12-2021

The Relationship of L2 Proficiency and L1 Reading Academic Performance Prior to Formal L1 Literacy instruction in Foreign Language Immersion Settings

Andrew David McKay Jr

**THE RELATIONSHIP OF L2 PROFICIENY AND L1 READING ACADEMIC
PERFORAMNCE PRIOR TO FORMAL L1 LITERACY INSTRUCTION IN FOREIGN
LANGUAGE IMMERSION SETTINGS**

By

Andrew David McKay Jr.

B.S., South Dakota State University, 2007
M.S., State University of New York College at Buffalo, 2013

A Dissertation Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Education

_____

Division of Curriculum and Instruction

Curriculum and Instruction Program
In the Graduate School
The University of South Dakota
December 2021

The members of the Committee appointed to examine
the dissertation of Andrew David McKay Jr. find it
satisfactory and recommend that it be accepted.

_____

Chairperson, Dr. Susan Gapp

_____

Dr. Garreth Zalud

_____

Dr. Dyanis Conrad

_____

Dr. Steven Chesnut

ABSTRACT

The purpose of this study is to explore the statistical nature of the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in foreign language immersion programs. Specifically, the interdependent relationship between L2 language proficiency levels and L1 reading academic performance in an environment where L1 is the majority language and L2 is the target language for instruction with students who have yet to receive formal reading instruction in the L1. The Continua Model of Biliteracy was used to frame the existing research on language acquisition. An experimental *ex post facto* design using archival data from 228 students in a Spanish immersion program in the Midwest was used. A Spearman rank correlation was used to compare L2 proficiency (STAMP 4Se assessment) to L1 reading academic performance (MAP K-2 reading assessment). The findings of this study showed a strong relationship between L1 reading academic performance and L2 composite scores, L2 reading proficiency, and L2 listening proficiency scores, respectively. A moderate correlation between L1 reading academic performance and L2 writing scores was established, as well as a weak correlation between L1 reading academic performance and L2 speaking proficiency. The researcher concluded that programs and teachers can use L2 proficiency composite and interpretive proficiency indicators as a way to make instructional decisions and monitor growth prior to formal L1 instruction and/or between and before L1 reading high stakes assessment periods.

This abstract of approximately 250 words is approved as to form and content. I recommend its publication.

Dissertation Advisor _____

Dr. Susan Gapp

Table of Contents

Contents

List of Tables

Table

List of Figures

Figure

Chapter 1

The demand for bilingual, biliterate people in our workforce is growing (Genesee, 2008; Wei, 2018). Bilingual workers and students have access to more opportunities than their monolingual counterparts and the growth of one and two-way immersion schools in North America (Genesee, 2008; Wei, 2018) reflects this growing need. Both types of programs rely on societal dominant language families to value bilingualism enough to fill the seats of the programs and to trust the program models enough that the reduction or delay of literacy instruction in the first language (L1) will not adversely impact their children's literacy development in the long term (Genesee, 2008; Wei, 2018).

Families assuming that programs using an additive language model will not negatively impact their child, can find this assumption broadly supported by research in bilingual education (August & Shanahan, 2008; Fred Genesee et al., 2006; National Research Council and Institute of Medicine et al., 1998; Slavin & Cheung, 2005). However, because much of the research and theory is centered on students who are acquiring a societal dominant second language (L2), foreign language immersion programs that delay literacy instruction in the societal dominant language initially may experience a unique set of tensions between the languages of instruction and assessment in the early years of the program.

Foreign Language Immersion Models

One model for developing bilingualism and biliteracy in elementary school students is the use of the foreign language immersion model, also referred to as full-immersion or one-way immersion. One-way immersion models in the United States are additive language programs where the societal dominant language (English in the US) is every student's first language (L1), and students are instructed in the target language (L2 - not English) for a large part of the day.

1

Note this model is distinct from programs commonly used to address the language needs of English learners in that one-way immersion is universally language additive and the group served has a homogenous L1, which is the societal dominant language.

The Center for Applied Linguistics (CAL) describes high quality bilingual programs as being composed of three pillars: Bilingualism & Biliteracy, Academic Achievement, and Sociocultural Competence (Howard et al., 2018). The first pillar, Bilingualism & Biliteracy is addressed pedagogically through content-driven language instruction. This means that students use the target language in order to acquire understanding of academic content, as opposed to the study of the language itself (Genesee, 2008). For this reason, many bilingual programs in the United States have adopted the American Council of Teachers of Foreign Languages' (ACTFL) language proficiency scale as a metric of student L2 development (Ruiz-Funes, 2020). A language proficiency scale is distinct from a language assessment. While language assessments typically assess some specific components of a language, a proficiency scale's purpose is to determine how communicative one is in the target language. ACTFL's (and other language proficiency measures) scale aims to determine the degree to which one can understand and be understood in the target language, regardless of topic as opposed to an assessment of the language itself. For this reason, the ACTFL scale is based on five modes of communication: speaking, listening, reading, writing, and interpersonal. Each mode has its own separate set of qualitative descriptors; however, the descriptions are organized into common levels: Novice (low, mid, high), Intermediate (low, mid, high), Advanced (low, mid, high), Superior and Distinguished (ACTFL, 2012). All modes (except interpersonal) are widely assessed using a range of instruments based on the proficiency scale developed by ACTFL (Ruiz-Funes, 2020).

One-way immersion programs in the United States defer early literacy instruction in

English, commonly until 2nd or 3rd grade, so that students can be immersed in and develop their L2 literacy and language skills. However, by definition these programs are additive language programs, meaning the goal is to develop both L1 and L2 simultaneously, and are still accountable to all their stakeholders for L1 (English in the United States) literacy outcomes on par with traditional programs. In the early grades, parents, teachers, and administrators are often left in limbo as to a student's L1 literacy development, however, L2 proficiency and literacy are constantly monitored. Teachers and programs that are required to use high stakes testing at the district or state level are in effect being evaluated on the basis of assessments given in a language different from the language used for instruction.

Framing Foreign Language Immersion Models

Cummins' (1976, 2017) Linguistic Interdependence hypothesis argues that language and literacy skills are associated across the languages of bilinguals. This concept was further articulated in transfer theory (MacSwan & Pray, 2005). The search for transfer across languages attempts to articulate the exact way in which these skills are interdependent. For example, there is consensus amongst the research that some L1 literacy skills are predictive of L2 language and literacy development. This exists for both English-language learners (ELL) (August & Shanahan, 2008; Cárdenas-Hagan et al., 2007; Mishra & Stainthorp, 2007) and majority language students acquiring a second language (D'Angelo et al., 2017; Shum et al., 2016; Siu & Ho, 2015). However, since teachers in foreign language immersion models are exclusively teaching in the L2 in the initial grades of elementary school, L1 predicting L2 success does not present a useful tool for instruction. Additionally, the precise ways in which these skills relate to each other are impacted by a host of factors both inside and outside of the classroom (MacSwan & Rolstad, 2005), differ across languages and contexts (Hornberger, 2002) and do not always account for

3

both language proficiency and academic content knowledge (MacSwan et al., 2017).

The Continua Model of Biliteracy (Hornberger, 2002) is a tool for framing the research and theory around bilingualism and language acquisition (see Appendix A). This tool provides us with four domains: Context, Development, Content, and Media. Each of these domains contains three intersecting continua. The continua of the Context domain are macro-micro, oral-literate, and bi(multi)lingual-monolingual. Development is made up of the production-reception, oral-written, and L1-L2 continua. The Content domain is characterized by minority-majority, vernacular-literary and contextualized-decontextualized. Finally, Media contains successive-simultaneous, similar-dissimilar and divergent-convergent. The Continua Model of biliteracy, "uses the notion of intersecting and nested continua to demonstrate the multiple and complex interrelationships between bilingualism and literacy" (Hornberger, 2004, p. 156).

Applying the framework presented by the Continua Model to English speaking students learning Spanish in a foreign language immersion setting allows us to contextualize the study for this specific bilingual model. For example, the oral-literate continua from the Context domain is present in the focus on the interdependence of L1 literacy as it relates to L2 proficiency. The Development domain can be characterized along the production-reception (L2 writing and speaking proficiency as it relates to L1 literacy), oral-written, and L1-L2 continua. The Content domain is represented along the minority-majority and contextualized-decontextualized continua. We are investigating the reading development for students who are learning a minority (Spanish) language in a highly contextualized format yet being assessed in the majority (English) language by using highly decontextualized L1 literacy assessments. Finally, the Media domain is characterized by the simultaneous-successive and similar-dissimilar continua. English speaking students learning Spanish in a foreign language immersion setting are taught using a successive

approach (partner language first) to literacy instruction. Spanish and English are highly similar in structure (Fred Genesee & Nicoladis, 2008; Proctor et al., 2010) and the goal of the program model is that students acquire English language and literacy both at home and school in the early years while acquiring almost all of the Spanish at school.

Statement of the Problem

Additive foreign language immersion models have become more widely available in the United States to meet the demand and need for a bilingual workforce (Genesee, 2008; Wei, 2018). While students learn content in L2, the successive, partner language first model of literacy instruction requires L1 literacy instruction to be deferred for a period of time (Howard et al., 2009). This delay in instruction is typically 2-4 years and exists even though state and district level requirements may require these programs to test their students L1 literacy development.

There are several studies on the academic outcomes for students in these programs (Bournot-Trites & Tellowitz, 2002; F. Genesee & Stanley, 1976; Lambert & Tucker, 1972; Swain & Lapkin, 1982; Turnbull et al., 2000). Of those studies, there are a few on L2 development in foreign language immersion programs (T. W. Fortune & Ju, 2017; T. W. Fortune & Tedick, 2015; Young, 2015). Furthermore, there is a wealth of research demonstrating a correlation between L1 language and literacy skills and L2 literacy development (August & Shanahan, 2008; Fred Genesee et al., 2006; Krenca et al., 2020; National Research Council and Institute of Medicine et al., 1998; Slavin & Cheung, 2005), particularly in the context of English-language learners. However, few studies have explored whether a relationship between L2 language development and L1 literacy development exists in foreign language immersion models (Proctor et al., 2010). Although "students in bilingual programs who speak a dominant societal language usually develop the same levels of proficiency in all aspects of the L1 as comparable

5

students in programs where the L1 is the exclusive medium of instruction" (Genesee, 2008, p. 6),

the precise ways the interdependent relationship operates between language and literacy across

the two languages is unclear. Specifically in the early years, before literacy instruction in the L1

is introduced to students.

Purpose of the Study

The purpose of this study is to explore the statistical nature of the linguistic

interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in

foreign language immersion programs. Specifically, the interdependent relationship between L2

language proficiency levels and L1 reading academic performance in an environment where L1

is the majority language and L2 is the target language for instruction with students who have yet

to receive formal reading instruction in the L1.

Research Questions

The following questions will be used in this study:

1. To what extent and in what way is L1 reading performance associated with L2 language
   proficiency composite scores?

2. To what extent and in what way is L1 reading performance individually associated with
   L2 reading, listening, speaking, and writing proficiencies?

Significance of Study

This study aims to elaborate on the relationships of L2 proficiency and L1 literacy in the

context described above, more narrowly defining the interdependent association of language and

literacy skills across the two languages for the group of students in the study. A stronger

understanding of the interdependent relationship of literacy and language acquisition for the

student population represented in this study presents implications for instruction, curriculum, and

assessment moving forward. Furthermore, should the relationship between a specific L2 communicative mode (speaking, writing, listening, reading) show to be more closely related to L1 reading academic achievement, this study provides a baseline for further academic inquiry into the relationship.

Definition of Terms

In instances where the definition does not contain a citation, the term was defined by the researcher.

L1: A person's first language.

L2: A label for any additional language a person uses or acquires.

L2 composite scores: In this study, L2 composite scores are generated by the STAndards-based Measure of Proficiency (STAMP) 4se by averaging the four scores for: Speaking, Listening, Reading, and Writing.

L2 proficiency: The language proficiency level for the second language using the American Council on the Teaching of Foreign Languages (ACTFL) scale with scores produced by the STAMP 4se assessment.

L1 reading academic achievement: Refers to a student's overall Rasch unIT (RIT) score on the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) K-2 Reading assessment.

Language Proficiency: The measure of a persons' ability to meaningfully, effectively, and authentically communicate in a given language (CASLS, 2012a).

Majority Language: "...a majority societal language (e.g., English in Canada, or Japanese in Japan) (Genesee, 2008, p. 3).

MAP: "MAP Growth tests are vertically scaled interim assessments that are also administered in

the form of CAT [computer adaptive testing]. MAP tests are constructed to measure student achievements from Grades K to 12th…" (NWEA, 2015, p. 3).

One way immersion program: "A dual language program in which students are primarily native English speakers learning a foreign language" (CAL, 2016).

Overall RIT: "MAP Growth scores are reported with Rasch Unit (RIT) scale with a range from 100 to 350. Each subject has its own RIT scale" (NWEA, 2015, p. 3). The overall RIT score is an average of the RIT scores of the four subcategories of reading assessed by the K-2 MAP Growth assessment: Language and Writing, Foundational Skills, Vocabulary Use and Functions, and Literature and Informational Text.

STAMP 4se: A general language proficiency assessment designed for elementary students. Test scores are reported to the closest ACTFL scale correlate (CASLS, 2012a, p. 3).

Limitations

The following are limitations to this study:

1. The sample used in the study are second grade students from a Spanish foreign language immersion school in the Midwest. This particular program model begins L1 literacy instruction in third grade. Findings from this study cannot be generalized to the larger population.

2. The study is limited to ex post facto data gathered previous to the 2019-2020 school year. The sample for this study contains all valid assessment data prior to the changes to the instructional model necessitated by COVID-19.

3. The ABC School District requires the administration of the MAP assessment in the fall, winter, and spring of kindergarten through fifth grade. Teachers and administrators follow district and school guidelines for proctoring the assessments.

4. Beginning in the fall of the 2018-19 school year, the ABC School required the administration of the STAMP 4Se to all second-grade students in the spring who are enrolled in the Spanish Immersion program. Previous to this year, students were randomly selected by the school district to take all or some of the STAMP 4Se assessment components.

5. Another limitation is presented in the methodology. In ex post facto design, one cannot assume causation or manipulate variables. Although we can determine the predictive nature of L2 proficiency on L1 literacy development, this does not prove causality. Additionally, any correlation found in question one is bidirectional in nature.

6. The total valid literacy and language proficiency assessments varies by domain (see limitation #4) which ruled out some statistical procedures in addition to impacting statistical power.

7. Students participating in the ABC school immersion program comprise a self-selected group as families choose to enter the specialty program as opposed to their neighborhood school upon entering kindergarten.

8. The writing portion of the language proficiency assessment was administered via paper-pencil to the 2018-2019 group, while previous groups completed the writing portion of the assessment via keyboards. This study does not consider the differences (if any) these two mediums have on student writing production.

9. While this study will look at the connection between the language of instruction and assessment in the early years of a foreign language immersion program, a host of additional factors, both inside and outside of school, play a role in children's language and literacy development across two languages and were not considered or controlled for

within the scope of this study.

Delimitations

1. The sample for the study will be based on students who were enrolled in the ABC school Spanish immersion program from kindergarten through second grade and who completed both the STAMP 4Se and MAP assessments in the spring of second grade.

2. The MAP reading assessment has two versions: a K-2 and a 2-5 version. While the vast majority of students begin taking the 2-5 version in the fall of 3rd grade, some students transition to the 2-5 version in the winter or spring of second grade. Students who transitioned to the 2-5 test at any point of second grade will be excluded from the study.

3. Because of the current interpretation of federal SPED code, students with 0525 coded onto their IEPs received literacy instruction in their L1 as part of their SPED services, these students will be excluded from the study.

Assumptions

1. It is assumed that both the reading assessment data and the language proficiency assessment data used for this study are an accurate assessment of the achievement of the individual students.

2. It is assumed that students received similar, effective instruction in K-2. Classroom teachers adhered to the one-way immersion model and taught literacy using similar curriculum and practices.

Organization of the Remainder of the Study

Chapter 1 outlined the purpose and need for this research. In Chapter 2, a review of the related literature is provided. Chapter 3 describes the methods and procedures used to carry out the research. Data analysis for each of the research questions is presented in Chapter 4. Chapter 5

contains the summary, any major findings, conclusions, discussion, and recommendations for further study.

Chapter 2

Review of Related Literature and Research

This chapter provides a description of the research into bilingual programs, L2 proficiency development, and literacy and academic development in bilingual students. It also presents a tool for framing the research across different program models and populations as well as a description of language development theory in bilingual and multilingual settings. The chapter is divided into the following sections: (a) bilingual education, (b) Continua Model of Biliteracy, (c) full immersion programs: academic achievement, (d) full immersion programs: L2 acquisition and development, (e) full immersion programs: L1 literacy development, (f) full immersion programs: L1 literacy development and L2 proficiency.

