How Does Learning L2-English in an Immersion Context Compare to Learning in a Majority-Language Context? Comparing Adolescent L1-Mandarin Speakers in Taiwan to Peers in Western Canada

by

Brian Vincent Rusk

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

Doctor of Philosophy

Department of Linguistics
University of Alberta

Examining Committee:

Johanne Paradis, Supervisor
Juhani Järvikivi, Supervisory Committee
Elena Nicoladis, Supervisory Committee
Stephanie Archer, Examiner
Katherine Demuth, External Examiner

ABSTRACT

This dissertation reports three independent studies undertaken to better understand the second language (L2) English abilities that speakers of first-language (L1) Mandarin develop in language immersion classrooms. Children learning an L2 as the majority language of their community and schooling often attain L2 proficiency that closely parallels that of L1 speakers of the target L2. In contrast, children who learn a non-community L2 through an immersion education program are noted for having non-target L2 abilities for both morphosyntax and vocabulary. Though the L2 abilities of young learners in the immersion and community-L2 contexts have been compared to an L1-based target language standard, little is known about how the L2 abilities these learners compare directly to each other. In addition, it is unclear how differences in these learners' out-of-school L2 environments may impact resulting L2 abilities. In order to address this under-researched area, this dissertation examined the L2 morphosyntactic and receptive vocabulary abilities of 37 adolescent L1-Mandarin participants in an English immersion program in Taiwan. Across the three studies, the immersion learners were compared to a group of adolescent L1-English speakers and a group of L1-Mandarin community-L2 learners, both comprised of participants living in western Canada. Individual studies also

compared the L2 groups' out-of-school L2 environments and examined whether variations in their L2 input and experience were associated with individual differences in L2 ability.

The first two studies assessed ability with L2 morphosyntax. The three participant groups' receptive ability with a range of grammatical morphemes was compared using a grammaticality judgment task, and a task combining eye-tracking with a picture-decision task compared the immersion learners' ability to comprehend plural-singular marking to the L1-English group. In each case, differences in English ability were found between the immersion learners and the comparison groups. In the third study, receptive vocabulary was gauged using an age-standardized assessment task for monolingual English speakers, finding the immersion learners had significantly smaller L2 vocabularies than the community-L2 participants. The third study also compared the out-of-school L2 environments of the immersion learners to their community-L2 peers, showing the groups had similar access to L2 input, but differed in opportunities for interactional L2 use. In the first and third studies, differences in L2 input and experience were found to impact L2 morphosyntactic ability and receptive vocabulary ability, but that these factors appeared to have a much greater impact on vocabulary as opposed to morphosyntactic ability. Findings suggest that while development of L2 receptive vocabulary is ongoing for these L2 learners, L2 development for the studied grammatical morphemes has likely plateaued. Results indicate there are substantial differences between the L2 abilities of learners in the immersion or community-L2 context and that out-of-school L2 input and experience can have an impact on abilities in either context.

PREFACE

The studies undertaken for this dissertation received ethics approval from the University of Alberta Research Ethics Board under the project named "English language skills of ESL and EFL immersion students in junior high", number Pro00074796, approved November 14th, 2017. Data for this project were collected in Taiwan but also included additional data collected in Canada under the project named "Age effects in child L2 acquisition", number Pro00057990 for which Dr. Johanne Paradis was the principal investigator. Data for the latter project was collected in Edmonton under the supervision of Dr. Paradis and organized by Keren Hernandez or Dr. Adrianna Soto-Corominas. Data collection in Vancouver was supervised by Dr. Stefka Marinova-Todd and organized by Xuan Zhang. The Alberta Language Environment Questionnaire-3 was created under the supervision of Dr. Paradis with contributions from Kayla Day and myself. Dr. Paradis assisted with the initial design of the grammaticality judgment task. The dissertation has also benefited from Dr. Paradis's comments on drafts throughout the writing process. The study presented in Chapter Three was published in Applied Psycholinguistics as "Rusk, B. V., Paradis, J., & Järvikivi, J. (2020). Comprehension of English plural-singular marking by Mandarin-L1, early L2-immersion learners." Dr. Paradis, Dr. Järvikivi and the anonymous journal reviewers contributed to the final version of this manuscript. With these exceptions, I was responsible for concept formation, experiment design, data collection, statistical analyses, and composition of the manuscript.

For Cordelia and Camille.

Of all the luck I have had in life,
getting to be your father will always be the greatest.

ACKNOWLEDGMENTS

I would like to express my gratitude to all of those who have supported me throughout my dissertation research project. I am particularly grateful to my supervisor, Dr. Johanne Paradis, for her guidance, mentorship, and for motivating me to continuously develop as a researcher. I would also like to thank the members of my supervisory committee, Dr. Juhani Järvikivi and Elena Nicoladis for their guidance and valuable feedback. As well, I would like to thank all the faculty within the department of linguistics. Over my years within the department, I benefited immensely from the collegial environment they have cultivated.

Next, I would like to thank all those who facilitated or helped me with data collection. In particular, I would like to thank Anne Sutcliffe for her support of the project which was key to getting dissertation data from participants in Taiwan. In addition, during data collection in Taiwan, Santiago Balado and Elaine Lien provided me with far more support than I expected to receive and it was greatly appreciated. I would also like to thank all of the participants who volunteered their time. Finally, I would like to thank my sister Meghan Rusk McKechnie for additional help with participant recruitment. Without participants, this type of research simply cannot happen and so all of these contributions were crucial.

I would also like to thank all of the other graduate students and post-docs I met within our department and whose friendships I will cherish, in particular Dr. Caelan Marrville, Dr. Yoichi Mukai, and Dr Adrianna Soto-Corominas. A special thank-you goes to Dr. Vincent Porretta whose generosity with sharing his expertise in both statistical analyses and eye-tracking methodology was an immense benefit to me at the early stages of study. And a final thank you goes to my friend Trent Clowater who offered to put up with me during the months I gathered data in Taiwan.

