP Comp. Ratings	Age	Age of Exposure to English	Length of Exposure to English
GFTA-3 Standard Score	Correlation Coefficient		.380	.092	-.204	-.318	-.255	.102
	Sig. (2-tailed)		.132	.726	.433	.214	.323	.697
	N		17	17	17	17	17	17
NWR Standard Score	Correlation Coefficient			-.244	**-.582***	.107	.032	.057
	Sig. (2-tailed)			.346	**.014**	.683	.902	.828
	N			17	17	17	17	17
SLP Accent Rating	Correlation Coefficient				**.701****	**-.675****	.264	-.450
	Sig. (2-tailed)				**.002**	**.003**	.305	.070
	N				17	17	17	17
SLP Comp. Ratings	Correlation Coefficient					**-.682****	-.114	-.167
	Sig. (2-tailed)					**.003**	.664	.522
	N					17	17	17
Age	Correlation Coefficient						.011	.319
	Sig. (2-tailed)						.966	.211
	N						17	17
Age of English Exposure	Correlation Coefficient							**-.892****
	Sig. (2-tailed)							**.000**

	N							17
Length of English Exposure	Correlation Coefficient							
	Sig. (2-tailed)							
	N							

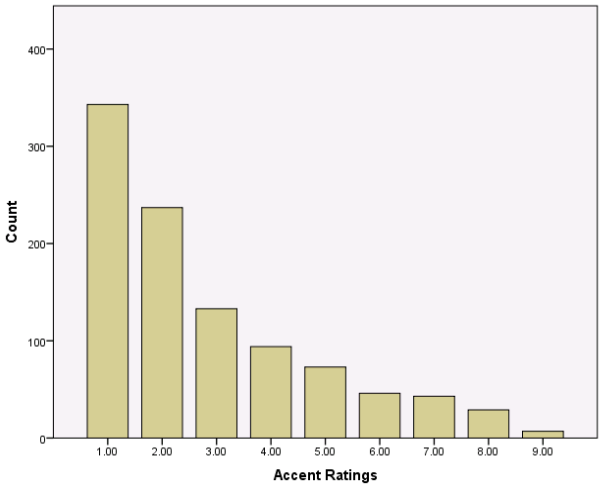
* Correlation is significant at the 0.01 level (2-tailed).

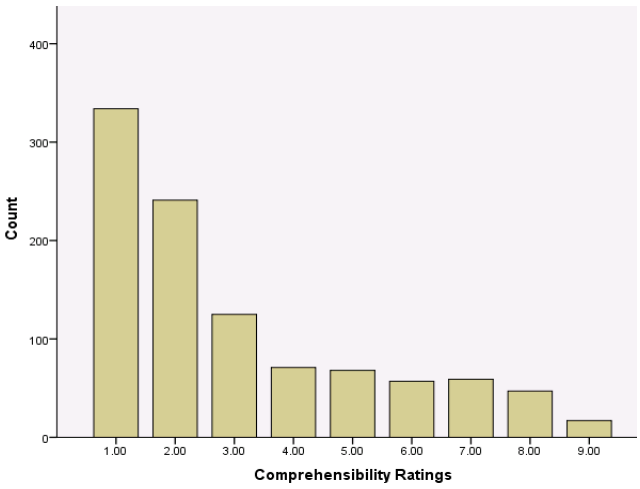
** Correlation is significant at the 0.05 level (2-tailed).

Appendix H

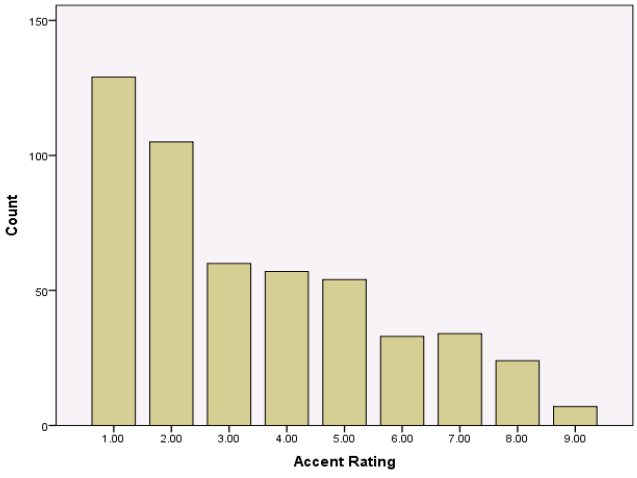
Histograms showing the frequency of responses for accentedness ratings and comprehensibility ratings for the combined bilingual and monolingual groups, as well as the individual bilingual and monolingual groups.

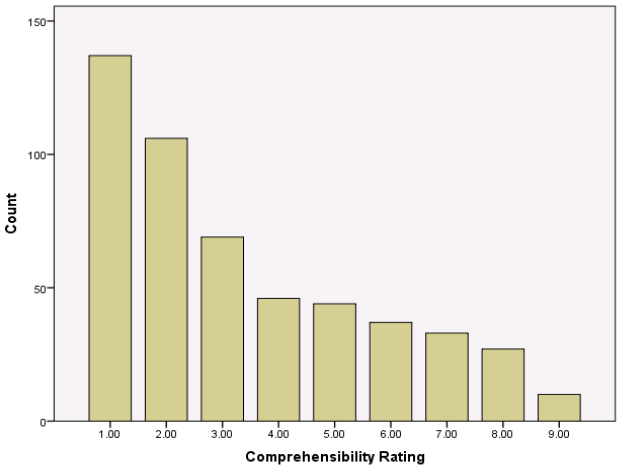
Combined Ratings N=34





Bilingual Participants N=17





Monolingual Participants N=17