Bilingual Education

Bilingual education is distinctive from traditional models of instruction in the United States in that at its core, the framework presents an additive language model in addition to grade level content instruction. The Center for Applied Linguistics (CAL) describes high quality bilingual programs as being comprised of three pillars: Bilingualism & Biliteracy, Academic Achievement, and Sociocultural Competence (Howard et al., 2018). The focus on bilingualism and biliteracy and sociocultural competence distinguishes bilingual education from traditional instructional models. Longitudinal research has consistently demonstrated that an additive language program model is the most effective model for bilingual and biliteracy development in all learners and these programs lead to both proficiency in the target language and increased content area achievement (Faulkner-Bond et al., 2012; Fred Genesee & Lindholm-Leary, 2010). Additionally, students who replace or lose a language tend to have less success in education over time (Housen, 2002; Thomas & Collier, 2003; Wei, 2018). The terms bilingual education or dual

12

language programs are umbrella terms that describe a wide array of program models with differing target demographics, program outcomes, and instructional delivery models. Besides being language additive in nature, these programs have other, specific structures in common.

Similarities Across Bilingual Programs

An additive language model is not the only unique similarity across differing bilingual education models. These programs also have a similar set of pedagogical principles and goals. Genesee (2008) describes the pedagogy of bilingual education as one that facilitates the development of an additional language through both the promotion of "bilingual competence" and the instruction being "content-driven" (p. 3) as opposed to the language-driven instruction of traditional high-school world language courses and programs.

Bilingual programs' goals are centered on the three pillars articulated by CAL and supported by research. The pedagogical focus on the Bilingualism and Biliteracy pillar was described above, bilingual competence and content driven instruction. The Academic Achievement pillar is based on the premise that bilingual programs have enriched, as opposed to remedial, instructional models (Fred Genesee & Lindholm-Leary, 2010; Montecel & Danini, 2002). And programs which have the Sociocultural Competence pillar in place demonstrate an improvement in both cross-cultural attitudes and self-esteem from students (De Jong & Bearse, 2011; Lindholm-Leary & Howard, 2008). Additionally, well implemented programs which are built on the three pillars have been shown to outperform traditional English-only instructional models (National Academies of Sciences, Engineering, and Medicine et al., 2017). Well implemented bilingual programs consist of additive language models and are built on the three pillars laid out from CAL; however, different instructional models across bilingual programs exist.

Differences Across Bilingual Programs

Bilingual programs currently have many instructional models based on allocation of instruction across languages, outcome goals, and student language demographics. Ideally, the differences in bilingual education programs are attributed to differences in student demographics and the instruction necessary to support students as they acquire a second language and move towards the program's intended outcomes (Howard et al., 2018; Johnson, 1997). A K-5 program, with no 6-12 pathway for students to continue building on the language they acquired as they progress has an inherently different goal at the program level than Utah's dual language state driven model which starts university coursework in high school (Valdez et al., 2016).

Outcome Goals

Starting with the end in mind, the stated intended outcomes for students of a bilingual program have an impact on all other aspects of the program. A program model that prepares students for university coursework in the target language will be different from a program focused on using the partner language to support student English proficiency. Program duration is also significant. Students need an average of 6 years in order to achieve high levels of proficiency and grade-level achievement in the target language (Fred Genesee & Lindholm-Leary, 2010; Thompson, 2017) and it's important that students stay in the same model throughout their k-12 career as students who end up participating in a mix of programs have the lowest outcomes compared to groups who stay in a single program or are in traditional monolingual programs (Borsato & Lindholm-Leary, 2006). Bilingual programs can be designed for different lengths of time and have different intended outcomes for the students who participate in them while still being additive in nature and upholding the three pillars described by CAL.

Student Language Demographics. In this context, demographics is used to refer to the language demographics of the student population and the relationship of the societal dominant and partner languages to each other, and to the student population, respectively. Student demographics, in this context does not refer to SES status, students with special needs or learning disabilities, or race. In fact, students with special education needs are usually accepted into the programs and perform well (Fred Genesee & Lindholm-Leary, 2013). Additionally, students who begin a bilingual program before they have an identified learning disability are not removed from the program based on having these needs identified (Howard et al., 2018). The only consideration to be made, most effectively at the individual student level, is in regard to native English-speaking students with serious language processing difficulties (Fred Genesee, 2007; Howard et al., 2018).

Specific research on economically disadvantaged African American students shows that participation in dual language programs increases outcomes when compared to peers in monolingual classrooms (Lindholm-Leary & Howard, 2008; Thomas & Collier, 2012). A well-implemented bilingual program thrives on diversity, as sociocultural competence is one of the three key pillars of the program model. It produces better results, when compared to monolingual instruction, for both English learners and native English speakers (Faulkner-Bond et al., 2012).

When thinking of student demographics in this context, Genesee (2008) articulates an important distinction to consider when discussing bilingual education as he describes a difference between "...programs for students who come to school speaking a majority societal language (e.g. English in Canada, or Japanese in Japan) as well as programs for students who come to school speaking a minority language (e.g. Spanish in the U.S., or Hungarian in

Slovakia)." (p. 3). While the goals of these programs are similar, the issues these two iterations face are both distinct and complex.

With this in mind, bilingual program models in the United States can be organized into three general groups based on the language demographics of the student populations they serve. Program models labeled as Developmental Bilingual typically serve a homogeneous group of students who are native or heritage speakers of the partner language (CAL, 2016). Two-way immersion (TWI) programs are designed for both native English speakers and native speakers of the partner language attending the same class and receiving the same instruction across the two languages, with neither group making up more than two-thirds of the total students (Howard et al., 2018). The third program model is Foreign Language Immersion (also referred to as full-immersion or one-way immersion) and this model is distinct from the previous two in that it is designed for native English-speaking students acquiring a second language (CAL, 2016). In this type of model, students spend most or all of their day in the target language. The students' L1 is homogenous and is usually the majority language for the area in which they live (Johnson, 1997). This model's development can be traced to English speakers living in English-speaking regions of Canada and learning French (Fred Genesee & Nicoladis, 2008). ABC school is most accurately categorized as a foreign language immersion model.

Allocation of Instruction. The allocation of instructional time across the two languages can be thought of in two, broad categories: language of initial literacy instruction and the ratio of total instructional time across the two languages. Bilingual education models can vary at the program level in these two areas. Ideally, the variance is driven by the program's stated outcomes, the student demographics, and best practices (Johnson, 1997). Furthermore, Howard et al. (2018) point out that, "No research has yet determined the best ratio of English to partner

16

language in instruction. Thus, this decision should be made with respect to student outcomes, family and community needs, and in connection with the resources…" (p. 16).

The question of initial literacy instruction can be broken down into three options: All students learn to read in the partner language first, all students learn to read in their native language first, or all students learn to read both languages simultaneously (Howard et al., 2009). The successive vs. simultaneous debate around the sequence of literacy instruction is a non-issue for students in foreign language immersion models as the program model itself dictates all students learn to read the partner language first, later acquiring literacy skills in the societal dominant language (Howard et al., 2018).

The ratio of total instructional time across the two languages is usually described with a ratio. Students attending a 90:10 model in the United States spend 90% of their day in the partner language and 10% of their day in English. In a 50:50 model, typical of TWI models, students spend 50% of their day in each language.

Dual language immersion programs in the U.S. typically range from 50:50 to 90:10, with foreign language immersion programs trending towards the 90:10 side of the range (Howard et al., 2018). In the case of the Spanish foreign language immersion program at ABC school, the ratio changes over time. Students in kindergarten through second grade are 90:10, students in grades third through fifth are 80:20, students in sixth through eighth are 50:50 and students in high school are 20:80. The amount of time in the target language is reduced over time. Additionally, no literacy instruction takes place in the L1 until third grade. Research has shown that although it may be counterintuitive, using the partner language as the sole source of initial literacy instruction for students in this context does not interfere with L1 long term literacy

development as standardized testing shows that these students score similarly on reading tests as their monolingual peers by 3rd or 4th grade (Fred Genesee & Lindholm-Leary, 2013; Vik, 2018).

Although bilingual programs can be lumped together under the commonality of being additive in nature and adhering to the three pillars of dual language developed by CAL, the programs' goals, the language demographics of the students and families they serve, and the allocation of instruction across the languages can be vastly different. For these reasons, it's important to frame the research on bilingual education as it's being presented.

Continua Model of Biliteracy

The Continua Model of Biliteracy (Hornberger, 2004), briefly mentioned in Chapter 1, is a tool that allows us to frame similarities and differences across bilingual programs. In Chapter 1, this tool was used to frame the Spanish Immersion program at ABC school. Moving forward, the tool will be used to frame research for the purposes of identifying program and demographic similarities and differences as they relate to the Spanish Immersion program at ABC school. Hornberger (2002) describes the potential uses of the tool as, "The continua of biliteracy is a comprehensive, ecological model I have proposed as a way to situate research, teaching, and language planning in multilingual settings" (p. 36).

The Continua Model of Biliteracy is organized into four domains with twelve total continua (see Appendix A). The domains are: Development, Media, Context, and Content. Each of these domains contains three intersecting continua. The Development domain provides a framework for thinking about how two or more languages develop in relation to each other. The continua in the Development domain are first language-second language, receptive-productive skills, and oral-written skills. The Media domain provides a medium for thinking about the languages themselves: how similar or dissimilar the linguistic structures are, to what degree the

scripts are more convergent or divergent, and how the exposure to the two languages varies from simultaneous to successive. The Context domain provides a framework for identifying the very context in which the two languages develop: from the micro to the macro level, the mono to multilingual, and the oral to literate continua. The final domain, Content, allows us to frame biliteracy development content from minority to majority perspectives, vernacular to literary styles, and contextualized to decontextualized texts. While it may be tempting to take any bilingual program model and categorize it into the Continua Model of Biliteracy, it's important to remember that the twelve continua are, in fact, continua and not categories. An example of this can be found in thinking about the language demographics of students in a Spanish-English TWI program in the United States. Students are typically organized as L1 English or L1 Spanish for the purposes of maintaining program effectiveness (Howard et al., 2018). For one subgroup of students, English speaking students with English speaking parents who have a very limited set of experiences with Spanish fit nicely into the L1 English category. On the other hand, L1 Spanish is potentially hard to categorize. As an example, some students may have initially spoken Spanish at home, but received most of their language comprehensible input outside of their home, through media or childcare experiences and are dominant in English by the time they arrive at school. This is an example of where a continuum is more useful than a category.

The Continua Model has been used to frame research across methodologies and program models (Gort, 2019; Proctor et al., 2010). Moving forward, the Continua Model will be used to frame research as it relates to the program at ABC school. This will be accomplished by specifically noting the similar and dissimilar continua in each of the four domains and situating the research as towards one end of the continua or the other.

Bilingual Interdependence Theory

Because the Continua Model of Biliteracy is, "...premised on a view of multilingualism as a resource" (Hornberger, 2004, p. 37) it situates itself nicely inside of Cummins' (1976, 2017) Linguistic Interdependence hypothesis and transfer theory as described by MacSwan et al. (MacSwan et al., 2017) which also frame bi and multilingualism as assets. This idea of interdependence across the two languages of bilinguals begins with Cummins' (1976) description of his "threshold hypothesis". Summarily, this hypothesis states that students must achieve a threshold in their L1 and L2 in order to receive the benefits of learning a second language. Conversely, if the threshold is not met, learning a second language can negatively impact L1 and L2 development, as well as content area subject matter knowledge. Evidence of a threshold exists (Cárdenas-Hagan et al., 2007; Fred Genesee & Nicoladis, 2008; Jared et al., 2011; Siu & Ho, 2015). Evidence for the threshold hypothesis is particularly strong for programs similar to ABC school in the broad domains of the continua model of contexts, development, and content (Bournot-Trites & Tellowitz, 2002). Additionally, evidence of the existence of a threshold is persistent in research relating to late immersion programs where students' low L2 proficiency is attributed to low acquisition of complex subject matter presented in L2 (Marsh et al., 2000; Swain, 1996). Yet, two large problems exist with this initial framework: Defining and validating the threshold(s) has proven to be problematic for researchers (Bournot-Trites & Tellowitz, 2002) and the hypothesis, taken alone could lead to unintended negative consequences, specifically in programs that are dissimilar from ABC school in the areas of the Continua Model of contexts and content for speakers of the minority language and their communities (MacSwan et al., 2017) where an L1 threshold is used as a program gate.

Cummins (1976) built on this initial idea with his "developmental interdependence hypothesis" which sees L1 and L2 language and conceptual knowledge as interdependent. This leads to the idea of a "common underlying proficiency" across the languages: a discrete skill or understanding that is accessible to students in one language, is accessible in the other. Cummins (2017) pulls these ideas together using a Spanish-English bilingual program as an example:

> In concrete terms, what this hypothesis implies is that in, for example, a dual language Spanish-English bilingual program in the United States, Spanish instruction that develops Spanish reading and writing skills is not just developing Spanish skills, it is also developing a deeper conceptual and linguistic proficiency that is strongly related to the development of literacy in the majority language (English). In other words, although the surface aspects (e.g., pronunciation, fluency, etc.) of different languages can be distinguished, there is an underlying cognitive/academic proficiency that is common across languages. This common underlying proficiency makes possible the transfer of cognitive/academic or literacy-related proficiency from one language to another. (p. 106)

Although widely adopted and broadly supported by research (Bournot-Trites & Tellowitz, 2002), a similar issue to the threshold hypothesis arises (MacSwan, 2000) in that Cummins lumps language proficiency and academic content knowledge together, implying that language developed around non-academic contexts is less valuable. The separation of language proficiency from academic content knowledge is a key aspect of quality dual-language immersion programs' structure and represented by two of the three separate, but equal pillars of high-quality programs described by CAL at the beginning of this chapter (Howard et al., 2018). Specifically, the pillars of Bilingualism & Biliteracy (representing proficiency) and Academic Achievement (representing academic content knowledge) represent both the separation and the

equal importance of these two ideas. Immersion programs which adhere to the best practices outlined by CAL cannot lump proficiency and academic content knowledge together as it runs against the very design of the program's model.

This issue can be accounted for with transfer theory. Krashen (1996), while emphasizing the role that transfer plays across two languages, did not draw from the language interdependence hypothesis while articulating transfer theory. Instead, Krashen argues that L1 instruction provides the means to access high level L2 content. For example, Krashen (1996) argued that successful EL students who attended monolingual schools tended to have L1 support somewhere outside the classroom. This L1 support provided access and context to the L2 instruction taking place in their classes. MacSwan and Pray (2005) developed this idea further in their articulation of transfer theory. In this denomination of the language interdependence hypothesis, language is viewed as the tool with which academic and conceptual understanding are accessed and articulated (MacSwan, 2017). The idea that multilingual (and bilingual) speakers have a single, integrated language system in which some resources are shared, and some are discreet is largely consistent with Cummins' language interdependence hypothesis (Cummins, 2008, 2017; García, 2011; MacSwan et al., 2017). Yet, it accounts for differences in language proficiency and academic content knowledge development. Moving forward, this specific articulation of the language interdependence hypothesis is the defining aspect of the broader terms of "transfer theory" and "language interdependence hypothesis".

MacSwan et al. (2017) conducted an empirical evaluation of major theoretical frameworks, which included both the version of transfer theory described above and the threshold hypothesis. The strongest evidence was found to support transfer theory, with comparatively less evidence for the threshold hypothesis. Framing the programs and participants

used in this research in the Continua Model of Biliteracy, MacSwan et. al.'s participants and context were similar to ABC school in the continua of structure and scripts. Specifically, because the two languages used in both contexts are Spanish and English. However, MacSwan et al. 's participants consisted of EL students living in the United States, which is dissimilar from the ABC school in the continua of L1-L2 and minority-majority. These differences have an implication on MacSwan et al.'s findings around the threshold hypothesis. They attribute the weak predictive nature of Spanish oral language as evidence against the threshold hypothesis. This finding, in this context, is consistent with existing research (Cárdenas-Hagan et al., 2007; Shum et al., 2016; Uchikoshi & Marinova-Todd, 2012). However, when L1 is the majority language, evidence of L1 oral proficiency predicting later L2 achievement exists (Siu & Ho, 2015). The role of the initial L1 oral proficiency is still considered an open question in full immersion contexts (Dicks & Genesee, 2017). MacSwan et al. (2017) found strong evidence consistent with transfer theory. Specifically, that Spanish literacy, along with English language proficiency predict academic achievement in English. Stated another way, L1 literacy, along with L2 language proficiency predicted L2 academic achievement.

Full Immersion Programs: Academic Achievement

Full Immersion programs concerned with the high academic achievement pillar and beholden to local stakeholders in the majority language are concerned with L1 academic achievement as this tends to be a factor in how those programs are evaluated (Howard et al., 2018).

Swain and Lapkin's (1982) research on early immersion students in Canada can be framed with the continua model as extremely similar to the program at ABC school in each of the twelve continua, the major difference being the target language (French instead of Spanish).

However, both languages are orthographically similar to English, which is the L1 for both groups of students. Swain and Lapkin (1982) found that early immersion students performed as well as their monolingual counterparts on standardized tests in math and science. In another study, which consisted of similar programs and demographics as Swain and Lapkin's work, Turnbull, Hart and Lapkin (2000) found similar scores between immersion and non-immersion students in grade 3. By grade 6, students in immersion programs outperformed their counterparts in both monolingual and enrichment programs in mathematics. This adds weight to an argument for the positive impacts of immersion education beyond the scope of literacy and language development.

Taking a broader scope, Thomas and Collier's (2003, 2012) longitudinal research on immersion programs found positive impacts for students in the areas of academic achievement, literacy development, and social development. Although Thomas and Collier's work includes mostly TWI and transitional programs, making it dissimilar from ABC school in the continua of L1-L2 and minority-majority for some of their participants, it does provide some additional evidence for the positive impact of immersion education on academic achievement.