I would also like to thank the Social Sciences and Humanities Research Council for its funding support in providing a doctoral fellowship award (SSHRC Award: 752-2016-2105).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Decades of research show that immersion classrooms are a particularly effective way for children to achieve strong communicative proficiency in a second language (L2) that is not widely spoken in the community (Genesee, 2006; Lyster & Genesee, 2019). Through these programs, participating students develop L2 abilities as they learn their mainstream educational curricula through use of the L2. Despite beginning at a young age and having extensive L2 exposure over many years, children who participate in these programs do not uniformly attain L2 abilities identical to those who acquire the language of immersion (*i.e.*, the target language) as a first language (L1; Lyster & Genesee, 2019). For these learners, gaps in L2 abilities are evident for both their lower morphosyntactic accuracy (Kowal & Swain, 1997; Lyster, 2007) and their more limited range of vocabulary (Harley, 1992). The studies of the present dissertation were conducted to identify ways in which the English abilities of immersion learners differ from other groups of English speakers, and to understand how these differences are connected to their L2 learning context.

Research generally supports the common observation that children have an advantage in L2 learning over adults. Nevertheless, understanding of the impact of age on L2 acquisition has become more nuanced over time. Both adults and children can become highly proficient L2 users, but when compared to older L2 learners, very young L2 learners

living where the L2 is the majority language of a community (henceforth community-L2 learners) are much more likely to attain L2 abilities that more closely parallel L1 speakers of the target L2 (DeKeyser, 2012; Marinova-Todd, Marshall, & Snow, 2000). In order to capitalize on the apparent ability of children to learn an L2 through authentic language use, and in response to the perceived limitations of traditional approaches to L2 instruction, early immersion programs in Canada and the US were founded on the assumption that children could rely on L2 input and their natural incidental language learning mechanisms to develop L1-like proficiency in an L2 (Fortune, Tedick, & Walker, 2008; Genesee, 1984). Essentially, it was hoped that the success of community-L2 learners could be reproduced within immersion classrooms. Despite these expectations, research has consistently shown that children in immersion contexts have predictable differences in their L2 abilities when compared to L1 speakers of the target language. Immersion learners still achieve high levels of L2 communicative proficiency. However, why these learners continue to demonstrate non-target L2 abilities (*i.e.*, they continue to make errors, have smaller vocabularies) even after many years of L2 exposure is an important topic for basic L2 research, as well as immersion pedagogy, yet has remained under researched.

The overarching purpose of this dissertation is to examine how L2 abilities are impacted by learning in an immersion context. Though the common wisdom is that children are inherently and uniformly gifted at language learning, the evolving understanding of L2 acquisition in children indicates that building L2 proficiency is not the effortless, rapid process that it is sometimes believed to be (Paradis, 2019). Research over the last decade in community-L2 contexts increasingly shows that differences in either cognitive ability or L2 experiences can influence the L2 development of individual learners

(Chondrogianni, 2018; Paradis, 2011). Learning an L2 either within an immersion context, where the language is not widely spoken in the community, or in a community-L2 context is certain to impact the type and frequency of experiences that learners have with the L2. Though these differences are expected to impact L2 abilities, it is currently unclear the extent to which immersion learners' non-target L2 abilities are unique to immersion contexts or are shared by other early high-proficiency L2 learners. Where other potential impacting factors are held constant, differences in L2 abilities between groups should be considered evidence of the impact of L2 context. However, due to the lack of direct comparisons between immersion learners and their community-L2 peers, there is little previously collected empirical data to address this issue.

In examining the impact of learning context, the present dissertation aims to address three gaps where there is a lack of research in immersion. The first gap is that little is known about how immersion learners compare to other young high-proficiency L2 learners. Within the following studies, the L2-English immersion participants are compared to both L1-English speakers and community-L2 learners of English. While both comparison groups have value, community-L2 learners are also bilinguals and as such, are similar to the immersion learners in having the influence of co-existing and interacting languages (Grosjean, 2008). Thus, for the goal of understanding how the immersion context impacts L2 acquisition, community-L2 learners are a particularly valuable comparison group.

The second gap in immersion research is that little is known about how L1-Mandarin speakers acquire L2 English in immersion. Most research for immersion contexts has examined the language pairings of L1 English and L2 French, or L1 English and L2

Spanish (Genesee and Lindholm-Leary 2013). Despite the interest that both adult and community-L2 English learners who speak Mandarin as their L1 have received, little research has examined this pairing of typologically distinct languages in immersion (Xiong & Feng, 2018).

The third addressed gap is the lack of immersion research on L2 receptive abilities. Studies often conclude that immersion learners are less target-like in their L2 production than for their receptive abilities (*e.g.*, Genesee, 1987; Knell, Siegel, & Lin, 2007; Lindholm-Leary, 2011; Swain & Lapkin, 1982; Turnbull, Lapkin, & Hart, 2001). However, the receptive abilities of immersion learners have typically been assessed using holistic language assessments of general listening or reading comprehension (*e.g.*, Genesee, 1981; Swain & Lapkin, 1982; Turnbull, Lapkin, Hart, & Swain, 1998). In tasks that measure holistic comprehension (*e.g.*, a multiple-choice reading comprehension test), participants can potentially incorporate contextual and semantic information rather than morphosyntactic information to arrive at the correct interpretation of the L2. In contrast, L2 production is assessed using elicitation, sentence repetition, or detailed error analysis of oral interview transcripts (*e.g.*, Bild & Swain, 1989; Day & Shapson, 1987; Harley, Cummins, Swain, & Allen, 1990; Harley & Hart, 1997; Thomas, Williams, Jones, Davies, & Binks, 2014; Turnbull, Lapkin, Hart, & Swain, 1998), where errors with specific morphosyntactic L2 constructions are tabulated. The result of this asymmetry of assessment is that less is known about the receptive morphosyntactic L2 abilities of immersion learners at a more fine-grained level.

Given these research gaps, this dissertation examines the receptive L2 abilities that L1-Mandarin immersion learners develop in comparison to both L1-English speakers and

L1-Mandarin community-L2 learners. In addition to the impact of L2 learning context, this dissertation also compares how differences between individual learners may result in individual differences in L2 ability within each L2 learning context. By examining how both context and differences between learners impact immersion learners' L2 ability, the aim is to provide a deeper understanding of the factors that impact L2 development in an immersion context.

1.1. Research on Immersion

Broadly speaking, language immersion is an educational approach in which children are required to use an L2 in order to learn new content; in other words, children learn some or all of their mainstream curricular content through their L2 instead of the L2 being its own subject (L2 is used throughout to denote any additional language begun in the preschool years or after). Programs that share this feature may also be referred to as *bilingual education*, or *content and language integrated learning* (CLIL, Cenoz, Genesee, & Gorter, 2014). These programs are designed to foster additive bilingualism through use of the L2 as the medium of instruction (Fortune, Christian, & Tedick, 2011; Swain & Johnson, 1997). Students typically begin an immersion program with similar, limited levels of L2 proficiency (Swain & Johnson, 1997). Immersion programs can have a range of starting points and are correspondingly termed *early* for those that start within the preschool to first grade age range, *middle* for those who begin in the middle of primary school (approximately third grade of a K-12 system), and *late*, generally a seventh-grade start (Genesee and Lindholm-Leary 2013). For the present dissertation, the focus is early-start immersion learners.

Cummins (1991) estimated that by the end of the 1980s, it was likely that more than a thousand empirical studies on language immersion had been done on Canadian French immersion alone. This research on Canadian French immersion has played an important role in the contemporary popularity of this educational approach, as these studies provided a wealth of compelling evidence that, in addition to fostering strong L2 skills, participation in these programs did not negatively impact academic achievement or L1 abilities (Genesee, 1987; Harley, 1986; Swain & Lapkin, 1982; Turnbull, Lapkin, & Hart, 2001). Parallel findings outside of Canadian French immersion reinforce the value of the immersion approach for establishing language skills in an L2 while also supporting the L1 abilities and academic achievement (Alanís, 2000; Cazabon, Nicoladis, & Lambert, 1998; Cheng, Li, Kirby, Qiang, & Wade-Woolley, 2010; Knell, Siegel, & Lin, 2007; Lindholm-Leary, 2011).

Language immersion classrooms are now found worldwide (Fortune, Christian, & Tedick, 2011), including East Asia where there is a large and increasing demand for early English-immersion programs (Lin & Johnson, 2016). In Canada, French immersion emerged from the St. Lambert 'experiment' where, in the early 1960s, a parents' group pushed for an educational alternative that would more effectively meet the language needs of an Anglophone minority in a suburb of a French-English bilingual city in the French-majority province of Quebec (Tucker, 2008). Similar concerns about the effectiveness of traditional L2 instruction have made immersion an increasingly sought-after option in Mainland China (Cheng, 2012; Gao & Wang, 2017). For these contexts, research documenting the academic outcomes of Chinese children in early English immersion programs concludes that these programs can be similarly effective to Canadian French

immersion, with learners obtaining strong communicative ability in English, without negative impacts to L1 skills or academic performance (Cheng, 2012; Cheng, Li, Kirby, Qiang, & Wade-Woolley, 2010; Knell, Siegel, & Lin, 2007; Qiang & Kang, 2011), though research on these contexts is still fairly limited (Xiong & Feng, 2018).

While the cited research indicates that L1-Mandarin students can develop L2-English proficiency without compromising L1 or academic abilities, conclusions about students' L2 abilities are based on holistic assessments. The bulk of immersion research has prioritized assessment of language, literacy, and academic skills in students' L1 rather than the L2, with fewer studies specifically on L2 development (Hermanto, Moreno, & Bialystok, 2012). This research focus is understandable given that few parents would be likely to opt for an L2 program that left students at a disadvantage for other essential abilities. Thus, despite the number of studies done on immersion contexts, few offer fine-grained information about L2 abilities and this is especially true for receptive abilities. In contrast to either L1-Mandarin adult or community-L2 child learners, the L2-English abilities that L1-Mandarin immersion learners develop with specific L2 constructions is essentially unresearched.

A defining characteristic of the immersion context is that the L2 is not one widely spoken in the community, and thus, out-of-school access to the L2 is a key difference between immersion learners and community-L2 learners. It is considered characteristic of immersion that L2 input and experience are largely confined to the classroom (Ballinger, Lyster, Sterzuk, & Genesee, 2017; Swain & Johnson, 1997) and, previous research has found that immersion learners rarely use the L2 for peer-to-peer communication (Tarone & Swain, 1995). Nevertheless, this does not preclude individual students from having

meaningful L2 exposure outside of the classroom. Thus, it is possible that greater out-of-school L2 use has an important impact on the L2 ability of individual immersion learners. Further, increasingly widespread access to the Internet may provide immersion learners with new and beneficial L2 experiences. As a result, the potential impact of out-of-school L2 exposure should be an important topic for immersion contexts.

Early on, Genesee (1978) noted that immersion research tended to be evaluations of programs rather than evaluations of individual learners and the L2 abilities that a learner acquires. Genesee pointed out that, while studies of variations between individual immersion learners were valuable for extracting a deeper understanding of the process of L2 learning, there were few studies of this type in the early wave of immersion research. The importance of identifying factors that mediate immersion learners' L2 proficiency was reiterated in the early 1990s (Lapkin, Swain, & Shapson, 1990). Nevertheless, contemporary research on differences between learners in immersion contexts has largely been limited to the development of L2 reading skills (*e.g.*, Erdos, Genesee, Savage, & Haigh, 2011; Jared, Cormier, Levy, & Wade-Woolley, 2011; Li, Kirby, Cheng, Wade-Woolley, & Qiang, 2012). Some research exists for the impact of learners' cognitive abilities at the early stages of L2 vocabulary development (*e.g.*, Nicolay & Poncelet, 2013; Woumans, Ameloot, Keuleers, & Van Assche, 2019). However, much less research exists for sources of individual differences in vocabulary or morphosyntax abilities at later stages.

1.1.1. Acquisition of L2 Morphosyntax in Immersion Contexts

Though it is commonly believed that children are inherently and uniformly gifted at learning an additional language, and despite the effectiveness of immersion classrooms, the

L2 abilities learners develop in these programs are not identical to those of L1 speakers, particularly for morphosyntactic accuracy (Cummins, 2009; Lyster & Genesee, 2019; Swain & Lapkin, 1982). Specifically for Canadian French immersion, learners typically have persistent non-target use of constructions even when they are frequent in the L2 input, such as grammatical gender, articles, and verbs (Allen, Swain, Harley, & Cummins, 1990). In early-start programs, these learners begin receiving large amounts of L2 input in the year they turn five years old and typically amass 6000 hours of L2 exposure by the end of eighth grade (Turnbull, Lapkin, Hart, & Swain, 1998). These learners quickly outpace peers with traditional L2 instruction and can perform on par with L1-French peers on tests of holistic listening comprehension by eighth grade (Swain & Lapkin, 1982). Though L2 communicative ability remains high, non-target morphosyntactic ability typically endures into the final years of the immersion program (Harley & Hart, 1997; Turnbull, Lapkin, Hart, & Swain, 1998; Wesche, 1993). Studies outside of French immersion have come to parallel conclusions about the morphosyntactic ability of immersion learners. Research on immersion contexts involving languages as varied as Cherokee (Peter, Sly, & Hirata-Edds, 2011), Irish (Ó Duibhir, 2011), and Spanish (Herschensohn, Stevenson, & Waltmunson, 2005) document challenges learners have with attaining morphosyntactic accuracy.