Historically, the results of research on academic achievement in full immersion programs have been positive (Bournot-Trites & Tellowitz, 2002). Additionally, empirical research has shown positive impacts of bilingualism on social and cognitive development (Hakuta & Diaz, 1985; Thomas & Collier, 2003) providing evidence for the positive impacts of bilingualism beyond the classroom.

Full Immersion Programs: L2 Acquisition and Development

Immersion programs in the United States typically have relied on guidelines developed by the American Council on the Teaching of Foreign Languages (ACTFL) to measure L2

proficiency (Ruiz-Funes, 2020). These guidelines describe L2 abilities across five modes: reading, writing, speaking, listening and interpersonal. Each mode has its own separate set of qualitative descriptors; however, the descriptions are organized into common levels: Novice (low, mid, high), Intermediate (low, mid, high), Advanced (low, mid, high), Superior and Distinguished. ACTFL (2012) describes the purpose of their guidelines thusly:

> These Guidelines present the levels of proficiency as ranges and describe what an individual can and cannot do with language at each level, regardless of where, when, or how the language was acquired. Together these levels form a hierarchy in which each level subsumes all lower levels. The Guidelines are not based on any particular theory, pedagogical method, or educational curriculum. They neither describe how an individual learns a language or prescribe how an individual should learn a language, and they should not be used for such purposes. (p. 3)

The ubiquitous use of ACTFL's guidelines in the United States to describe and rate L2 proficiency has led to a host of assessments based on the guidelines. Some assessments have been developed by ACTFL, such as the oral proficiency interview (OPI) and the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL) (ACTFL, n.d.). While other organizations have adopted ACTFL's proficiency guidelines to develop their own assessments. Examples of this include the University of Oregon's Center for Applied Second Language Studies' (CASLS) Standards-Based Measurement of Proficiency (STAMP) (Avant Assessment, 2015) and the Center for Applied Linguistics' (CAL) Early Language Learning Oral Proficiency Assessment (ELLOPA), the Cal Oral Proficiency Exam (COPA), and Student Oral Proficiency Assessment (SOPA) (CAL, 2000). All of these assessments measure some or all of the modes of communication described in ACTFL's proficiency scale, excluding the

interpersonal mode. The complexities of designing a reliable language interpersonal assessment tool that can be used across a range of program types and languages has proven to be too complex.

Research on L2 proficiency development across a broad swath of elementary bilingual programs generally shows learners increasing across the proficiency levels from grade to grade, specifically in the early grades as the threshold for moving from level to level is smaller (Potowski, 2007; Young, 2012; Ruiz-Funes, 2020, Fortune & Tedick, 2015; Padilla, Fan, Xu, & Silva, 2013; Fortune & Ju, 2017; Padilla et al., 2013; Potowski, 2007; Young, 2015). Almost all students in immersion programs reach the intermediate level across all four modes by the end of high school, compared to less than half of students in traditional high school foreign language study programs (CASLS, 2012a). More specifically, research around L2 proficiency on Canadian full-immersion programs (similar in all aspects of the continua model to ABC school, other than the L2 is French instead of Spanish) have demonstrated that students typically acquire more complex L2 proficiency skills each subsequent year they are in the program (Genesee, 2008; Spada, 1997).

The most recent data from Avant Assessment (2020) on the scores of Spanish immersion students on the STAMP 4Se assessment, which is a language proficiency assessment designed for students in grades 2-5, demonstrates the gains that students make in their L2 proficiency across all four modes in each year of immersion. A key, interpretive table, and the composite results from the 2018-19 school year are represented in Figure 1. Results from the 2018-2019 school year for each of the four modes assessed are represented in Figure 2.

Figure 1

Interpretive Table and 2018-19 Composite Results by Grade of the STAMP 4SE Assessment.

KEY: The numbers for each skill are averages for that particular test group/grade and indicate the average level of achievement for that group. A "2.7" for example, would indicate that this group is approaching Novice-High "3".

1	2	3	4	5	6	7	8	9
Novice			Intermediate			Advanced		
Low	Mid	High	Low	Mid	High	Low	Mid	High
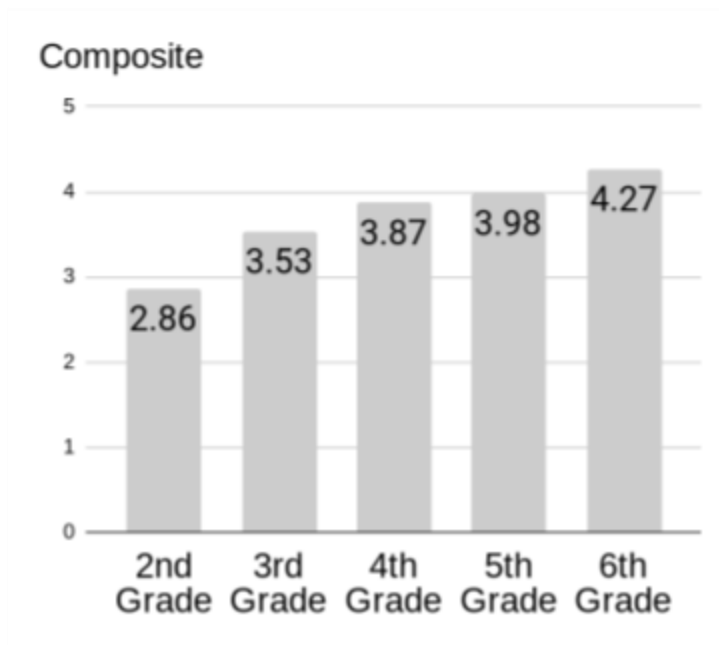
Composite

Note. The 4Se has a 'ceiling' score of 6+ and the 'Composite' represents an average of the 4 domains. Redrawn from Avant Assessment. (2020). Avant STAMP results: National averages 2018–2019.

Figure 2

2018-19 Results by Grade Level of the STAMP 4SE Assessment for each mode.

Reading

Grade	Score
2nd Grade	2.91
3rd Grade	3.49
4th Grade	3.98
5th Grade	4.14
6th Grade	4.58

Listening

Grade	Score
2nd Grade	3.33
3rd Grade	4.21
4th Grade	4.58
5th Grade	4.64
6th Grade	4.94

Writing

Grade	Score
2nd Grade	2.97
3rd Grade	3.27
4th Grade	3.52
5th Grade	3.7
6th Grade	3.92

Speaking

Grade	Score
2nd Grade	2.8
3rd Grade	3.33
4th Grade	3.56
5th Grade	3.61
6th Grade	3.68
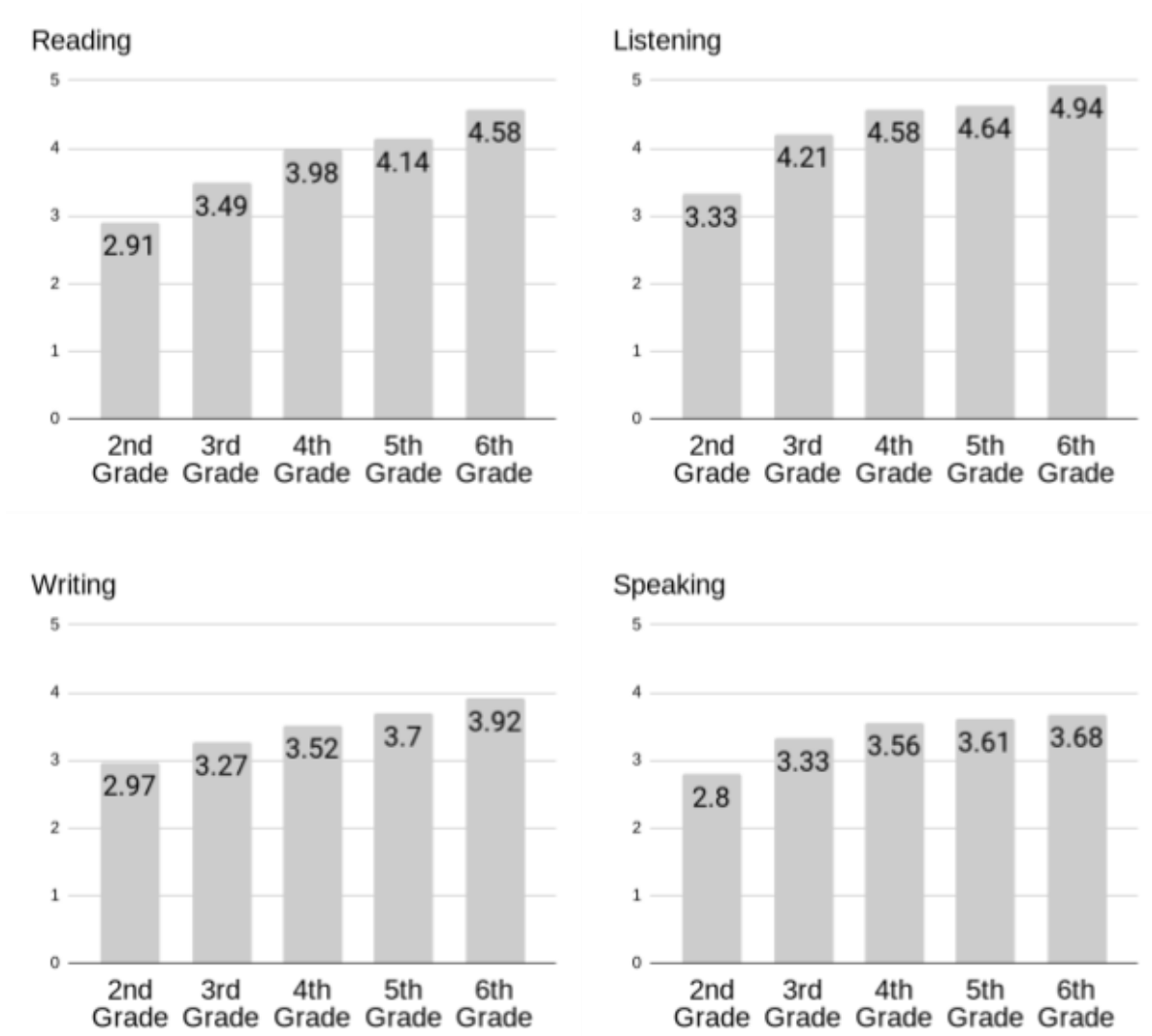
Note. The 4Se has a 'ceiling' score of 6+. Redrawn from Avant Assessment. (2020). Avant STAMP results: National averages 2018–2019.

  Although this data cannot be disaggregated by the type of immersion model (TWI vs. Full, for example) and is not longitudinal in nature, it does provide clear evidence that students in immersion programs acquire increasingly complex L2 proficiency each year of school across all

four modes. However, using this data at either the student or program level proves to be problematic. Avant Assessment (2020) describes an increase in the composite score of .4 to .7 for each student as typical of immersion programs. However, this general guideline is based on the 200,000 'valid' assessments given across seven languages and includes a range of immersion program models. Fortune and Tedick (2015) name the complexities of dealing with the information presented above at the program level, "The current research literature from one-way immersion programs in the United States also reveals several inadequacies, including a majority focus on single program, cross-sectional evaluation studies, aggregate data reporting that combines different languages and program types..." (p. 6).

It's important to note that some authors have described a 'plateau effect' specifically around the productive skills (speaking and writing) in late elementary and middle school (Fortune & Tedick, 2015) for students in full immersion program models. A strong example of this effect appears in Figure 2 in Speaking for grades 4, 5 and 6. This is generally attributed to the complexity of language at the Intermediate and Advanced levels of the ACTFL guidelines. Because each level subsumes all other proficiency levels, ACTFL (2012) often presents a visual of their levels in the shape of an ice-cream cone. Where the Novice levels form the tip, and each progressive level is stacked on top of the previous level. This visual is intended to communicate the added complexity of the upper levels as compared to the lower levels. The distance between Novice Low and Novice Mid is much smaller than the difference between Advanced Mid and Advanced High on the scale. Studies of immersion students' L2 proficiency development in the later levels shows that although students' use of linguistic features demonstrate an increasingly more complex and accurate skillset over time, they simply take longer to cross the threshold to the next level because of the complexity of the later levels (Potowski, 2007; Spada, 1997;

Fortune & Tedick, 2015; Fortune & Ju, 2017; Padilla et al., 2013). This could describe the small increase in speaking scores in grades 4, 5 and 6 above. Although students are acquiring more complex and accurate speaking skills, the vast majority are not doing so at a rate that allows them to cross to the next level in a single school year.

These findings around L2 proficiency development are broadly consistent with research into bilingual programs that can be framed on the continua model of biliteracy as almost identical with the model at ABC school. Fortune and Tedick (2015) looked at the L2 oral proficiency development of 218 students in four one-way Spanish immersion programs in the Midwest. Facilitators used instruments both initially designed by and then modified in collaboration with CAL to ensure the contexts presented in the assessments would be accessible for the students. The SOPA was used for students in grades k-4 and the COPE for students in grades 5-8. Both of these instruments use the same nine-point scale and provide information about oral proficiency in four areas: oral fluency, vocabulary, grammar, and listening comprehension. The results showed a median score of Novice-High at the end of kindergarten and a median score of Intermediate-High at the end of second grade in the area of oral Spanish proficiency.

Ruiz-Funes (2020) reported specific L2 Spanish proficiency results as part of a broader investigation into L2 literacy and language development using the SOPA and COPE as L2 Spanish assessment instruments. The population for this study (n=9) consisted of the English-speaking students enrolled in a TWI program. They were assessed over a period of six years, from kindergarten through fifth grade. The assessments took place at the end of kindergarten, 2nd, and 5th grade. Because of the differing program models, this population should be framed differently on the continua model of biliteracy than ABC school as the students from the Ruiz-

30

Funes study should be placed further to the simultaneous side of the simultaneous-successive exposure continua. Additionally, the students at ABC school receive approximately twice the amount of exposure to the target language in grades k-2 as compared to the student population from the Ruiz-Funes study. Ruiz-Funes (2020) found consistent growth in L2 proficiency and linguistic complexity in each testing instance. Most of the students were unratable at the end of kindergarten and scored in the Novice-high/Intermediate-low range at the end of 5th grade. The continual growth, when compared to the plateau described by Fortune and Tedick in their work above, can most likely be attributed to the fact that the students L2 proficiency did not develop at the same rate as the students in the full immersion setting, therefore they did not reach the plateau before 5th grade.

L2 proficiency for majority language speakers in immersion programs is a central piece of immersion programs. The focus on bilingualism and biliteracy is described as one of the three pillars by CAL (Howard et al., 2018) and was articulated earlier in the chapter. The opportunity to acquire a second language is a motivating factor for many parents to enroll their children in these programs (Genesee, 2008; Wei 2018). However, bilingualism and biliteracy is just one of three pillars that are central to high quality immersion programs. Another pillar is high academic achievement and one component of high academic achievement, specifically for students in elementary school, is developing the ability to read.

Full Immersion Programs: L1 Literacy Development

L1 literacy development, when referring to full immersion programs in the United States refers to English literacy. The program instructional model, as previously noted, relies on a successive approach to literacy instruction. All the literacy instruction initially takes place in the target language. Students do not typically receive literacy instruction in the L1 until second or

third grade. The vast majority of the research that can be framed on the continua model of biliteracy as similar to that of the ABC school has taken place in Canada on French immersion programs.

Broadly speaking, findings have consistently demonstrated that students' L1 literacy development lags behind monolingual peers until explicit instruction in L1 literacy takes place. Once students are instructed in L1 literacy, they not only catch-up to their monolingual peers, but they also surpass them in L1 literacy achievement (Fred Genesee, 1979; F. Genesee & Stanley, 1976; Lambert & Tucker, 1972; Swain & Lapkin, 1982; Turnbull et al., 2000). Thus, creating a strong argument for the positive impact of immersion on L1 literacy achievement. While these findings have relied on standardized tests and compared immersion students to their monolingual peers, empirical studies have not typically analyzed the specific effects of one-way immersion on L1 literacy development. Bialystock (2007) took note of the surprising lack of research in this area given the probable impact of immersion education on literacy development in both L1 and L2.

Bruck and Genesee (1995) investigated the specific transfer effect of phonological awareness from L2 to L1 by comparing 91 French immersion students to 72 students in monolingual classrooms in kindergarten and grade 1 using phonological awareness assessments proctored in English. At the end of kindergarten, the French immersion group out-performed the monolingual group in onset-rime awareness. At the end of grade 1, the monolingual group scored higher on phoneme awareness tasks which the authors attributed to the group receiving instruction in English phonemes as part of their literacy instruction. However, the French immersion group outperformed the monolingual group on syllable segmentation tasks at the end

of grade 1. The authors reported that L2 syllable-awareness skills resulted in a corresponding development of syllable awareness in L1, even before instruction in L1 took place.

As opposed to looking at the comparative development of monolingual and bilingual students' literacy skills, Lee and Chen (2019) explored the predictive nature of discrete literacy skills within the two languages of a group of French immersion students at the end of grades 2 and 3. A total of 66 students completed both testing sessions, 33 of the students were labeled as L1 English and 33 of the students were labeled as L1 other than English or French. In addition to the specific program model beginning English instruction in 4th grade, as opposed to 3rd grade at ABC school, 50% of the participants' L1 other than English presents important program model distinctions from ABC school, specifically in the continua of L1-L2, simultaneous-successive, and minority-majority.