In accounting for non-target morphosyntactic ability, immersion researchers propose that acquisition is tied to the specific L2 experience learners have within their classrooms (*e.g.*, Harley, Cummins, Swain, & Allen, 1990; Lyster, 2008). Though immersion learners receive large amounts of L2 exposure, it is considered characteristic of this context that exposure is limited to the classroom (Swain & Johnson, 1997). This means that learners get a lot of exposure to an academic register of the L2, but very little to informal

speech that would be used among peers (Tarone & Swain, 1995). This pattern of using the L2 within the classroom, but the L1 for social interaction outside the classroom has been found to be characteristic of a wide variety of immersion programs around the world (Ballinger, Lyster, Sterzuk, & Genesee, 2017).

The limitation in sociolinguistic variation in the input can have implications for the frequency of morphosyntactic constructions experienced (Lyster, 2007; Peter, Hirata-Edds, & Montgomery-Anderson, 2008). For example, French immersion learners have been found to overuse the informal second-person singular form *tu* in contexts where the formal *vous* would be more appropriate, and this has been linked to the infrequency of the sociolinguistic function of *vous* in classroom input (Swain & Carroll, 1987). In contrast, instances of *vous* as a marker of second-person plural are common in the input and immersion learners generally use it correctly. Because classroom input is not always representative of language use in a French majority community, this lack of exposure to a particular form-function mapping can result in non-target L2 use (Swain & Lapkin, 1990).

Sociolinguistic context can also have implications more directly on morphosyntax and may explain some of the difficulty that Canadian French immersion learners' experience with L2 verb forms. Studies examining the distribution of verbs that French immersion teachers produce for their students from grades three to six have shown input can be heavily skewed toward present and imperative forms with past, future, and conditional being much less frequent (Harley, Allen, Cummins, & Swain, 1987; Izquierdo, 2007). It has been proposed that the bias toward present and imperative forms is likely to impact verb form accuracy. Specifically for conditional forms, such as in response to a question like *Qu'est-ce que tu ferais si tu avais beaucoup d'argent?* 'What would you do if

you had a lot of money?' immersion learners in tenth grade only used a correct conditional verb form in 56% of responses, whereas L1-French tenth grade students produced an accurate conditional form in 98% of responses, and even the L1-French first graders were able to use a conditional form in 95% of responses (Harley & Swain, 1984). Swain (1988) pointed out that it is likely that immersion learners are never exposed to many verbs in less frequent forms like the conditional even into the later stages of immersion. Thus, she proposed 'restricted input' as a partial explanation for inaccurate verb marking with less frequent forms.

Naturally, if particular morphosyntactic constructions are rare or non-existent in the L2 input provided to immersion learners, they are unlikely to be acquired. However, non-target ability is still observed in the upper grades for Canadian French immersion for the highly frequent auxiliary verbs ÊTRE and AVOIR, and gender marking (Harley, 1993). By this point of their programs, these learners will have experienced thousands of hours of L2 input (Turnbull, Lapkin, Hart, & Swain, 1998), and should not be lacking in accurate exemplars of these L2 constructions. In explaining these discrepancies, Harley (1993) suggests that immersion learners misinterpret the L2 input that they hear in terms of linguistic knowledge that they already possess (*i.e*, their L1).

As Harley (1993) indicates, a learner's L1 can have an important influence on L2 development in an immersion context, and in many cases these influences can be traced to specific properties of the L1. Previous research with L1-Mandarin community-L2 English learners indicates that grammatical morphology is often acquired more slowly by these learners, with differences between them and L2 learners with other L1 backgrounds evident even after many years (*e.g.*, Jia & Fuse, 2007; Paradis, 2011; Paradis, Tulpar, &

Arppe, 2016). Given the evidence that 1) immersion learners' acquisition of L2 morphosyntax can be subject to a long developmental trajectory and 2) non-target abilities can stretch past the early acquisition years for community-L2 learners, it could be predicted that these L2-English constructions would be particularly vulnerable for immersion learners. However, the L2 morphosyntactic abilities of L1-Mandarin speakers in L2-English immersion is currently under researched.

1.1.2. Acquisition of L2 Vocabulary in Immersion Contexts

In addition to non-target morphosyntactic ability, previous research has shown that immersion learners lag behind monolinguals in their knowledge of L2 vocabulary, though they establish larger L2 vocabularies than peers in traditional L2 instruction (Harley & Jean, 1999). Early French immersion students in second and fifth grade (their third or sixth years of immersion, respectively) have been shown to score more than one standard deviation below the standardized mean score on a monolingual-normed standardized test of French receptive vocabulary (Hermanto, Moreno, & Bialystok, 2012). Comparison of scores across these two cross-sectional time points suggests that learners in this context were not closing the receptive vocabulary gap between them and monolinguals. As compared to L1-French speaking peers, sixth grade immersion learners have lower lexical richness in their L2-French writing, avoiding use of French verbs with meanings incongruous with English counterparts and making proportionally greater use of high-frequency verbs (Harley & King, 1989). This latter strategy of resorting to words with more general meanings where L1 speakers would use a more exact term is observed in immersion learners' oral abilities throughout their programs and is likely used to

compensate for more limited vocabulary knowledge (Harley, 1992). However, this strategy is not unique to immersion learners as it has also been observed for community-L2 learners in their first few years of L2 exposure (Golberg, Paradis, & Crago, 2008).

For immersion learners, the benefit of more years of immersion on vocabulary knowledge is evident in the early stages of immersion (Harley & Jean, 1999; Kalia, Lane, & Wilbourn, 2018; Nicolay & Poncelet, 2013), but, like community-L2 learners, rapid early vocabulary development appears to eventually slow down (*e.g.,* Soto-Corominas, Paradis, Rusk, Marinova-Todd, & Zhang, 2020). When compared to early immersion learners, late-start learners (*i.e.*, those who start in seventh grade, rather than kindergarten) begin to catch up to early-start peers by tenth grade (Harley & Jean, 1999). However, the impact of more exposure may depend on the measure of vocabulary used, as findings for eleventh grade learners showed that early learners were significantly more able to identify words in the L2 than their late starting peers (Harley & Hart, 1997).

In contrast, research shows that community-L2 learners are steadily catching up with monolinguals (Farnia & Geva, 2011; Golberg, Paradis, & Crago, 2008) and potentially meet or exceed the PPVT mean standard score around four to six years of L2 exposure (Paradis & Jia, 2017; Soto-Corominas et al., 2020). Research with community-L2 learners also shows that the child's L1-L2 pairing matters, with more typological similarity between the languages providing a benefit to L2 receptive vocabulary development (Blom et al., 2020; Goriot et al., 2018). For L2 vocabulary learning in general, learning a language with a large number of cognates can facilitate acquisition (De Groot & Keijzer, 2000; Willis & Ohashi, 2012). This can also benefit the acquisition of grammatical morphology as learners can analyze unknown morphemes on known words such as learning the bound progressive

morpheme by comparing the transformation in English of USE to *us-ing* with Spanish *us-ando* (August, Carlo, Dressler, & Snow, 2005). However, in contrast to languages with greater historical and geographical links, Mandarin and English share very few cognates, and thus learners with this language pairing are less likely to benefit from this type of lexical overlap (Sheng, Lam, Cruz, & Fulton, 2016).

The research discussed above for vocabulary acquisition in an immersion context involves pairings of European languages, and mostly L1 English to L2 French. Thus, it is unclear how these findings generalize to more typologically distant languages that share few cognates. Like ability with L2 morphosyntax, little is known about the L2 vocabulary ability of L1-Mandarin, L2-English immersion learners.

1.2. Non-target Comprehension Vs. Production

For L2 morphosyntax, the exact nature of how immersion learners' L2 ability may be non-target remains unclear. Within the literature on language immersion classrooms, the assertion that students have superior proficiency in receptive language ability (*i.e.*, listening or reading comprehension) as compared to productive ability (*i.e.*, speaking or writing) is often made (*e.g.*, Cummins, 1998; Kowal & Swain, 1997; Lyster, 2007). On one level, this assertion may simply mean that immersion learners are more easily identified as L2 speakers by their language production rather than their language comprehension, a conclusion well-supported by evidence. On the other hand, this assertion may reflect a belief that there is an asymmetry between immersion learners' receptive and productive linguistic abilities. While a cursory review of the literature may appear to support an asymmetry, the findings may also be explained by the methodology used to examine either

receptive or productive ability. To assess receptive ability, immersion studies typically use measures such as general listening or text comprehension that assess comprehension on a holistic level (*e.g.*, Genesee, 1981; Harley & Hart, 1997; Siegel et al., 2010; Swain & Lapkin, 1982; Turnbull, Lapkin, Hart, & Swain, 1998), as opposed to assessing comprehension of a specific morphosyntactic construction. In contrast to holistic comprehension measures, production is often evaluated using tasks such as elicitation, sentence repetition, or detailed error analysis from oral interview transcriptions (*e.g.*, Bild & Swain, 1989; Day & Shapson, 1987; Harley, Cummins, Swain, & Allen, 1990; Harley & Hart, 1997; Thomas, Williams, Jones, Davies, & Binks, 2014; Turnbull, Lapkin, Hart, & Swain, 1998). These production tasks detect non-target use of L2 constructions at a level of granularity that holistic comprehension tasks generally do not. For example, an L2 learner who fails to produce the third-person agreement marker when saying **the dog run at the park* has made a detectable error. However, when listening, the learner's failure to interpret the agreement marker is unlikely to interfere with holistic comprehension of the situation being described. For educational research, holistic assessment tasks will often be superior for evaluating comprehension skills relevant for academic success, but these give little specific information about L2 morphosyntactic ability. Currently, there is little research that has given immersion learners' receptive L2 morphosyntactic ability the same scrutiny as productive ability.

Though holistic methods have predominated in assessment of immersion learners' receptive L2 abilities, one notable exception is Lew-Williams (2017). This study reported three separate eye-tracking experiments with L1-English participants in Spanish two-way immersion. The experiments measured participants' ability to identify upcoming nouns

using either Spanish determiners with abstract grammatical gender; grammatical gender that coincided with the notional gender of the noun (*e.g.*, the Spanish translations of 'the man' or 'the woman'); or number information. To test receptive ability with these constructions, the experiments in Lew-Williams (2017) compared conditions where, for the grammatical-gender experiment, participants viewed side-by-side pictures of objects or animals that either belonged to the same grammatical gender category (*e.g.*, *pelota*, "ball$_{[fem.]}$" and *galleta*, "cookie$_{[fem.]}$") or different categories (*e.g.*, *pelota*, "ball$_{[fem.]}$" and *zapato*, "shoe$_{[masc.]}$"). At the same time, participants listened to simple sentences in which one of the pictured nouns was presented with its associated determiner, either *la* or *el*. When the gender of the nouns is the same, the determiner provides no information about which noun will follow. However, when the gender of the nouns is different, the determiner can serve as a cue to the noun that will follow. If participants interpret grammatical gender, then eye-tracking should show that they look to the picture named in the stimulus sentence earlier in the different-gender condition as opposed to the same-gender condition. The other two experiments worked similarly, contrasting notional gender (*e.g.*, *la niña*, "the$_{[fem.]}$ girl$_{[fem.]}$" and *la señora*, "the$_{[fem.]}$ woman$_{[fem.]}$"; or *la niña*, "the$_{[fem.]}$ girl$_{[fem.]}$" and *el niño*, "the$_{[masc.]}$ boy$_{[masc.]}$"), or number-marked determiners (*e.g.*, *el carro*, "the$_{[sing.]}$ car$_{[sing.]}$" and *el zapato*, "the$_{[sing.]}$ shoe$_{[sing.]}$," or *el carro* and *los zapatos*, "the$_{[plu.]}$ shoes$_{[plu.]}$").

Results for both six- and ten-year-old immersion participants were compared to their respective L1-Spanish classmates. The L2-Spanish participants in fourth or fifth grade, like L1-Spanish peers, were able to use notional gender and number-marked determiners to identify referents, and six-year-old L2 participants already demonstrated the ability to interpret number-marked determiners. However, unlike their L1-Spanish peers, neither L2

age group could use grammatical gender this way, suggesting that the acquisition of these immersion learners was similar to late L2 learners, who also often demonstrate a similar inability to use gender information (Grüter, Lew-Williams, & Fernald, 2012; Hopp, 2013; Lew-Williams & Fernald, 2010). The results of Lew-Williams (2017) indicate that when receptive ability with specific L2 constructions is examined closely, L2 ability can also be non-target. However, the topic of immersion learners' receptive L2 abilities including either how they perceive or interpret L2 input remains under researched.