Students were tested in the spring of 2nd and 3rd grade in the following areas using the following instruments: Nonverbal reasoning (Matrix Analogies Reasoning Test), Phonological Awareness (Elision subtest of the Comprehensive Test of Phonological Processing in English) and a parallel experimental measure in French, Rapid automatized naming (Rapid Digit Naming subtest of the Comprehensive Test of Phonological Awareness in English) and a parallel experimental measure in French, Word Reading accuracy (Letter-Word identification from the Woodcock-Johnson Tests of Achievement in English) and a parallel experimental measure in French, Receptive vocabulary (Peabody Picture Vocabulary Test in English) and the French version of the Peabody test, Word reading fluency (Sight Word Efficiency subtest of the Test of Word Reading Efficiency in English) and a parallel experimental measure in French, and finally, Reading Comprehension was assessed using the Gates-MacGinitie Reading Test in English and the French tasks were generated from translating an alternate version of the same test.

The researchers constructed regressions by entering non-verbal reasoning as the first step and then within-language phonological awareness and word reading accuracy in order to control for known predictors of reading comprehension. In order to find the independent impact on reading comprehension, within-language vocabulary and word reading fluency were mean-centered and entered in the model in the last steps. The researchers found that an interaction between word reading fluency and vocabulary emerged in grade 3 and was significantly related to reading comprehension in the concurrent language. Although these two factors independently contributed to reading comprehension in grade 2, there was no interaction concurrently or longitudinally. Although Lee and Chen (2019) were able to explore the simultaneous development of reading comprehension of bilingual students within each language, this study did not examine any cross-linguistic impact of discrete literacy skill(s) development on the partner language.

Thus far, we've explored the research on literacy development in Canadian French immersion programs as we can place these programs on the continua model of biliteracy close to that of the ABC school in almost all of the domains, however, French and Spanish although orthographically similar, are not the same. Spanish has a more consistent orthographic structure than French and English. Reading (2009) explored the differential effects of French and Spanish on L1 English students in immersion programs. In this study, the authors used the results from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment, proctored in English, to compare students in monolingual English classrooms (n=47) to students in Spanish immersion (n=33) and French immersion (n=21) in kindergarten and first grade. The specific model of immersion at this setting is distinct from the model used at ABC school in that the model used for this research would be placed closer to the simultaneous side of the

simultaneous-successive continua as students in the study in both immersion settings began receiving English instruction for 60 minutes a day in first grade. At the beginning of kindergarten, the French and Spanish groups' scores were similar, however, by the end of kindergarten the Spanish groups' median scores were consistently higher than the French groups, with significant differences emerging in the beginning of first grade. At the end of first grade, both Spanish and French groups scored at or above the first-grade benchmarks, however the Spanish group scored significantly higher than the French group. Additionally, there was a significant difference between the number of students needing L1 reading intervention between the two groups as the French group had a higher number of children in reading interventions. The authors attribute the differential effects of French and Spanish on L1 development to two possible factors. The first is motivation. Because of the consistent orthographic structure of Spanish, children learning to read in Spanish were able to depend on a regular sound-letter correspondence which gave them an extra dose of confidence when approaching reading tasks in either language. The second is the consistent orthographic structure of Spanish when compared to English and French. Reading (2009) addresses both factors, "The regularities that occur often in Spanish served as an assist in learning to detect regularities in English, which are not as easily apparent. The regular letter-sound correspondences of Spanish may have provided a type of extra phonemic awareness and phonics instruction for children learning this language" (p. 139). So even though French and English may be placed closer together on the similar structure - dissimilar structure continua of the continua model of biliteracy than Spanish and English, the results of the differential effects from this study don't necessarily indicate that is inherently better for students in either L1 or L2. If anything, this small study points to the complexities of generalizing bilingual education research across languages and programs.

Reading (2009) aptly describes the current state of literacy research in the context of full immersion programs, "Although much research confirms that L1 literacy skills transfer to L2 and affect literacy development in L2, there are fewer research studies investigating the effect of immersion education and learning L2 on the literacy development of L1" (p. 120). While this presents a small pool of knowledge to draw from, an even smaller pool presents itself when the continua model of biliteracy is used to frame research as having a similar development, context, content and media as that of the program at ABC school.

Full Immersion Programs: L1 Literacy Development and L2 Proficiency

There is some research into the interdependent development of literacy across the two languages of bilingual students in immersion programs as it relates to proficiency. Proctor, August, Snow, and Bar (2010) reported the results of a 3-year cross-sectional study that involved 91 fourth graders from three cities in the United States as part of a larger study on Spanish literacy development in the United States. The program models represented in this study are more closely categorized as transitional and/or TWI than full immersion. The student population is made up of a heterogeneous L1 group, with only 7 reporting mostly or only English spoken at home. This makes framing the program and populations on the continua of biliteracy not only significantly different from the more homogenous grouping of full-immersion programs in general and ABC school specifically, but also most likely best represented by thinking of this group as representing a range on the continua, specifically the L1-L2, minority-majority, simultaneous-successive, bilingual-monolingual, and micro-macro continua.

Participants were assessed in both languages in the areas of alphabetic knowledge (Computer-based Academic Assessment System), Vocabulary Knowledge (Woodcock Picture Vocabulary test), Listening Comprehension (Woodcock Listening Comprehension test), and

Reading Comprehension (Woodcock Passage Comprehension test). Notably, the Spanish assessments used in this study consisted of officially translated published versions of the assessments as opposed to researcher translated or experimental Spanish versions of the tasks.

Using data from the above assessments, Proctor et al. (2010) modeled a restricted set of proficiency and literacy variables cross-linguistically as a method of proposing and testing a continuum of interdependence. Cross-linguistic decoding predicted reading comprehension in both English and Spanish and Spanish and English reading comprehension were closely related to each other. Proctor et al. (2010) conclude that, "...biliterate outcomes may be optimized by literacy instruction delivered in L1 and L2 simultaneously" (p. 17). However, they acknowledge that this does not articulate the exact nature in which the transfer of skills across the two languages takes place or the ways in which proficiency may impact literacy development. Additionally, 3 of the 4 assessments measured literacy skills, while listening comprehension was the only measure used for language proficiency. Furthermore, students in this population bring a heterogeneous set of skills to the study as previously noted and all of the students in the study had already received literacy instruction in both languages at the time of the assessments.

The second and more closely situated on the continua model of biliteracy school to that of ABC school looked at reading strategy use across and within the two languages as it related to the students' L2 proficiency in fourth and fifth graders in French immersion, in a model where L1 literacy instruction begins in fourth grade. Frid and Friesen (2020) examined the L2 proficiency and reading comprehension strategy use across both French and English in a group (n=70) of French immersion students. 66 of the students reported L1 English, placing the vast majority of the population of this study on similar continua of the continua model of biliteracy as the students as the students at ABC school, excluding the target languages.

Frid and Friesen (2020) assessed the group of fourth and fifth graders in the areas of Vocabulary (Peabody Picture Vocabulary Test - English and French versions), Word Reading Efficiency (Test of Word Reading Efficiency in English and a French version) and a researcher developed Comprehension and Strategy Use Task in which students read different stories in English and French and then recorded themselves answering researcher developed 'think aloud prompts' which were analyzed later. The results to the think aloud prompts were organized along a scale of less to more complex and placed into three categories: surface strategies, text-based strategies, and situation model strategies. Regression analysis was used to view the results of student strategy use as they related to vocabulary and word reading within and across the two languages.

Unsurprisingly, students scored significantly higher on the vocabulary and word reading assessments in English than French. Additionally, students used more complex comprehension strategies more frequently in English than French. However, correlations between language proficiency and the complexity of the comprehension strategy used tended to be small. As students tended to use similar comprehension strategies across both languages, adding support to the linguistic interdependence hypothesis on the interdependent nature of reading comprehension strategies. When a comprehension strategy (i.e., summarizing) is available in one language, it's available for recruitment in the other. Situation modeling (the most complex comprehension strategy category) proved to be a valid predictor of reading comprehension performance in both English and French.

Specifically looking at cross-language predictors in this data set, English non-word reading, and English vocabulary knowledge predicted French reading comprehension. French inferences and French word reading predicted English comprehension. Frid and Friesen (2020)

note, "Although cross-language predictors produced significant regression models, they accounted for less variance than within language models" (p. 38). Which may lead one to question whether L2 proficiency has an impact on L1 literacy development. However, there are two important distinctions to keep in mind between this study and ABC school: first, the students had already received literacy instruction in L1 at the time of the study and second, the assessments labeled as language proficiency in this study (vocabulary and word reading) are probably more accurately described as discrete literacy skill assessments than language proficiency tools which measure the degree to which one can understand and be understood in the target language.

Conclusion

This chapter provided a description of the research into bilingual programs, L2 proficiency development, and literacy and academic development in bilingual students. The research has established that students who stay in high quality bilingual programs meet or exceed their monolingual peers over time on standardized tests in L1 academics. Additionally, these students generally lag behind their monolingual peers in the area of L1 literacy until instruction in the L1 is introduced, after which they generally meet or exceed their monolingual peers within a single school year. Furthermore, students in bilingual programs, specifically in the early grades, acquire an increasingly complex set of L2 proficiency skills across all four modes of communication. While much of the current research focuses on the comparative nature of monolingual to bilingual development in the areas of proficiency, academic achievement and literacy development, or the comparative development of literacy within the two languages of bilinguals, the current research on the interdependent nature of L1 literacy skills as they relate to

L2 proficiency is scant, and generally focuses on comprehension, after students have received literacy instruction in L1.

Chapter 2 provided a review of the related literature. Chapter 3 will describe the methods and procedures used to carry out the research. Data analysis for each of the research questions will be presented in Chapter 4. Chapter 5 will contain the summary, any major findings, conclusions, discussion, and recommendations for further study.

Chapter 3

Methodology

The purpose of this study is to explore the statistical nature of the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in foreign language immersion programs. Specifically, the interdependent relationship between L2 language proficiency levels and L1 reading academic performance in an environment where L1 is the majority language and L2 is the target language for instruction with students who have yet to receive formal reading instruction in the L1.

Research Questions

The research questions below were designed to address the problem statement:

1. To what extent and in what way is L1 reading performance associated with L2 language proficiency composite scores?

2. To what extent and in what way is L1 reading performance individually associated with L2 reading, listening, speaking, and writing proficiencies?

Design

This study used a non-experimental *ex post facto* design. The study used archival data from the school district which is stored, tracked, and monitored by the district's assessment management team. Gay and Airasian (2003) describe this specific type of non-experimental design as correlational research. The use of existing data sets in correlational education research is both "efficient" and "economical" (Muijs, 2004) and when the variables identified for use in the research are selected from a theoretical basis the results are more meaningful (Gay & Airasian, 2003). This design, however, precludes the researcher from both randomization and manipulating the independent variable (Thompson & Panacek, 2006). Non-experimental design

is useful in cases where the concurrent control group that is necessary for experimental or quasi-experimental research is not possible or ethical (Thompson & Panacek, 2007). Given both the existing data sets and research questions, non-experimental *ex-post facto* design is both "common" and "ideal" (Thompson & Panacek, 2007), specifically when the research is intended to identify factors that can be further validated later using experimental design (Thompson & Panacek, 2007; Gay & Airasian, 2003). Both research questions are quantitative in nature and call for both descriptive and inferential statistics (Gay & Airasian, 2003).

Population

The population for this study is students in a one-way Spanish immersion program prior to receiving L1 literacy instruction. The sample is composed of students from a school district in the Midwest, which consists of approximately 24,000 students in 36 schools. Specifically, the sample group is composed of all 2nd grade students at the district's ABC Spanish Immersion school with valid STAMP 4Se and K-2 Reading MAP assessment scores in the spring of 2nd grade prior to the 2019-2020 school year. Therefore, 2nd grade students at the ABC Spanish Immersion school from the 2017-2018 and 2018-2019 school years were included in the sample. This school's instructional model is most accurately classified as a one-way Spanish immersion program. The foreign language immersion model adopted by ABC school includes a partner language first model of literacy instruction (Howard et al., 2009). In this model for literacy instruction, students learn to read in the partner language first, Spanish in the case of ABC school. All the students in the population speak English as their first language and have thus far received all their literacy instruction at school in Spanish.

Sample

The total sample meeting the study's criteria is made up of 228 students. Of those students, 124 are male and 104 are female. Fourteen percent of the students in the sample qualify for free/reduced lunch and 7% have a 504 plan. Ten percent of the students in the sample receive special education services. Of the 24 students in the sample who receive special education services, 4 students receive services based on a specific reading disability. These students were identified by an 0525 code on their individual education plan (IEP). These 4 students received literacy instruction in English prior to 3rd grade and were removed from the sample. All the students in the sample speak English as their first language and have thus far received all their literacy instruction at school in Spanish. This accounts for 0% of the students in the sample being identified for district EL services. Differences in the actual test result numbers from the total number of students in the sample exist because not all children took all four sections of the STAMP 4Se assessment each year and some students were identified to move to the Reading MAP 2-5 assessment (as opposed to the Reading MAP K-2 assessment) in the spring of their 2nd grade year per criteria set by the participating school district from which the sample was drawn.

Data Collection

The criteria for the study included all second-grade students who had valid K-2 Reading MAP and STAMP 4Se assessment results in the spring of 2nd grade prior to the 2019-2020 school year and who did not have an 0525 coded IEP.

After the Institutional Review Board (IRB) approval (Appendix C), the researcher submitted an application for Research Approval to the assistant superintendent of academic achievement of the ABC school district (Appendix B). Once this approval was received, the researcher submitted a request to the district's assessment management team. The district's assessment management team located, obtained and organized the archival data that met the

criteria for this study and shared it with the researcher. During this process, the district's

assessment management team de-identified the data, generating random identification numbers

to represent each student. Thus, the researcher never had access to identifying student

information. Additionally, the district's assessment management team password protected the

file before giving it to the researcher. Only the district's management team and the researcher

have access to the password protected file.

Instrumentation

L2 language proficiency and L1 reading academic data was gathered using two separate

instruments. The following sections contain descriptions of each instrument in addition to a

description of the reliability and validity of the instruments.

STAMP 4Se

L2 language proficiency data was gathered using the STAMP 4Se assessment. STAMP is

an acronym for Standards based Measurement of Proficiency (CASLS, 2012a). The STAMP 4S

and 4Se have been used as a proficiency measurement of target language acquisition in research

across a range of immersion models and languages (Amato, 2011; Blanton, 2015; Davin et al.,

2014; Burkhauser et al., 2016). The standards referred to in the assessment's title are the ACTFL

proficiency guidelines (ACTFL, 2012). The STAMP 4Se assessment is designed for students in

grades 2-6. It produces a separate rating, in reference to the ACTFL scale, for each of the four

modes of communication it assesses: reading, speaking, listening, and writing. Each mode has its

own separate set of qualitative descriptors; however, the descriptions are organized into common

levels: Novice (low, mid, high), Intermediate (low, mid, high), Advanced (low, mid, high),

Superior and Distinguished. Additionally, each subsequent level subsumes all other proficiency

levels. A student who receives a score of 2 on the reading portion of the STAMP assessment is

rated as Novice-Mid on the ACTFL proficiency scale, a score of 3 would translate to Novice-High, and so on. A composite score is also generated by averaging the scores across the four modes of communication.

Support for an assessment's validity can be derived from the processes used to develop the assessment (Santos, 2019). In the case of STAMP assessment development, this is generally a 10-step, six-month process. One step of this process involves developing items specifically around the ACTFL Descriptors and NCSSFL-ACTFL Can-Do Statements, while a completely separate step involves revision of items for quality. This process involves both assessment experts, such as professional item writers and statisticians and a specific panel of target language experts (Santos, 2019). For the 4se assessment, this process specifically involved a pilot of items from 2006 through 2007. The pilot consisted of 5 forms delivered to over 7,000 testing instances. A Rasch analysis of the items used in the pilot was conducted in order to identify inconsistent test items, which were then revamped or discarded. In 2007 and 2008, a field test that consisted of 13,000 assessments was delivered, which was used to ensure the testing algorithm was working properly and to inform future changes to items (CASLS, 2012a).

While this is the general process for STAMP assessments across all the languages assessed, AVANT Assessment does produce technical documents for each language. In the specific case of the Spanish assessment, a field test was developed in which over 5,000 students participated. Based on the results of this field test, AVANT Assessment reported 90% accuracy in identifying students' "true" proficiency levels on the Spanish version of the STAMP assessment (CASLS, 2012b).

Listening Subtest. The listening section of the STAMP 4Se assessment consists of dialogues (or monologues) which are delivered twice in the target language and are followed

by a question, which is answered by clicking on a picture, either from a set of four pictures or by identifying the specific part of a picture to which the question is referring. The questions are designed to assess the test-taker's ability to both understand the general topic of the dialogue, as well as extract detailed information from the dialogue. The test is adaptive in that after each sequence of ten questions, the program chooses the next set of prompts and questions based on student response accuracy (CASLS, 2012a).

Reading Subtest. The reading section, like the listening section of the STAMP 4Se is also delivered adaptively and computer scored. In this section, the assessment is used to rate test-takers' reading comprehension in the target language by presenting a passage intended to mimic authentic reading tasks that require no specialized content knowledge. Test-takers respond by selecting from a set of four pictures, clicking on the target area of a single picture, or by reading and answering a multiple-choice question about the passage in the target language (CASLS, 2012a).

Writing Subtest. The writing section of the STAMP 4Se assessment is not adaptive and is manually rated. In this section, test-takers are given two prompts designed to elicit a response in the target language. Each prompt is given separately and presented both in text in the target language and aurally in English. Test-takers are then prompted to write their response in the target language (CASLS, 2012a).

Speaking Subtest. The speaking section, like the writing section of the STAMP 4Se is not adaptive and is manually rated. In this section, test-takers are given two prompts designed to elicit a response in the target language. The prompts are not interactive, given separately, and presented both in text in the target language and presented aurally in English. Test-takers record

their responses using software embedded in the testing system which is then rated by a person (CASLS, 2012a).

Because the speaking and writing responses are rated manually, Avant Assessment (2017) has developed an inter-rater reliability system in which 20% of all responses are sent to a 2nd person for a blind rating. This continual inter-rater reliability ensures that the acceptable score of inter-rater reliability (Gall et al., 2007) demonstrated in the pilot studies (CASLS, 2012b), continues year to year. Speaking and writing raters for the STAMP 4Se are required to have a bachelor's degree, score at least a 90% on Avant's rater certification training, and score at the Advanced level in the target language. If there is a difference in scores between the first two raters, the response is sent to a 3rd rater. The results are tracked and used for training and retraining of the raters. In their last published inter-rater reliability statistics for Spanish, Avant Assessment (2017) reported a 93.84% inter-rater reliability rating. The STAMP 4Se is a psychometrically validated standardized language proficiency assessment scored by external evaluators (Falsgraf, 2009), thereby removing teachers, administrators, and researchers from the scoring process.

MAP

Reading academic data was gathered using the K-2 Reading MAP assessment from the Northwest Evaluation Association (NWEA). The assessment is both computer-adaptive and norm-referenced. This assessment is currently being used in 49 states (NWEA, 2017) and the ABC school's district requires the administration of this assessment three times per year (fall, winter and spring) beginning in kindergarten.

The K-2 Reading assessment consists of 43 test items that count towards a test-taker's score, in addition to some field test items which do not count. The scores are broken out into four

instructional areas on the K-2 MAP assessment: Foundational Skills, Language and Writing, Literature and Informational Texts, and Vocabulary Use and Functions with 6-8 items per instructional area per test. Like the listening and reading portions of the STAMP 4Se, the MAP is computer adaptive, choosing questions in real time based on student responses. Results are communicated through the use of a stable RIT scale, so that assessment results can be tracked within and across grades (NWEA, 2017). Additionally, this stable scale provides the means for educators to compare their students' scores to scores from across the nation. The total score for each subcategory is averaged to produce an "overall RIT" score in addition to the scores for each sub-category.

The National Center on Response to Intervention (2019) rates the K-2 MAP test, specifically when used with second grade students in the spring, as a tool which provides "convincing evidence" for use as both a screening and progress monitoring tool. This rating, as it pertains to a screening tool, is based on classification accuracy, reliability, validity, sample size, and bias analysis. The rating, as it pertains to a progress monitoring tool, is based on reliability, validity, bias analysis, sensitivity (reliability and validity of slope), alternate forms, and decision rules.

NWEA (2017) reports a +/- three-point standard error of measure, a marginal reliability correlation score of 0.944, and a content validity correlation score of 0.968. Student scores from the NWEA Reading MAP assessment have also been found to be both correlated to and predictive of the ELA Smarter Balanced Assessment (SBA) (Bjorklund-Young & Borkoski, 2016; NWEA, 2015) which is a state required assessment for students at the ABC school beginning in the spring of 3rd grade. Specifically, the correlation of student scores from the MAP and SBA was found to range from .80-.89 and Reading MAP scores were found to

48

correctly predict ELA SBA proficiency 84% of the time (NWEA, 2015). Bjorklund-Young and Borkoski (2016) were able to replicate the statistical validity of the MAP assessment using data provided by NWEA.

Data Analysis

As both research questions are quantitative in nature, descriptive and inferential statistics were used (Gay & Airasian, 2003). The goal is to further articulate the interdependent nature of language and literacy development as it pertains to the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the context of a full immersion program.

The study first explored the relationship between L2 language proficiency composite scores and L1 reading academic performance in the context of a full-immersion program prior to students receiving literacy instruction in their L1. The relationship was further articulated by looking for relationships between each L2 proficiency communicative mode and L1 literacy outcomes.

Research Question One (Composite)

The first research question, which explored the relationship between L1 reading academic performance and L2 language proficiency composite scores, was answered using a Spearman rank correlation. The total number of students who had valid K-2 Reading MAP and STAMP 4Se assessments in the spring of 2nd grade prior to the 2019-2020 school year and who did not have an 0525 SPED coded IEP was n=30.

A Spearman rank correlation is appropriate when three assumptions are met (Gay & Airasian, 2003). The first assumption is that the data is ordinal or interval in nature (Muijs, 2004). The ACTFL language proficiency scale is ordinal in nature, while the reading academic achievement score generated from the Reading MAP K-2 test is interval in nature, meeting the

first assumption. The second assumption is that the variables are paired observations (Wackerly et al., 2002). A single student taking two assessments meets this assumption. The last assumption is that the two data sets are monotonic in nature. In order to meet this assumption, a scatterplot was generated in SPSS before conducting the Spearman rank correlation in order to verify that the two data sets are monotonic in nature (Good & Hardin, 2003).

Once these three assumptions were met, the researcher used SPSS to run a Spearman rank correlation. Two variables were created from the archival data obtained from the ABC school district. L2 language proficiency and L1 reading academic performance. L2 language proficiency is represented by composite scores from the STAMP 4Se assessment. Composite means are computed from each of the four language domains measured on the STAMP 4Se assessment. L1 reading academic achievement is represented by the overall RIT score on the K-2 MAP reading assessment. The output table generated by SPSS was examined in order to determine whether a correlation between L2 language proficiency composite scores and L1 reading academic performance exists and if the correlation is statistically significant (Muijs, 2004). A post hoc power analysis was conducted with G*Power statistical software using an exact test family and correlation: bivariate normal model two-tailed statistical test as parameters.

Research Question Two (Individual)

The second research question, which explored the relationship between L1 reading academic performance and individual L2 proficiency scores (reading, listening, speaking, writing), used a Spearman rank correlation. For each of the comparisons, two variables were created from the archival data obtained from the ABC school district. Additionally, a scatterplot was generated in SPSS before conducting each Spearman rank correlation in order to verify that the two data sets are monotonic in nature (Good & Hardin, 2003).

The first comparison was between L1 reading academic performance and L2 reading proficiency. L2 reading proficiency is represented by scores from the reading section of the STAMP 4Se assessment. L1 reading academic achievement is represented by the overall RIT score on the K-2 MAP reading assessment. Overall RIT scores from K-2 MAP reading assessment were used to represent L1 reading academic achievement in each of the comparisons for question two. The total number of students who had valid K-2 Reading MAP and STAMP 4Se reading subtest assessment scores in the spring of 2nd grade prior to the 2019-2020 school year and who did not have an 0525 SPED coded IEP was n=83.

The sample used for the pairwise comparison between L1 reading performance and L2 listening proficiency was n=83. L2 listening proficiency is represented by scores from the listening section of the STAMP 4Se assessment. The sample used for the pairwise comparison between L1 reading performance and L2 speaking proficiency was n=76. L2 speaking proficiency is represented by scores from the speaking section of the STAMP 4Se assessment. The sample used for the pairwise comparison between L1 reading performance and L2 writing proficiency was n=73. L2 writing proficiency is represented by scores from the writing section of the STAMP 4Se assessment.

After meeting the assumptions and creating variables, the researcher used SPSS to run a Spearman rank correlation for each of the comparisons. Each of the output tables generated by SPSS was examined in order to determine whether a correlation between each of the respective L2 proficiency sub tests and L1 reading academic performance exists and if the correlation is statistically significant (Muijs, 2004). A post hoc power analysis was conducted with G*Power statistical software using an exact test family and correlation: bivariate normal model two-tailed statistical test as parameters.

Summary

Chapter 3 described the methods and procedures used to carry out the research. Data analysis for each of the research questions will be presented in Chapter 4. Chapter 5 will contain the summary, any major findings, conclusions, discussion, and recommendations for further study.

Chapter 4

Results

Introduction

This chapter provides the results of the data analysis and the findings of the study. The purpose of this study is to explore the statistical nature of the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in foreign language immersion programs. Specifically, the interdependent relationship between L2 language proficiency levels and L1 reading academic performance in an environment where L1 is the majority language and L2 is the target language for instruction with students who have yet to receive formal reading instruction in the L1.

Research Questions

The research questions below were designed to address the problem statement:

1. To what extent and in what way is L1 reading performance associated with L2 language proficiency composite scores?

2. To what extent and in what way is L1 reading performance individually associated with L2 reading, listening, speaking, and writing proficiencies?

Archival Data

The criteria for the study included all 2nd grade students at the district's ABC Spanish Immersion school with valid STAMP 4Se and K-2 Reading MAP assessment scores in the spring of 2nd grade prior to the 2019-2020 school year without an 0525 IEP. The district's assessment management team located, obtained, and organized the archival data for 2nd graders from the 2016-17, 2017-2018, and 2018-2019 school years. The researcher received data representing 182 students of the 228 possible students rostered in the 2017-2018 and 2018-2019 school years. The

2016–2017-year data contained 10 cases of students completing the STAMP 4Se assessment, however, it was removed because this was prior to the district's implementation of the MAP assessment and therefore these students did not have corresponding scores. Of the 172 remaining students, 139 had valid Reading MAP K-2 assessment scores. From these, 5 students with an 0525 IEP received literacy instruction in English prior to the end of 2nd grade and were therefore removed from the sample.

The total number of students who completed all four sections of the STAMP 4Se assessment (necessary for computing the composite score used in question one) and the Reading MAP K-2 assessment was thirty. The total number of students who completed the Reading MAP K-2 assessment and the STAMP 4Se reading section was 83. The total number of students who completed the Reading MAP K-2 assessment and the STAMP 4Se writing section was 73. The total number of students who completed the Reading MAP K-2 assessment and the STAMP 4Se listening section was 83. The total number of students who completed the Reading MAP K-2 assessment and the STAMP 4Se speaking section was 76.

There were several factors that led to the differences in the number of scores between the two assessments. The first is that the school district did not proctor all the sections of the STAMP 4Se assessment to each student in the two years the data represents. The reasons for this are twofold: first, these represent the first years of implementing the STAMP 4Se in the program and a soft implementation was used. And second, the district had technical questions about implementing the writing section via technology vs. paper pencil which were answered by piloting the assessment for the first 3 years. Specifically, approximately 50 students, who were randomly selected by the district, took the reading, listening, and speaking sections in the 2017-2018 data set. While the 2018-2019 data set contains 97 students who completed the reading,

54

listening, and speaking sections with 50 who were randomly selected by the district to also

complete the writing section. Additionally, not every student who was proctored the assessment

received a score for each section as represented by the lower number of students who received

scores in the speaking and writing sections. These students received an 'unratable' rating either

due to technical issues with recording or lack of production on that portion of the assessment.

Furthermore, the district has a process for identifying students who hit the 'ceiling' of the

Reading K-2 MAP assessment and moving them to the Reading 2-5 MAP assessment in the

winter and spring assessment windows of 2nd grade. Some of the students with valid STAMP

4Se scores were removed from the data set because they took the Reading 2-5 MAP assessment.

Findings

Research Question One (Composite)

Research question one addressed the relationship between L1 reading academic

performance and L2 language proficiency composite scores. A Spearman rank correlation was

used to address this question by creating two variables: L2 language proficiency (composite

score from STAMP 4Se) and L1 reading academic performance (Reading K-2 MAP RIT score).

The total number of cases from the archival data which met the study criteria was thirty.

A Spearman rank correlation has three assumptions. The first two assumptions are met

through the study design (variable type and paired observations) while the third assumption is

that the data is monotonic in nature. Preliminary analysis, which used a visual inspection of the

data on a scatterplot generated from SPSS determined that the data was, in fact, monotonic in

nature, meeting the third assumption. The scatterplot is located in Appendix D (Figure D1).

Data regarding the relationship of L2 language proficiency composite scores and L1

reading academic performance are summarized in Table 1. There was a statistically significant,

strong positive correlation between L2 language proficiency composite scores and L1 reading

academic performance, $r_s(28) = .821$, $p = < .001$. A correlational coefficient of .821 represents a

strong correlation (Cohen, 1988) between L2 language proficiency composite scores and L1

reading academic performance.

A post hoc power analysis was conducted using an exact test family and correlation:

bivariate normal model two-tailed statistical test as parameters. The correlational coefficient of

.821, an alpha level of .05, and sample size of 30 were used in the analysis. The post hoc analysis

revealed the power for this question exceeded .99 as the correlation coefficient of .821 exceeded

the upper critical r value of .36.

Table 1

Correlations Between L1 Reading RIT and L2 Proficiency Composite Scores

	N	Post Hoc Power	L1 Reading RIT	
			r_s value	p
L2 Composite Score	30	>.99	.821	<.001

Research Question Two (Individual)

Research question two looked at the relationship between L1 reading performance and

individual L2 proficiency modes (reading, listening, speaking, writing). A Spearman rank

correlation was used to address this question by creating two variables for each comparison. L1

reading academic performance (Reading K-2 MAP RIT) and the respective L2 proficiency rating

(subtest score from STAMP 4Se). Results from the Spearman correlations, as well as post hoc

power analysis for each comparison are summarized in Table 2.

L1 Reading and L2 Reading. The first comparison looked at L1 reading performance

and L2 reading proficiency. The total number of cases from the archival data which met the

study criteria was 83. During preliminary analysis, visual inspection of a scatterplot generated

from the data determined that the data was monotonic in nature, meeting the third assumption.

The scatterplot is located in Appendix D (Figure D2).

Data regarding the relationship of L2 reading proficiency subtest scores and L1 reading

academic performance are summarized in Table 2. There was a statistically significant, strong

positive correlation between L2 reading proficiency and L1 reading academic performance,

$r_s(81) = .810, p = < .001$. A correlational coefficient of .810 represents a strong correlation

(Cohen, 1988) between L2 reading proficiency subtest scores and L1 reading academic

performance.

A post hoc power analysis was conducted using an exact test family and correlation:

bivariate normal model two-tailed statistical test as parameters. The correlational coefficient of

.810, an alpha level of .05, and sample size of 83 were used in the analysis. The post hoc analysis

revealed the power for this question at 1 as the correlation coefficient of .810 exceeded the upper

critical r value of .21.

L1 Reading and L2 Listening. The next comparison was between L1 reading academic

performance and L2 listening proficiency. The total number of cases from the archival data

which met the study criteria was 83. During preliminary analysis, visual inspection of a

scatterplot generated from the data determined that the data was monotonic in nature, meeting

the third assumption. The scatterplot is located in Appendix D (Figure D3).

Data regarding the relationship of L2 listening proficiency subtest scores and L1 reading

academic performance are summarized in Table 1. There was a statistically significant, strong

positive correlation between L2 listening proficiency and L1 reading academic performance,

$r_s(81) = .560, p = < .001$. A correlational coefficient of .560 represents a strong correlation (Cohen, 1988) between L2 listening proficiency subtest scores and L1 reading academic performance.

A post hoc power analysis was conducted using an exact test family and correlation: bivariate normal model two-tailed statistical test as parameters. The correlational coefficient of .56, an alpha level of .05, and sample size of 83 were used in the analysis. The post hoc analysis revealed the power for this question exceeded .99 as the correlation coefficient of .56 exceeded the upper critical r value of .35.

L1 Reading and L2 Speaking. The next comparison was between L1 reading academic performance and L2 speaking proficiency. The total number of cases from the archival data which met the study criteria was 76. During preliminary analysis, visual inspection of a scatterplot generated from the data determined that the data was monotonic in nature, meeting the third assumption. The scatterplot is located in Appendix D (Figure D4).

Data regarding the relationship of L2 speaking proficiency subtest scores and L1 reading academic performance are summarized in Table 1. There was a statistically significant, small positive correlation between L2 speaking proficiency and L1 reading academic performance, $r_s(74) = .280, p = < .014$. A correlational coefficient of .280 represents a small correlation (Cohen, 1988) between L2 speaking proficiency subtest scores and L1 reading academic performance.

A post hoc power analysis was conducted using an exact test family and correlation: bivariate normal model two-tailed statistical test as parameters. The correlational coefficient of .28, an alpha level of .05, and sample size of 76 were used in the analysis. The post hoc analysis

revealed the power for this question at .69 as the correlation coefficient of .28 was marginally higher than the upper critical r value of .22.

 L1 Reading and L2 Writing. The final comparison was between L1 reading academic performance and L2 writing proficiency. The total number of cases from the archival data which met the study criteria was 73. During preliminary analysis, visual inspection of a scatterplot generated from the data determined that the data was monotonic in nature, meeting the third assumption. The scatterplot is located in Appendix D (Figure D5).

 Data regarding the relationship of L2 writing proficiency subtest scores and L1 reading academic performance are summarized in Table 1. There was a statistically significant, moderate positive correlation between L2 writing proficiency and L1 reading academic performance, $r_s(71) = .386, p = < .001$. A correlational coefficient of .386 represents a moderate correlation (Cohen, 1988) between L2 speaking proficiency subtest scores and L1 reading academic performance.