1.3. Comparison of Grammatical Morphology in Mandarin and English

Even for early L2 learners in community-L2 contexts, L1 can have an impact on acquisition of morphosyntax (Blom, Paradis, & Sorenson Duncan, 2012; McDonald, 2000; Paradis, 2011), and L1-L2 differences are implicated for immersion learners as well (Harley, 1993; Lew-Williams, 2017). Differences in morphosyntax between L1 and L2 are particularly relevant to the present dissertation because the two languages involved, Mandarin as an L1 and English as an L2, are typologically distinct. Differences in grammatical morphemes are particularly salient, and currently, there are numerous studies in the L2 acquisition literature focusing on how L1-Mandarin/Chinese speakers acquire L2-English form-function mappings (*i.e.*, constructions) that are absent in the L1 (*e.g.*, Jia & Fuse, 2007; Jiang, 2007; Luk & Shirai, 2009; Paradis, Tulpar, & Arppe, 2016; Trenkic, Mirković, & Altmann, 2014).

One key difference is that Mandarin, an isolating language (Li & Thompson, 1989), lacks bound inflectional morphemes. Whether a community-L2 learner's L1 is isolating or inflecting has been shown to affect the rate at which the child develops particular

morphosyntactic constructions, with speakers of inflecting L1s acquiring L2 inflectional morphemes more rapidly than those whose L1 is an isolating language (Blom & Paradis, 2013; Blom, Paradis, & Sorenson Duncan, 2012; Paradis, 2011). Specifically for English verbal morphology, non-target L2 ability can endure beyond the first few years of L2 acquisition for L1 speakers of isolating languages. In a longitudinal study, Paradis, Tulpar, and Arppe (2016) tested the L2-English ability of L1-Chinese participants. Elicited production was used to test BE, DO, third-person singular, and past tense (both regular and irregular); and grammaticality judgments were used for subject-verb agreement and dropped '*-ing'* marking. At the outset of the study, the participants already had more than four years of L2-English exposure and were tested every year for three years from 8;5 to 10;5 years of age. Despite having an early age of beginning L2 exposure and more than six years of English exposure by the final round of data collection, 11 of the 18 participant children had not reached a criterion score set for six-year-old L1-English speakers. Further, the authors concluded that development appeared to be plateauing. These findings are also supported by Jia and Fuse (2007) who tracked 10 L1-Mandarin participants who ranged in age of first English exposure from five to 16 years of age. These participants' production of English inflectional morphology was measured at multiple points over five years of English exposure. Both Paradis, Tulpar, and Arppe (2016) and Jia and Fuse (2007) found that participants were accurate with '*-ing"* marking but were less accurate with past tense and subject-verb agreement morphemes.

The impact of L1 on community-L2 learners can also last until adulthood. In a study of adults who began L2 English as children, McDonald (2000) found that L1 predicted whether community L2 learners ultimately had L2-English ability that was comparable to

L1-English speakers. In this study, a range of English morphosyntactic constructions were examined using a grammaticality judgment task. Early L1-Spanish speakers matched the accuracy of the L1-English group for all English constructions. However, for the L1-Vietnamese group, accuracy was lower than the L1-English group for past tense, third person, plural, and articles. All of these grammatical morphemes were expected to be difficult for Vietnamese speakers given that Vietnamese, like Mandarin, is an isolating language, that also lacks articles.

For English articles, it is unclear whether results similar to McDonald (2000) would be found for other L1-speaking community-L2 groups. The impact of L1 on community-L2 learners' acquisition of L2-English articles has been found to be limited during the early stages of development (Zdorenko & Paradis, 2008) and six- to eight-year-old L1-Chinese speakers in New York public schools have been found to be close to 90% accurate with articles on an elicitation task (Dulay & Burt, 1974).

In contrast to articles, but similar to verbal morphology discussed above, plural-singular marking has been identified as a potentially problematic L2-English construction for L1-Chinese speakers in both adult/late L2 acquisition (Jiang, 2007; Luk & Shirai, 2009), as well as, early community-L2 acquisition (Jia, 2003; Xu Rattanasone, Davies, Schembri, Andronos, & Demuth, 2016). In L1 acquisition of English, plural marking is one of the earliest acquired grammatical morphemes (Brown, 1973; de Villiers & de Villiers, 1973), and the earliest acquired among grammatical morphemes identified above as potentially subject to L1 influence (*i.e.*, markers for subject-verb agreement, past tense, plural, and articles). Because plural-singular marking is early acquired in L1 acquisition, but potentially delayed in L2 acquisition for speakers of languages without a transferable L1

construction, L2 ability with plural-singular marking was a key focus of the present dissertation.

English and Mandarin differ in both how and when nouns are marked for number, and therefore, like many other grammatical morphemes in English, plural-singular marking is not a construction that can be transferred from Mandarin (Cheng & Sybesma, 1999; Li & Thompson, 1989; Li, 1999). As a result, immersion learners need to learn it through their L2 exposure. Nouns can be quantified in both languages, but a key difference is that marking number is non-optional for English count nouns (Corbett, 2000). In contrast, nouns unmarked for number are very common in typical Mandarin usage (Cheng & Sybesma, 1999; Li & Thompson, 1989; Li, 1999), and thus, can be considered to have *general number* (Corbett, 2000; Rullmann & You, 2006) meaning that, unlike English, they can be expressed without reference to number and are thus ambiguous in this regard when used outside of a discourse context. The lack of obligatory number marking in Mandarin can be seen in the two potential glosses of Example 1.1.

1.1. 狗 有 尾巴
gǒu *yǒu* wěibā
dog have tail
'The dogs have tails' or 'The dog has a tail'

While Mandarin has a morpheme '*-men*' that can follow pronouns and some common nouns referring to humans (Li, 1999), it lacks the extensive productivity of the bound '*-s*' in English. Productive ways of making explicit reference to a single item or multiple items are classifier constructions taking the form NUMERAL + CLASSIFIER + NOUN with the noun taking no bound morphology, meaning the head noun is the same whether the numeral is *one*, *two* or *twenty*. Demonstratives can also be combined with classifiers to form words equivalent to

this and *these* in English. In addition, Mandarin has a small number of nouns that can appear directly after a numeral without a classifier (*e.g.*, 一天 [yī tiān], 'one day'). A classifier construction for a count noun is given in Example 1.2. (CL denotes a classifier.)