 A post hoc power analysis was conducted using an exact test family and correlation: bivariate normal model two-tailed statistical test as parameters. The correlational coefficient of .386, an alpha level of .05, and sample size of 73 were used in the analysis. The post hoc analysis revealed the power for this question at .92 as the correlation coefficient of .386 exceeded the upper critical r value of .23.

Table 2

Correlations Between L1 Reading RIT and Individual L2 Proficiency Ratings

	N	Post Hoc Power	L1 Reading RIT	
			r_s value	p
L2 Reading Proficiency	83	1	.810	<.001
L2 Listening Proficiency	83	>.99	.560	<.001
L2 Speaking Proficiency	76	.69	.280	.014
L2 Writing Proficiency	73	.92	.386	<.001

Organization of the Study

This chapter contained the results of the findings for the two research questions. Chapter 5 will provide a summary of the study, conclusions, and a discussion of the results. Recommendations for practice and for future study will also be included in Chapter 5.

Chapter 5

Summary, Conclusions, Discussion, and Recommendations

This chapter provides a summary of the study, conclusions, and a discussion of the results. The study ends with recommendations for both future practice and study.

Summary of the Study

Purpose

The purpose of this study is to explore the statistical nature of the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in foreign language immersion programs. Specifically, the interdependent relationship between L2 language proficiency levels and L1 reading academic performance in an environment where L1 is the majority language and L2 is the target language for instruction with students who have yet to receive formal reading instruction in the L1.

Research Questions

The research questions below were designed to address the problem statement:

1. To what extent and in what way is L1 reading performance associated with L2 language proficiency composite scores?

2. To what extent and in what way is L1 reading performance individually associated with L2 reading, listening, speaking, and writing proficiencies?

Literature Review

Bilingual education is distinctive from traditional models of instruction in the United States in that at its core, the framework presents an additive language model in addition to grade level content instruction. The Center for Applied Linguistics (CAL) describes high quality bilingual programs as being comprised of three pillars: Bilingualism & Biliteracy, Academic

Achievement, and Sociocultural Competence (Howard et al., 2018). The terms bilingual education or dual language programs are umbrella terms that describe a wide array of program models with differing target demographics, program outcomes, and instructional delivery models. These various program models are frequently lumped together in literature (Fortune & Tedick, 2015) and in order to account for that Continua Model of Biliteracy (Hornberger, 2002) was adopted to frame the literature used in the context of this study.

This study looks at a foreign language immersion model, also referred to as full-immersion or one-way immersion. One-way immersion models in the United States are additive language programs where the societal dominant language (English in the US) is every student's first language (L1), and students are instructed in the target language (L2 - not English) for a large part of the day. The current literature on dual language programs, as it relates to this study, can be organized into four categories: Academic achievement, L2 acquisition, L1 literacy, and the relationship of L1 literacy and L2 proficiency.

Historically, the results of research on academic achievement in full immersion programs have been positive (Bournot-Trites & Tellowitz, 2002). Additionally, empirical research has shown positive impacts of bilingualism on social and cognitive development (Hakuta & Diaz, 1985; Thomas & Collier, 2003) providing evidence for the positive impacts of bilingualism beyond the classroom.

Research on L2 acquisition across a broad swath of elementary bilingual programs generally shows learners increasing across the proficiency levels from grade to grade, specifically in the early grades as the threshold for moving from level to level is smaller (Potowski, 2007; Young, 2012; Ruiz-Funes, 2020, Fortune & Tedick, 2015; Padilla, Fan, Xu, & Silva, 2013; Fortune & Ju, 2017; T. W. Fortune & Ju, 2017; T. W. Fortune & Tedick, 2015;

Padilla et al., 2013; Potowski, 2007; Young, 2015). More specifically, research around L2 proficiency on Canadian full-immersion programs (similar in all aspects of the continua model to ABC school, other than the L2 is French instead of Spanish) have demonstrated that students typically acquire more complex L2 proficiency skills each subsequent year they are in the program (Genesee, 2008; Spada, 1997).

Broadly speaking, findings have consistently demonstrated that students' L1 literacy development lags monolingual peers until explicit instruction in L1 literacy takes place. Once students are instructed in L1 literacy, they not only catch-up to their monolingual peers, but they also surpass them in L1 literacy achievement (Fred Genesee, 1979; F. Genesee & Stanley, 1976; Lambert & Tucker, 1972; Swain & Lapkin, 1982; Turnbull et al., 2000). While these findings have relied on standardized tests and compared immersion students to their monolingual peers, empirical studies have not typically analyzed the specific effects of one-way immersion on L1 literacy development (Bialystock, 2007).

There is limited research into the interdependent development of literacy across the two languages of bilingual students in immersion programs as it relates to proficiency. There is even less when the Continua Model of Biliteracy is used to frame the program models as those similar to the ABC school's foreign language immersion model. Proctor et al. (2010) found Spanish and English reading comprehension to be closely related, however, the language demographics of the students in this study were dissimilar from those at the ABC school and listening comprehension was the only proficiency skill included in the study.

The second and more closely situated on the continua model of biliteracy school to that of ABC school looked at reading strategy use across and within the two languages as it related to the students' L2 proficiency in fourth and fifth graders in French immersion, in a model where

L1 literacy instruction begins in fourth grade. Frid and Friesen (2020) examined the L2 proficiency and reading comprehension strategy use across both French and English in a group of French immersion students. Students in this study scored significantly higher on vocabulary, word reading, and used more complex comprehension strategies in English than in French. Correlations between language proficiency and the complexity of the comprehension strategy used tended to be small. As students tended to use similar comprehension strategies across both languages, adding support to the linguistic interdependence hypothesis on the interdependent nature of reading comprehension strategies. Frid and Friesen (2020) also found within language predictors to be stronger than cross-language predictors which may lead one to question whether L2 proficiency has an impact on L1 literacy development. However, there are two important distinctions to keep in mind between this study and ABC school: first, the students had already received literacy instruction in L1 at the time of the study and second, the assessments labeled as language proficiency in this study (vocabulary and word reading) are probably more accurately described as discrete literacy skill assessments than language proficiency indicators.

Because the Continua Model of Biliteracy is, "...premised on a view of multilingualism as a resource" (Hornberger, 2004, p. 37) it situates itself nicely inside of Cummins' (1976, 2017) Linguistic Interdependence hypothesis and transfer theory as described by MacSwan et al. (MacSwan et al., 2017) which also frame bi and multilingualism as assets. This idea of interdependence across the two languages of bilinguals begins with Cummins' (1976) description of his "threshold hypothesis". Summarily, this hypothesis states that students must achieve a threshold in their L1 and L2 in order to receive the benefits of learning a second language. Conversely, if the threshold is not met, learning a second language can negatively impact L1 and L2 development, as well as content area subject matter knowledge. Evidence of a

threshold exists (Cárdenas-Hagan et al., 2007; Fred Genesee & Nicoladis, 2008; Jared et al., 2011; Siu & Ho, 2015). Evidence for the threshold hypothesis is particularly strong for programs similar to ABC school in the broad domains of the continua model of contexts, development, and content (Bournot-Trites & Tellowitz, 2002). Additionally, evidence of the existence of a threshold is persistent in research relating to late immersion programs where students' low L2 proficiency is attributed to low acquisition of complex subject matter presented in L2 (Marsh et al., 2000; Swain, 1996). Yet, two large problems exist with this initial framework: Defining and validating the threshold(s) has proven to be problematic for researchers (Bournot-Trites & Tellowitz, 2002) and the hypothesis, taken alone could lead to unintended negative consequences, specifically in programs that are dissimilar from ABC school in the areas of the Continua Model of contexts and content for speakers of the minority language and their communities (MacSwan et al., 2017) where an L1 threshold is used as a program gate.

This study addresses the first of these problems by exploring the statistical nature of the linguistic interdependence hypothesis (Cummins, 1976, 2017) in the spring of 2nd grade of students in foreign language immersion programs. Specifically, the interdependent relationship between L2 language proficiency levels and L1 reading academic performance in an environment where L1 is the majority language and L2 is the target language for instruction with students who have yet to receive formal reading instruction in the L1.

Methodology

This study used a non-experimental *ex post facto* design. The study used archival K-2 MAP reading assessment and STAMP 4Se data as indicators of L1 reading academic performance and L2 proficiency, respectively. The population for this study is students in a one-way Spanish immersion program prior to receiving L1 literacy instruction. The sample is

composed of students from a school district in the Midwest, which consists of approximately 24,000 students in 36 schools. Specifically, the sample group is composed of all 2nd grade students at the district's ABC Spanish Immersion school with valid STAMP 4Se and K-2 Reading MAP assessment scores in the spring of 2nd grade prior to the 2019-2020 school year. The total sample meeting the study's criteria is made up of 228 students.

Spearman rank correlation was used to address both research questions. First the relationship between L1 reading academic performance and L2 language proficiency composite scores was examined and then the relationship between L1 reading academic performance and each of the L2 language proficiency modes (reading, listening, speaking, writing) was examined. As both research questions are quantitative in nature, descriptive and inferential statistics were used.

Findings

Chapter 4 presented the findings of this study. The major findings include the following:

1. There was a statistically significant, strong positive correlation between L2 language proficiency composite scores and L1 reading academic performance, $r_s(28) = .821$, $p = < .001$.

2. There was a statistically significant, strong positive correlation between L2 reading proficiency and L1 reading academic performance, $r_s(81) = .810$, $p = < .001$.

3. There was a statistically significant, strong positive correlation between L2 listening proficiency and L1 reading academic performance, $r_s(81) = .560$, $p = < .001$.

4. There was a statistically significant, small positive correlation between L2 speaking proficiency and L1 reading academic performance, $r_s(74) = .280$, $p = < .014$.

5. There was a statistically significant, moderate positive correlation between L2 writing proficiency and L1 reading academic performance, $r_s(71) = .386$, $p = < .001$.

Conclusions

The following conclusions come from the statistical analyses of the data and findings of this study.

1. L2 proficiency composite are strongly correlated to L1 reading academic achievement.

2. L2 proficiency interpretive skills are strongly correlated to L1 reading academic achievement.

3. There is less correlation between L2 proficiency productive skills (speaking and writing) and L1 reading academic achievement.

Discussion

In additive foreign language immersion models, students learn content in L2. The successive, partner language first model of literacy instruction requires L1 literacy instruction to be deferred for a period of time (Howard et al., 2009). This delay in instruction is typically 2-4 years and exists even though state and district level requirements may require these programs to test their students' L1 literacy development. Programs and K-2 teachers in this situation are teaching and assessing in one language, with high stakes testing occurring in another language.

This study aimed to elaborate on the relationships of L2 proficiency and L1 literacy in the context described above, more narrowly defining the interdependent association of language and literacy skills across the two languages for the group of students in the study. Which, if any, L2 proficiency skills are associated with L1 reading academic performance? Strong correlations between specific L2 proficiency skills and L1 reading academic performance would give

programs, teachers, and parents in this situation indicators to look for between and before assessments that take place in L1.

The first conclusion based on the findings of this study establishes a strong relationship between L2 proficiency composite scores and L1 reading academic achievement. This is similar to findings across a broad swath of bilingual program models for students of varying language demographics (August & Shanahan, 2008; Fred Genesee et al., 2006; Krenca et al., 2020; National Research Council and Institute of Medicine et al., 1998; Slavin & Cheung, 2005). Establishing that this strong relationship also exists in this specific bilingual program model and context allowed the researcher to further explore the independent L2 proficiency modes' relationships with L1 reading academic performance. Although this is not a revelation considering the amount of research across all the variations of bilingual programming models that already exists supporting this relationship, it was particularly important to verify its existence in this context prior to L1 literacy instruction taking place. Establishing the existence of this relationship prior to formal L1 literacy instruction has implications for teachers and programs similar to the one in this study as it provides some impetus on programs and staff for student L1 development prior to formal instruction taking place.

The second conclusion that emerged from this study found that L2 proficiency interpretive skills (reading and listening) were strongly correlated to L1 reading academic performance. In fact, L2 reading proficiency proved to have the strongest correlation to L1 reading academic performance. Stronger than both L2 proficiency composite scores and L2 listening proficiency scores. The implications of this are unknown as the study design did not account for interdependent L2 proficiencies' impact on L1 reading academic achievement.

The strong association between L2 reading proficiency and L1 reading academic performance is consistent with research comparing L1 and L2 reading performances across a range of program types and contexts (Bruck & Genesee, 1995; Lee & Chen, 2019; Proctor et al., 2010; Frid & Friesen, 2020). This conclusion establishes similar findings for programs in this context, specifies the relationship as that of L2 reading proficiency and L1 reading academic performance, and further extends this relationship's existence prior to formal L1 literacy instruction. It's important to point out the distinction between L2 reading proficiency and L1 reading academic performance. L2 reading proficiency solely assessed a student's ability to interpret L2 text while L1 reading academic performance assessed a wide variety of discreet literacy skills (foundational skills, vocabulary acquisition, and comprehension tools accessible when reading informational and literature text). Reading's (2009) findings on the cross linguistic nature and specific accessibility and reliability of Spanish early literacy skills' positive impact on English application offers some explanation as to why these may be so strongly associated. However, this study not only establishes a similar relationship for one-way immersion programs, but it also does so prior to literacy instruction in English.

In the context of Frid and Friesen's (2020) research on French/English proficiency and reading comprehension, these results are surprising as Frid and Friesen (2020) found within language predictors to be stronger than cross-language predictors. However, Reading's (2009) findings seem to directly contradict Frid and Friesen (2020). If we were to ignore the differing languages and contexts, this study's results would add support to Reading's (2009) work; however, I think there is something more complex happening. It could be that L2 reading proficiency is strongly associated with L1 reading academic performance at the word level for students in this context. Meaning that learning letter sounds, left to right, segmenting and

blending are specifically strong cross-linguistic skills, but that the more abstract work of accessing reading comprehension tools is less cross-linguistic or, possibly, could be beholden to a threshold by the L1 or L2. This would add some clarity to the language interdependence hypothesis' (Cummins, 2008, 2017; García, 2011; MacSwan et al., 2017) idea that multilingual (and bilingual) speakers have a single, integrated language system in which some resources are shared, and some are discreet.

The third conclusion of this study established a moderate correlation between L2 writing proficiency and L1 reading academic performance and a weak correlation between L2 speaking proficiency and L1 reading academic performance. Both L2 productive skills tend to lag behind L2 receptive skills (Figure 2) and students' L1 literacy skills in foreign language programs with successive literacy instruction tend to lag behind their peers in traditional programs (Fred Genesee, 1979; F. Genesee & Stanley, 1976; Lambert & Tucker, 1972; Swain & Lapkin, 1982; Turnbull et al., 2000). The interesting aspect to these results is that these two lags do not appear to be related to each other. Again, this provides some clarity for programs similar to the one in this study. As programs work to develop or intervene on both L2 productive skills and L1 literacy development in students, it's critical to understand that these two areas do not seem be related to each other. The implication being each of the two areas require a separate response by the program or staff.

Recommendations

Recommendations for Practice

1. Programs similar to the one in this study can utilize L2 proficiency composite and interpretive scores in K-2 students as a reliable screening tool prior to L1 literacy instruction taking place.

2. Programs similar to the one in this study should present L2 proficiency results in addition to L1 reading academic performance results to guardians when communicating student progress, particularly in instances prior to formal L1 instruction taking place.

3. Teachers in programs similar to the one in this study can use L2 proficiency composite and interpretive indicators as a way to make instructional decisions and monitor growth between and before L1 reading high stakes assessment periods.

Recommendations for Future Study

1. A study should be conducted that examines the directional nature of L1 reading academic performance and L2 proficiency in a context similar to the one used in this study.

2. A study should be conducted that examines the predictive nature of L2 proficiency on L1 reading academic performance in a context similar to the one used in this study.

3. A study should be conducted on the relationship of L2 proficiency composite scores and independent modes to discreet L1 reading skills in a context similar to the one used in this study.

4. A study should be conducted after formal L1 literacy instruction takes place in a similar context to determine if L1 literacy instruction impacts the association of L2 proficiencies and L1 reading academic performance.

5. Further study is needed to analyze the interdependent impact L2 proficiency modes have on L1 literacy development in similar contexts.

71

References

ACTFL. (n.d.). *The ACTFL assessment of performance toward proficiency in languages: AAPPL measure FAQs*. https://www.actfl.org/assessment-professional-development/assessments-the-actfl-testing-

ACTFL. (2012). *ACTFL proficiency guidelines [Electronic version]*. https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf

Amato, S. (2011). *Variables that predict world language achievement in one new jersey high school as measured by the STAMP test* (Order No. 3472684). Available from ProQuest Dissertations & Theses Global. (890227969). https://usd.idm.oclc.org/login?url=https://www-proquest-com.usd.idm.oclc.org/dissertations-theses/variables-that-predict-world-language-achievement/docview/890227969/se-2?accountid=14750

August, D., & Shanahan, T. (2008). *Developing reading and writing in second-language learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*. Taylor & Francis.

Avant Assessment. (2015). *STAMP 4S benchmark and rubric guide*. Eugene, OR: Avant Assessment.

Avant Assessment. (2017). *Report on Avant STAMP inter rater reliability and Avant's rating process*. https://d3itqxtdxl1nz0.cloudfront.net/pdfs/Report-on-Avant-STAMP-4S-Inter-Rater-Reliability-and-Avants-Rating-Process-3.pdf

Avant Assessment. (2020). *Avant STAMP results: National averages 2018–2019*. https://static1.squarespace.com/static/5b5110cda9e0283700e2a0fa/t/5e5fefa4c855db313304d60e/1583345582441/STAMP+2018_2019+National+Averages_withcopyright.pdf

Blanton, M. (2015). *The effect of two foreign language teaching approaches, communicative language teaching and teaching proficiency through reading and storytelling, on motivation and proficiency for spanish III students in high school* (Order No. 3708358). Available from ProQuest Dissertations & Theses Global. (1697922464). https://usd.idm.oclc.org/login?url=https://www-proquest-

com.usd.idm.oclc.org/dissertations-theses/effect-two-foreign-language-teaching-
approaches/docview/1697922464/se-2?accountid=14750

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning, 57*, 45–77.

Bjorklund-Young, A., & Borkoski, C. (2016). *Do formative assessments influence student learning? Research on i-Ready and MAP*. Institute for Education Policy: Johns Hopkins School of Education.

Borsato, G., & Lindholm-Leary, K. (2006). Academic Achievement. In F. Genesee, B. Saunders, D. Christian, & K. Lindholm-Leary (Eds.), *Educating English language learners* (pp. 157–179). Cambridge University Press.

Bournot-Trites, B. M., & Tellowitz, U. (2002). *Report of current research on the effects of second language learning on first language literacy skills*. Atlantic Provinces Educational Foundation. https://ab.cpf.ca/wp-content/blogs.dir/1/files/apef_report.pdf

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, 22(2), 307–324.

Burkhauser, S., Steele, J. L., Li, J., Slater, R. O., Bacon, M., & Miller, T. (2016). Partner-language learning trajectories in dual-language immersion: Evidence from an urban district. *Foreign Language Annals*, *49*(3), 415-433. http://dx.doi.org.usd.idm.oclc.org/10.1111/flan.12218

CAL. (2000). *Rating scales for cal oral proficiency exam (COPE) and student oral proficiency assessment (SOPA)*: Revised 2000. Center for Applied Linguistics (CAL).

CAL. (2016). *TWI glossary: Glossary of terms related to Dual Language/TWI in the United States*. Cal.org. http://www.cal.org/twi/glossary.htm

Cárdenas-Hagan, E., Carlson, C. D., & Pollard-Durodola, S. D. (2007). The cross-linguistic transfer of early literacy skills: the role of initial L1 and L2 skills and language of instruction. *Language, Speech, and Hearing Services in Schools, 38*(3), 249–259.

CASLS. (2012a). *STAMP 4Se (Standards-based Measurement of Proficiency - 4 Skills Elementary) Technical Report* (J. Snyder (ed.)). Center for Applied Second Language Studies.

https://d3itqxtdxl1nz0.cloudfront.net/pdfs/avant-stamp4Se-technical-document.pdf

CASLS. (2012b). *STAMP 4S (Standards-based Measurement of Proficiency - 4 Skills) Spanish Technical Report* (J. Snyder (ed.)). Center for Applied Second Language Studies. https://static1.squarespace.com/static/5b5110cda9e0283700e2a0fa/t/5e8616dc892cf236f2e95c7b/1585845989464/spanish-avant-stamp-technical-document.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: NY: Psychology Press.

Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practice, 2*(1), 1–20.

Cummins, J. (1976). *The influence of bilingualism on cognitive growth: a synthesis of research findings and explanatory hypotheses*. https://eric.ed.gov/?id=ED125311

Cummins, J. (2008). Teaching for transfer: Challenging the two solitudes assumption in bilingual education. *Encyclopedia of Language and Education, 5*, 65–75.

Cummins, J. (2017). Teaching for transfer in multilingual school contexts. *Bilingual and Multilingual Education, 3*, 103–115.

Davin, K. J., Rempert, T. A., & Hammerand, A. A. (2014). Converting data to knowledge: One district's experience using large-scale proficiency assessment. *Foreign Language Annals*, *47*(2), 241-260. http://dx.doi.org.usd.idm.oclc.org/10.1111/flan.12081

D'Angelo, N., Hipfner-Boucher, K., & Chen, X. (2017). Predicting growth in English and French vocabulary: The facilitating effects of morphological and cognate awareness. *Developmental Psychology, 53*(7), 1242–1255.

De Jong, E. J., & Bearse, C. I. (2011). The same outcomes for all? High school students reflect on their two-way immersion program experiences. In D. Tedick, D. Christian, & T. Fortune (Eds.), *Immersion education: practices, policies, possibilities* (pp. 104–122). Multilingual Matters Bristol.

Dicks, J., & Genesee, F. (2017). Bilingual education in Canada. *Bilingual and Multilingual Education*, 453–467.

Falsgraf, C. (2009). The ecology of assessment. *Language Teaching, 42*(4), 491-503.

> http://dx.doi.org.usd.idm.oclc.org/10.1017/S0261444808005570

Faulkner-Bond, M., Waring, S., Forte, E., Crenshaw, R. L., Tindle, K., & Belknap, B. (2012). *Language Instruction Educational Programs (LIEPs): A review of the foundational literature*. Office of Planning, Evaluation and Policy Development, US Department of Education. https://eric.ed.gov/?id=ED531981

Fortune, T. W., & Ju, Z. (2017). Assessing and exploring the oral proficiency of young Mandarin immersion learners. *Annual Review of Applied Linguistics, 37*, 264–287.

Fortune, T. W., & Tedick, D. J. (2015). Oral proficiency assessment of English-proficient k--8 Spanish immersion students. *The Modern Language Journal, 99*(4), 637–655.

Frid, B., & Friesen, D. C. (2020). Reading comprehension and strategy use in fourth- and fifth-grade French immersion students. *Reading and Writing, 33*(5), 1213–1233.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction*. Boston: Pearson-Allyn & Bacon.

García, O. (2011). *Bilingual education in the 21st century: A global perspective*. John Wiley & Sons.

Gay, L. R., & Airasian, P. W. (2003). *Educational research: competencies for analysis and applications* (7th ed.). Merrill/Prentice Hall.

Genesee. (2008). What do we know about bilingual education for majority-language students? In Bhatia, T. K., & Ritchie, W. C. (Ed.), *The Handbook of Bilingualism* (pp. 547–576).

Genesee, F. (1979). Acquisition of reading skills in immersion programs. F*oreign Language Annals, 12*(1), 71–77.

Genesee, F. (2007). French immersion and at-risk students: A review of research evidence. *The Canadian Modern Language Review, 63*(5), 655–687.

Genesee, F., & Lindholm-Leary, K. (2010). Improving education for english learners: research-based approaches. In California. Department of Education (Ed.), *Alternative educational programs for English language learners* (pp. 323–382). California Department of Education.

Genesee, F., & Lindholm-Leary, K. (2013). Two case studies of content-based language education. *Journal of Immersion and Content-Based Language Education, 1*(1), 3–33.

Genesee, F., Lindholm-Leary, K., Christian, D., Saunders, W., & Saunders, B. (2006). *Educating English language learners: A synthesis of research evidence*. Cambridge University Press.

Genesee, F., & Nicoladis, E. (2008). Bilingual first language acquisition. In *Blackwell Handbook of Language Development* (pp. 324–342). Blackwell Publishing Ltd.

Genesee, F., & Stanley, M. H. (1976). The development of English writing skills in French immersion school programs. *Canadian Journal of Education / Revue Canadienne de L'éducation, 1*(3), 1–17.

Good, P. I., & Hardin, J. W. (2003). *Common errors in statistics (and how to avoid them)*. Wiley-Interscience.

Gort, M. (2019). Developing bilingualism and biliteracy in early and middle childhood. *Urbana, 96*(4), 229–243.

Hakuta, K., & Diaz, R. M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. *Children's Language, 5*, 319–344.

Hornberger, N. H. (2002). Multilingual language policies and the continua of biliteracy: An ecological approach. *Language Policy, 1*(1), 27–51.

Hornberger, N. H. (2004). The continua of biliteracy and the bilingual Educator: Educational linguistics in practice. *International Journal of Bilingual Education and Bilingualism, 7*(2-3), 155–171.

Housen, A. (2002). Processes and outcomes in the European schools' model of multilingual education. *Bilingual Research Journal, 26*(1), 45–64.

Howard, E. R., Sugarman, J., Christian, D., Lindholm-Leary, K. J., & Rogers, D. (2018). *Guiding principles for dual language education*. Center for Applied Linguistics Washington, DC.

Howard, E., Sugarman, J., Rogers, D. & Olague, N. (2009). *Program models and the language of initial literacy in two-way programs*. Center for Applied Linguistics. http://www.cal.org/twi/initialliteracy.pdf

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2011). Early predictors of biliteracy

development in children in French immersion: A 4-year longitudinal study. *Journal of Educational Psychology, 103*(1), 119–139.

Johnson, R. K. (1997). *The Hong Kong education system*. Immersion Education: International Perspectives, 171. https://books.google.com/books?hl=en&lr=&id=m6gagHDTsCEC&oi=fnd&pg=PA171&dq=Johnson+R+K+(1997)+The+Hong+Kong+education+system+Late+immersion+under+stress&ots=wyjL0uzfv7&sig=yNPblpwoBNJ_0s4vbziVWdJhkd4

Krashen, S. D. (1996). *Under attack: The case against bilingual education*. Language Education Assocs.

Krenca, K., Gottardo, A., Geva, E., & Chen, X. (2020). English phonological specificity predicts early French reading difficulty in emerging bilingual children. *Annals of Dyslexia, 70*(1), 27–42.

Lambert, W. E., & Tucker, G. R. (1972). *Bilingual education of children: The St. Lambert experiment*. https://eric.ed.gov/?id=ED082573

Lee, K., & Chen, X. (2019). An emergent interaction between reading fluency and vocabulary in the prediction of reading comprehension among French immersion elementary students. *Reading and Writing, 32*(7), 1657–1679.

Lindholm-Leary, K., & Howard, E. R. (2008). Language development and academic achievement in two-way immersion programs. In T. Fortune & D. Tedick (Eds.), *Pathways to bilingualism: Evolving perspectives on immersion education* (Vol. 66, p. 177). MULTILINGUAL MATTERS LTD.

Macswan, J. (2000). The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22*(1), 3–45.

MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal, 54*(1), 167–201.

MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program. *Bilingual Research Journal, 29*(3), 653–678.

Macswan, J., & Rolstad, K. (2005). Modularity and the facilitation effect: Psychological mechanisms of

transfer in bilingual students. *Hispanic Journal of Behavioral Sciences, 27*(2), 224–243.

MacSwan, J., Thompson, M. S., Rolstad, K., McAlister, K., & Lobo, G. (2017). Three theories of the

    effects of language education programs: An empirical evaluation of bilingual and English-only

    policies. *Annual Review of Applied Linguistics, 37*, 218–240.

Marsh, H., Hau, K.-T., & Kong, C.-K. (2000). Late immersion and language of instruction in Hong Kong

    high schools: Achievement growth in language and nonlanguage subjects. *Harvard Educational*

    Review, 70(3), 302–347.

Mishra, R., & Stainthorp, R. (2007). The relationship between phonological awareness and word reading

    accuracy in Oriya and English: A study of Oriya-speaking fifth-graders. *Journal of Research in*

    Reading, 30(1), 23–37.

Montecel, M. R., & Danini, J. (2002). Successful bilingual education programs: development and the

    dissemination of criteria to identify promising and exemplary practices in bilingual education at the

    national level. In *Bilingual Research Journal* (Vol. 26, Issue 1, pp. 1–21).

    https://doi.org/10.1080/15235882.2002.10668696

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. Sage Publications.

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Division of

    Behavioral and Social Sciences and Education, Board on Science Education, Board on Children,

    Youth, and Families, & Committee on Fostering School Success for English Learners: Toward New

    Directions in Policy, Practice, and Research. (2017). *Promoting the educational success of children*

    and youth learning English: promising futures. National Academies Press.

National Center on Intensive Intervention. (2019). *Academic screening tools chart*. American Institutes

    for Research. https://charts.intensiveintervention.org/ascreening#disagg

National Research Council and Institute of Medicine, Division of Behavioral and Social Sciences and

    Education, Commission on Behavioral and Social Sciences and Education, & Committee on

    Developing a Research Agenda on the Education of Limited-English-Proficient and Bilingual

    Students. (1998). *Educating language-minority children* (August, D. & Hakuta (ed.)). National

Academies Press.

NWEA. (2015). *Linking the smarter balanced assessments to NWEA MAP assessments*. ERIC

Clearinghouse.

NWEA. (2017). *Common Core State Standards MAP Growth K-2 Reading & Mathematics content*.

Northwest Evaluation Association.

Padilla, A. M., Fan, L., Xu, X., & Silva, D. (2013). A Mandarin/English two-way immersion program:

Language proficiency and academic achievement. In *Foreign Language Annals* (Vol. 46, Issue 4, pp.

661–679). https://doi.org/10.1111/flan.12060

Potowski, K. (2007). Characteristics of the Spanish grammar and sociolinguistic proficiency of dual

immersion graduates. *Spanish in Context, 4*(2), 187–216.

Proctor, C. P., August, D., Snow, C., & Barr, C. D. (2010). The interdependence continuum: A

perspective on the nature of Spanish–English bilingual reading comprehension. *Bilingual Research

Journal, 33*(1), 5–20.

Reading, S. (2009). Differential effects of French and Spanish immersion education on English literacy

skills. *Bilingual Research Journal, 31*(1-2), 115–145.

Ruiz-Funes, M. T. (2020). The development of oral ability in L2 Spanish by English-home language

learners in a TWI program. *Foreign Language Annals, 53*(1), 96–127.

Santos, V. D. (2019). *The development of a STAMP Test: Support for test validity*. Avant Assessment

LLC. Retrieved from

https://static1.squarespace.com/static/5b5110cda9e0283700e2a0fa/t/6042be5bc5ca9556c2bfb852/16

14986847844/Development-of-a-STAMP-Test-Nov_1_2019.pdf

Shum, K. K.-M., Ho, C. S.-H., Siegel, L. S., & Au, T. K.-F. (2016). First-language longitudinal predictors

of second-language literacy in young L2 learners. *Reading Research Quarterly, 51*(3), 323–344.

Siu, C. T.-S., & Ho, C. S.-H. (2015). Cross-language transfer of syntactic skills and reading

comprehension among young Cantonese-English bilingual students. *Reading Research Quarterly,

50*(3), 313–336.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for

    English language learners. *Review of Educational Research, 75*(2), 247–284.

Spada, N. (1997). Form-Focussed instruction and second language acquisition: A review of classroom

    and laboratory research. *Language Teaching, 30*(2), 73–87.

Swain, M. (1996). Discovering successful second language teaching strategies and practices: From

    programme evaluation to classroom experimentation. *Journal of Multilingual and Multicultural*

    Development, 17(2-4), 89–104.

Swain, M., & Lapkin, S. (1982). Evaluating bilingual education: A Canadian case study. *Multilingual*

    Matters 2. Multilingual Matters Ltd.: Clevedon, England.

Thomas, W. P., & Collier, V. P. (2003). The multiple benefits of dual language. Educational *Leadership:*

    Journal of the Department of Supervision and Curriculum Development, N.E.A, 61(2), 61–61.

Thomas, W. P., & Collier, V. P. (2012). *Dual language education for a transformed world*. Dual

    Language Education of New Mexico/Fuente Press.

Thompson, C. B., & Panacek, E. A. (2006). Research study designs: Experimental and quasi-

    experimental. *Air Medical Journal, 25*(6), 242–246.

Thompson, C. B., & Panacek, E. A. (2007). Research study designs: Non-experimental. *Air Medical*

    Journal, 26(1), 18–22.

Thompson, K. D. (2017). *English learners' time to reclassification: An analysis. Educational Policy,*

    31(3), 330–363.

Turnbull, M., Hart, D., & Lapkin, S. (2000). *French immersion students' performance on Grade 3*

    provincial tests: Potential impacts on program design. Final Report Submitted to Education Quality

    and Accountability Office (EQAO), Ottawa: OISE-UT, Modern Language Centre.

    http://on.cpf.ca/wp-content/blogs.dir/1/files/eqao-final-report-2001.pdf

Uchikoshi, Y., & Marinova-Todd, S. H. (2012). Language proficiency and early literacy skills of

    Cantonese-speaking English language learners in the U.S. and Canada. In *Reading and Writing* (Vol.

    25, Issue 9, pp. 2107–2129). https://doi.org/10.1007/s11145-011-9347-2

Valdez, V. E., Freire, J. A., & Delavan, M. G. (2016). The gentrification of dual language education. *The Urban Review, 48*(4), 601–627.

Vik, T. L. (2018). *A comparison of fourth-grade students taught in English-only classrooms and Spanish-only classrooms in the same district* [University of South Dakota]. https://search.proquest.com/openview/fb13607178a0febca46f1f3da5bd2872/1?pq-origsite=gscholar&cbl=18750&diss=y

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2002). *Mathematical statistics with applications* (6th ed.). Duxbury.

Wei, L. (2018). A quality Chinese-English bilingual education program: defining success. *Journal of Language Teaching and Research, 9*(6), 1191–1198.