1.2. 三隻 狗
sānzhī *gǒu*
three-CL dog
'Three dogs'

Again, however, while a classifier construction can disambiguate the number of items to which reference is made, it is not required for a noun to be interpreted as denoting either one or many. As a result of this crosslinguistic difference, Mandarin-speaking L2-English learners have to learn to consistently mark this contrast in production and monitor for it in comprehension.

In comparison to other grammatical morphemes, there are reasons that suggest plural-singular marking should not be overly difficult to acquire. First, in L1 acquisition of English, plural '*-s*' is one of the earliest grammatical morphemes produced consistently by children learning English as an L1 (Brown, 1973; de Villiers & de Villiers, 1973), and generally approaches adult-like mastery around seven years of age (Berko, 1958; Graves & Koziol, 1971). Second, plural-singular marking is semantically transparent, as the singular and plural forms denote what should be a salient observable difference in number. In contrast, grammatical gender's lack of a semantically transparent function has been argued to be a potential obstacle to acquisition in immersion (Harley, 1993; Lew-Williams, 2017). If semantic transparency makes a difference, then plural-singular marking should be comparatively easier to acquire. Third, plural-singular marking in English is highly frequent as its use is non-optional on count nouns. It cannot always be assumed that even

adult L1 speakers will attain completely consistent comprehension for all L1 morphosyntactic constructions (Dąbrowska, 2018). However, for plural-singular marking, research indicates that attaining a target threshold of accurate use is not unrealistic for L2 speaking children.

Despite reasons to expect immersion learners should be able to acquire English plural-singular marking, the construction has been proposed as a potentially problematic one for Mandarin speakers, and speakers of other classifier languages (Jiang, 2007; Luk & Shirai, 2009). Further, evidence indicates that in the early stage of L2 English acquisition, preschool-aged community-L2 learners whose L1 is Chinese lag behind speakers of other L1s and English monolingual children for comprehension of plural-singular marking (Xu Rattanasone, Davies, Schembri, Andronos, & Demuth, 2016). In addition, longitudinal research shows L1-Mandarin children and adolescents can still produce English plural-singular marking variably even after five years of exposure (Jia, 2003). This typological difference with English is shared by both Korean (Kwon & Zribi-Hertz, 2004) and Japanese (Nakanishi & Tomioka, 2004), which also use classifiers when marking number optionally on nouns. For children with these L1s who are learning English in community-L2 contexts, production of plural '*-s*' appears to come later in L2 English than in L1 English relative to other grammatical morphemes (Hakuta, 1976; Shin & Milroy, 1999). Given that acquisition of morphosyntax in general is regarded as a weakness for immersion learners (Harley, Cummins, Swain, & Allen, 1990) and evidence that this construction may be impacted by L1-L2 differences, it is unclear whether the immersion participants will show target-like comprehension of English plural-singular marking.

1.4. Individual Differences in Language Ability

A fundamental premise of this dissertation is that the specific L2 experiences immersion learners have within their learning context impact their L2 acquisition. Previous research has drawn a connection between immersion learners' L2 ability, and both the language experiences they have in school and their lack of experiences outside of school (*e.g.*, Cummins, 2009; Lyster, 2008; Swain & Carroll, 1987; Tarone & Swain, 1995). However, because little is known about how the L2 abilities of these learners compare to other high-proficiency L2 learners, it is unclear the degree to which non-target L2 ability is the result of experiential factors associated with the immersion context, or to L2 acquisition or bilingualism more generally. Further, research in community-L2 contexts increasingly shows that differences in L2 input or experience between learners can result in differences between the L2 abilities of individual L2 learners (*e.g.*, Chondrogianni & Marinis, 2011; Paradis, 2011; Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017). However, there is little research that addresses this in immersion contexts. Thus, it is difficult to draw conclusions about how the immersion context impacts L2 ability, or the factors that result in individual L2-ability differences between immersion learners.

A clearer understanding of how the language experience of immersion learners impacts L2 ability is relevant to contemporary research on language acquisition. For decades, generative approaches to language acquisition have been dominant. These approaches originate from the theoretical work of Noam Chomsky who writes, "Grammar. .. [is] acquired by virtually everyone, effortlessly, rapidly, in a uniform manner" (Chomsky, 1976, p. 144). However, in contrast to the rapid and universal development of linguistic ability proposed by generative approaches, accumulating evidence shows that the process

is not as rapid, uniform, or effortless in either an L1 or L2 as previously believed. Individual differences in rate of L1 development can result from variations in cognitive ability (Kidd & Donnelly, 2020) or differences in language input and experience (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Walker et al., 2020). In adulthood, an individual's ability with L1 morphosyntactic constructions can vary based on both cognitive and language input and experience variables (Dąbrowska, 2018). For children exposed to multiple languages, the situation is further complicated because language experience is divided between multiple languages within often vastly different learning contexts. In community-L2 contexts, L2 input factors like the proficiency of interlocutors or the balance of input between the languages impact development (Hoff, 2020; Paradis & Jia, 2017). Despite the expectation that the ultimate attainment of community-L2 learners will closely parallel the linguistic abilities of L1 speakers, research shows that even the attainment of these early learners can depend on language learning aptitude (Abrahamsson & Hyltenstam, 2009); the learner's L1 (McDonald, 2000); and/or type of L2 experience (Flege, Yeni-Komshian, & Liu, 1999). The emerging picture is that, in either L1 or L2, individual language development and outcomes are sensitive to a multitude of factors.

Research on early community-L2 learners has sought to identify influential factors in order to better understand the processes and mechanisms underlying L2 acquisition (Chondrogianni, 2018; Paradis, 2011). This research often groups impacting factors as either learner-internal or learner-external (*e.g.*, Chondrogianni & Marinis, 2011; Paradis, 2011; Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017; Sorenson Duncan & Paradis, 2020). Internal factors include cognitive abilities like memory or analytical skills or accumulated/entrenched L1 knowledge. External factors are ones that influence the quality

and quantity of an individual's L2 input and experiences and can include such things as the frequency of L2 use at home among parents or siblings, language use with peers, or access to media in the L2. The extent to which these external factors contribute to a richer L2 environment should be expected to correlate strongly with an individual's L2 learning context.