Young, A. (2015). *Student oral proficiency in grade three Spanish immersion: Linguistic diversity, student interaction and differentiated scaffolding*. Available from ProQuest Dissertations & Theses Global. (1725140232). Retrieved from https://usd.idm.oclc.org/login?url=https://search-proquest-com.usd.idm.oclc.org/docview/1725140232?accountid=14750

Appendix A

The Continua Model of Biliteracy

traditionally **less** powerful ←——————→ traditionally **more** powerful

Contexts of biliteracy

micro ←——————————————————→ macro

oral ←——————————————————→ literate

bi(multi)lingual ←————————————————→ monolingual

Development of biliteracy

reception ←——————————————————→ production

oral ←——————————————————→ written

L1 ←——————————————————→ L2

Content of biliteracy

minority ←——————————————————→ majority

vernacular ←——————————————————→ literary

contextualized ←——————————————————→ decontextualized

Media of biliteracy

simultaneous exposure ←——————→ successive exposure

dissimilar structures ←——————————→ similar structures

divergent scripts ←——————————————→ convergent scripts

District Research Proposal

Application To Conduct Research

Name: Andrew McKay   Date: 5/16/2021

School: XXXXXX

Title of Project: The Relationship of L2 Proficiency and L1 Reading Academic Performance Prior to Formal L1 Literacy instruction in Foriegn Language Immersion Settings

Research Questions:

1. What relationships exist between L2 language proficiency and L1 reading academic performance?

2. To what extent do L2 proficiency subtest scores in reading, speaking, listening, and writing predict L1 reading academic performance?

Methodology:

Ex post facto design, using existing MAP and STAMPse assessment data from 2nd grade students prior to spring of 2020. Quick notes: using spring of 2nd grade data is interesting because it is the point at which students have had all of their formal initial literacy instruction in Spanish, right before they receive formal English instruction. Also, pre 2020 to avoid any questions of COVID instruction impacting the data set.

The STAMPse is used for 'language proficiency' while the MAP is used for 'reading academic performance' in the research questions.
Question 1 will use a Spearman rank correlation to determine if/to what degree these two data sets are related.
Question 2 will use multiple regression to see if one of the STAMPse subcategories is strongly tied to the overall RIT score on the MAP assessment.

Plan for data collection:
After the Institutional Review Board (IRB) approval, the researcher will submit an application for Research Approval to the assistant superintendent of academic achievement of the ABC school district. Once this approval is received, the researcher will submit a request to the district's assessment management team. The district's assessment management team will locate, obtain and organize the ex post facto data that met the criteria for this study and share it with the researcher. During this process, the district's assessment management team will de-identify the data, generating random identification numbers to represent each student. Thus, the researcher will never have access to identifying student information. Additionally, the district's assessment

management team will password protect the file before giving it to the researcher. Only the district's management team and the researcher will have access to the password protected file.

Benefits to School/District:
Boldly paraphrasing and simplifying my chapter 2: We know that for students in immersion programs, both locally and nationally, L1 literacy is initially behind and then quickly closes. Nationally, this happens by 5th grade. Locally, we tend to see the closure happen in third grade after students receive English literacy instruction in our ESS program. XXX's 2018 dissertation on English writing at the end of 4th demonstrated hard data for this closure taking place. This phenomenon is also widely written about in theoretical research: L1 literacy development in immersion programs initially lags – something happens – and then by upper elementary these students outperform their monolingual peers, even when controlling for a range of variables (SES, demographics, instructional models). My hope is to get a glimpse into that "something happens" point, right before students start literacy instruction in English. Is their Spanish proficiency related to their English literacy at this point? If so, which mode of communication is most closely related (speaking, listening, reading, or writing)? Understanding this relationship a little better has implications for the teachers, administrators, and parents of students in the Spanish Immersion program.

Attach, if applicable, survey instruments, letters to participants, consent forms, IRB approval, Chapter 1, Chapter 3.

IRB Approval
Chapter 1
Chapter 3

Prior to data collection, submit application and all appropriate documents to:

XXXX

XXXX

Appendix C

IRB Approval

UNIVERSITY OF
SOUTH DAKOTA

Date:
May 7, 2021

The University of South Dakota
414 E. Clark Street
Vermillion, SD 57069

PI: Susan Gapp
Student PI: Andrew McKay
Re: Initial - IRB-21-109, *The Relationship of L2 Proficiency and L1 Reading Academic Performance Prior to Formal L1 Literacy Instruction in Foreign Language Immersion Settings*

The University of South Dakota Institutional Review Board has rendered the decision below for this study. The approval is effective starting May 7, 2021 and will expire on May 7, 2022.

Decision: Approved
Category: Expedited 5
Associated Approvals: Existing Data Analysis, Research Involving Children, Waiver of the Process of Parental Consent

Dear Susan Gapp,

The study submission for the proposal referenced above has been reviewed and approved according to the procedures of the University of South Dakota Institutional Review Board.

Your study has been granted a waiver of the process of informed consent.

Prior to initiation, promptly report to the IRB, any proposed updates/amendments (e.g., protocol amendments/revised informed consents) in previously approved human subjects research activities.

Any research-related injuries (physical or psychological), adverse side effects, or other unexpected problems encountered during the conduct of this research study needs to be reported to the IRB within **5 days** of notification of the occurrence.

Any modifications to the approved study must be submitted for review through Cayuse IRB. All approval letters and study documents are located within the study details in Cayuse IRB.

You have approval for this project through May 7, 2022. When this study is completed please submit a closure form through Cayuse. If the study is to last longer than one year, a continuation form needs to be submitted through Cayuse at least **14 days** prior to the expiration of this study.

If you have any questions, please contact: humansubjects@usd.edu or (605)658-3743.

Sincerely,

The University of South Dakota Institutional Review Board

Ann Waterbury

Ann Waterbury. M.B.A
Director, Office of Human Subjects
University of South Dakota
(605) 658-3767

Appendix D

Figure D1

Scatterplot of L2 Proficiency Composite Scores and L1 Reading Academic Performance
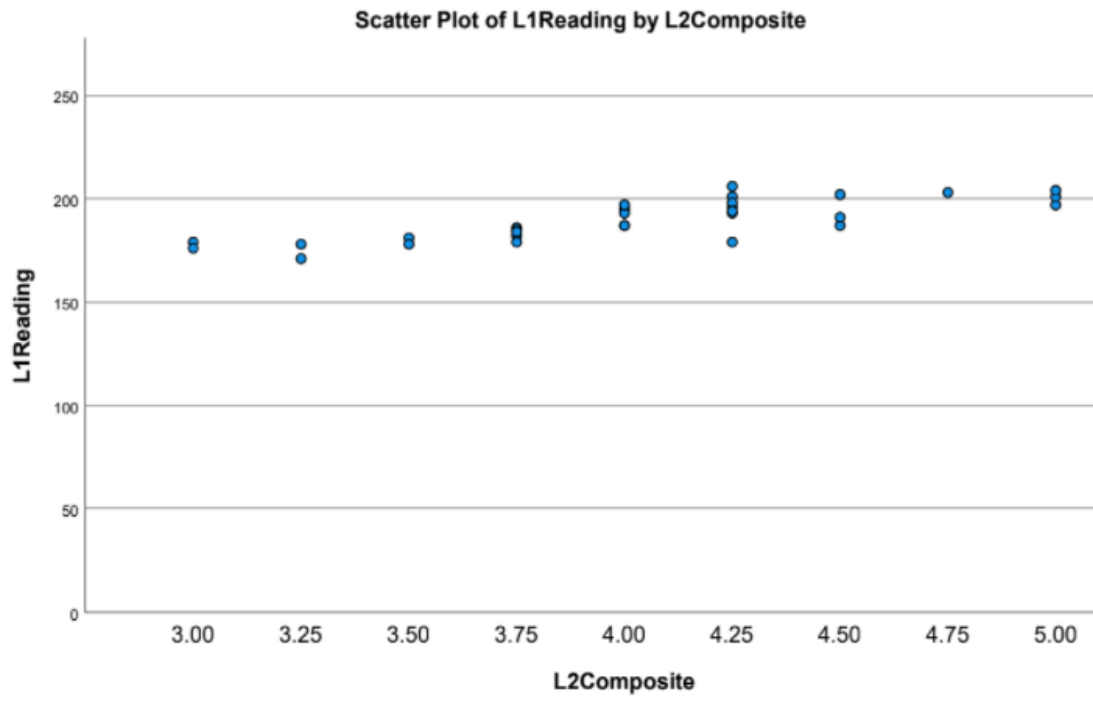

Scatter Plot of L1Reading by L2Composite

Figure D2

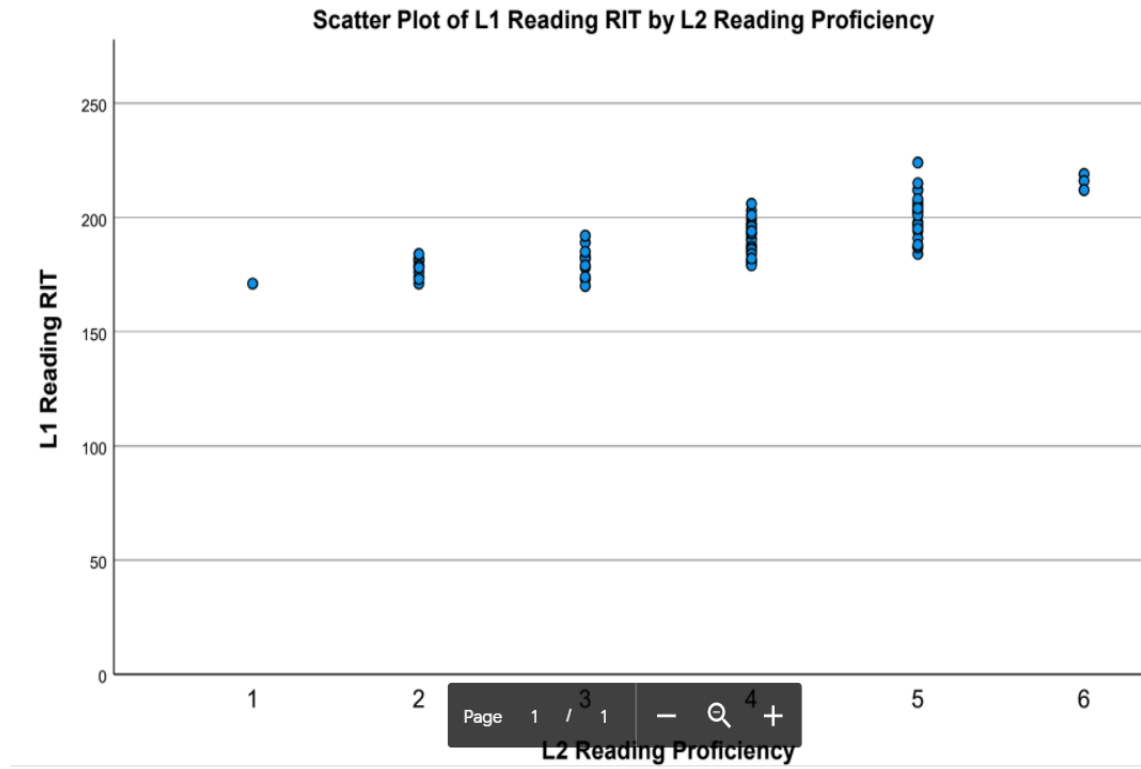*Scatterplot of L2 Reading Proficiency and L1 Reading Academic Performance*

Figure D3

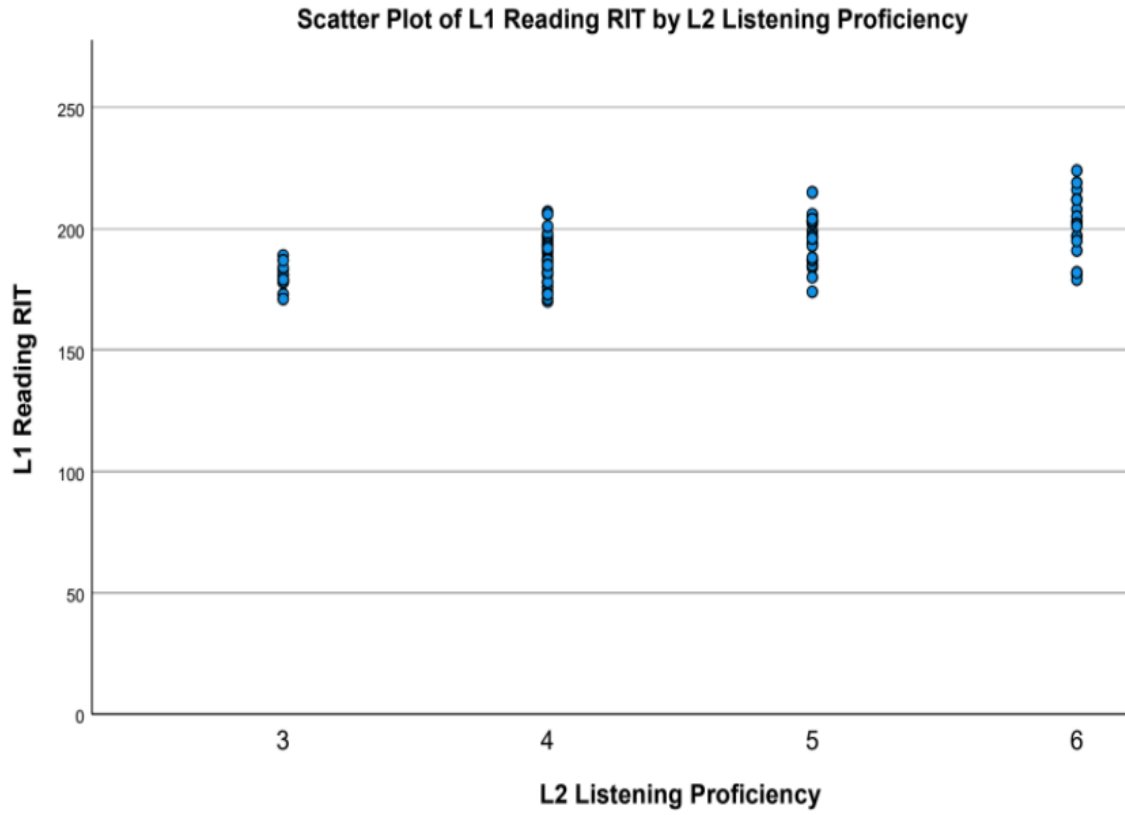*Scatterplot of L2 Listening Proficiency and L1 Reading Academic Performance*

Figure D4

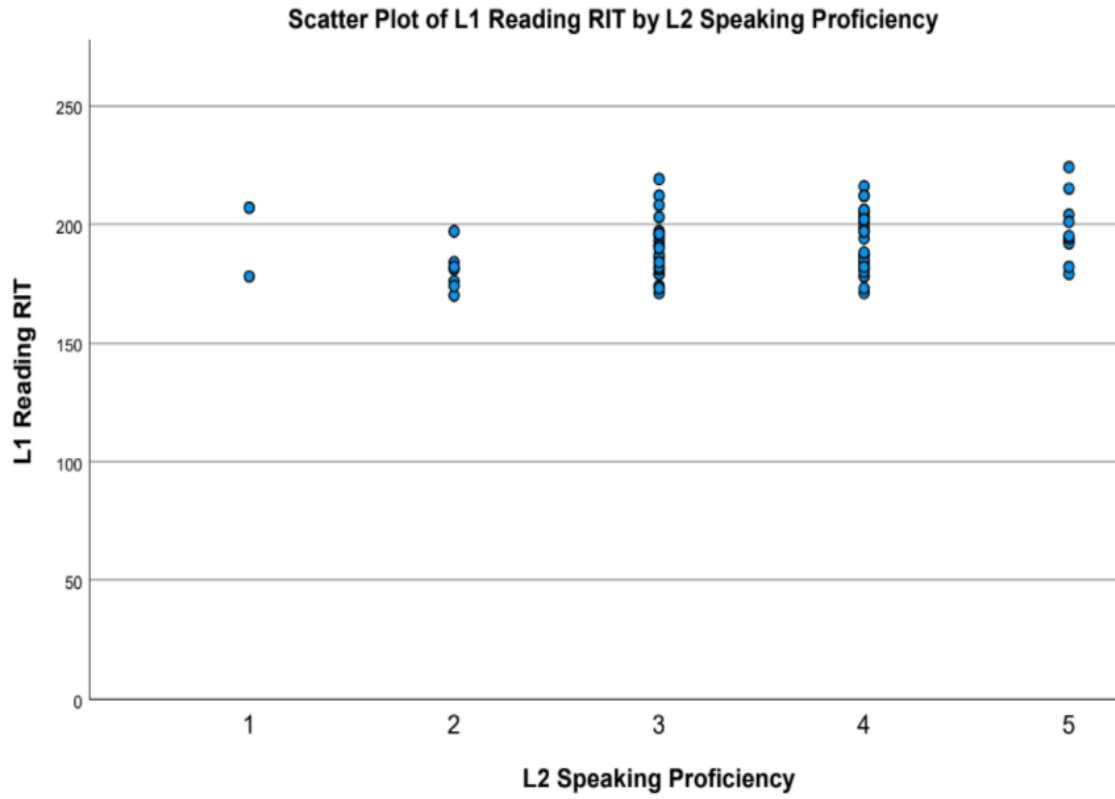*Scatterplot of L2 Speaking Proficiency and L1 Reading Academic Performance*

Figure D5

Scatterplot of L2 Writing Proficiency and L1 Reading Academic Performance

Scatter Plot of L1Reading by L2Writing