To date, there have been very few studies in immersion contexts that have examined the impacts of learner factors and how these can result in individual differences in L2 abilities. In addition, existing studies have generally focused on learner-internal rather than learner-external differences. For vocabulary, higher rates of vocabulary acquisition have been shown for those with higher verbal memory scores and better cognitive flexibility (Nicolay & Poncelet, 2013), as well as better inhibitory control (Woumans, Ameloot, Keuleers, & Van Assche, 2019). For immersion learners with many more years of exposure, language aptitude abilities reflecting better memory for text predict L2 vocabulary recognition (Harley & Hart, 1997). Factors that impact immersion learners' general L2 proficiency and/or comprehension include learner-internal factors like nonverbal reasoning (Genesee & Hamayan, 1980), intelligence (Genesee, 1976), receptive vocabulary (Vandergrift & Baker, 2015; Vandergrift & Baker, 2018), and phonological awareness (Li, Cheng, & Kirby, 2012). Similarly, listening comprehension has been shown to benefit from measures of language aptitude reflecting better memory ability and language analytic skills (Harley & Hart, 1997). In addition, the learner-external factor of greater L2 use with acquaintances/peers can impact general L2 proficiency (Hamayan, Genesee, & Tucker, 1977).

1.5. The Importance of Research on Immersion Contexts

In contrast to what is typical for research on immersion, the immediate motivation for this dissertation is not to improve immersion pedagogy, but rather to contribute to a better understanding of the process of L2 acquisition in immersion and other L2 learning contexts in general. Research in immersion has made multiple compelling contributions to the field of L2 research (see Swain, 2000 for a review), yet has largely been under appreciated as a setting for general L2 research. One of these key contributions has been demonstrating that L2 input alone is insufficient for development of completely target-like L2 ability particularly for morphosyntactic accuracy in language production (Lyster & Tedick, 2014; Swain & Lapkin, 1995). While research in immersion contexts supports the importance of L2 input, it has also established that years of daily authentic (albeit sociolinguistically limited) input does not result in a replication of L1-like ability with an L2 (Harley & Hart, 1997; Kowal & Swain, 1997; Turnbull, Lapkin, Hart, & Swain, 1998; Wesche, 1993). Thus, immersion served as a counter argument to proposals, popular throughout the 1980s, that claimed only authentic/comprehensible input was necessary for L2 acquisition (*e.g.*, Krashen, 1985).

Though research for a single L2 context cannot be immediately generalized beyond that context, there are several reasons that immersion learners are particularly valuable to understanding how L2 proficiency develops. Immersion contexts create high proficiency L2 speakers within a more controlled environment than that of other L2 learners (Swain, 2000). Learners within a single immersion program generally have the same starting age, cumulative L2 exposure, current L2 exposure, L1-L2 pairing, and are not typically self-selected, meaning that the decision to undertake L2 study is not their own. Much more is

known about the L2 exposure background of immersion learners than for other groups of high proficiency L2 learners, such as community-L2 learners who have extensive and diverse exposure outside their classrooms as they learn the majority language of their community (Paradis & Jia, 2017). This is especially true in comparison to L2 participants recruited from undergraduate participant pools where details about prior L2 exposure beyond age of beginning are often undocumented, obscuring potentially important differences reflecting the diversity of life/learning experiences these L2 learners have with the L2. It is important that these sources of variation are known, and it is in these immersion contexts where variables of language experience are most easily documented and potentially manipulated. Essentially, immersion contexts are likely as close to 'laboratory' conditions as L2 research can ethically come. Though a better understanding of L2 development in immersion contexts has a clear value to L2 pedagogy, the potential of these contexts for investigating more general L2 research topics has been under appreciated.

1.6. Theoretical Framework

The purpose of this dissertation is to examine the L2 abilities of immersion learners and factors that these abilities. In order to investigate this topic, empirical evidence is gathered from an under researched immersion context. As valuable as this type of evidence can potentially be, its value is diminished if it cannot be cohesively integrated within a theoretical framework. At some level, all theories must link L2 experience with L2 ability. However, this connection is vital to usage-based theories where L2 acquisition is thought to be driven by input and experience (Wulff & Ellis, 2018). From a usage-based approach,

language ability is the cognitive organization of the individuals' concrete experiences with language from which learners inductively learn abstract linguistic knowledge (Bybee, 2006; Lieven, 2016). This emphasis on learning from actual language use means that what is acquired is ultimately a reflection of the social-interactional contexts in which the learner has been exposed to the language (Beckner et al., 2009; Wulff & Ellis, 2018). In the concluding chapter of this dissertation, research in the usage-based paradigm is discussed both in terms of how the findings of the present dissertation are consistent with these approaches, and how these approaches contribute to a better understanding of L2 acquisition in the immersion context. Specifically, elements of usage-based approaches are used to explain why differences in L2 ability emerge and why they potentially persist in immersion.

1.7. General Introduction to the Immersion Participants

The central participants of these studies were 37 adolescent Taiwanese middle-school participants recruited from a single private high school in northern Taiwan, and studying in grade seven, eight, or nine (mean age = 13 years, 9 months, *SD* = 9 months). These participants spoke Mandarin as their L1, though a few also spoke an additional language, as well. The most common of these was Taiwanese, the local dialect of Min Nan Chinese. These participants had been learning in an immersion context taught by a foreign teacher from an English-speaking country (*e.g.,* the UK, US, Canada, etc.) since preschool (mean age of beginning English immersion = 4;9, *SD* = 28 months). After preschool, they attended a bilingual elementary school where half their content instruction was in Mandarin from a local teacher, and half in English. From the beginning of grade seven their entire school day

was in English. On average, they had attended an immersion program for eight years and two months (*SD* = 27 months). Because even L1-Mandarin/Chinese community-L2 learners have been shown to have non-target L2 ability with L2-English grammatical morphemes beyond five or six years of L2 exposure (Jia & Fuse, 2007; Paradis, Tulpar, & Arppe, 2016), adolescent learners were chosen to ensure that the immersion learners had a reasonable amount of time for L2 development.

Some collected data for individual participants were excluded entirely from analysis for reasons including exposure to English from birth, fewer than five years of immersion exposure (usually transfer-in from an instructed-L2 program), or, in one case, being an outlier for having very low receptive vocabulary. In addition to the immersion learners, the study reported in Chapter Two includes two age-matched comparison groups, one of L1-English speakers and the other, community-L2 learners. The study in Chapter Three compares the immersion learners to just the L1 speakers and in Chapter Four, the immersion learners are compared with just the community-L2 participants. Details for these comparison groups are given in the respective chapters.

1.8. Research Questions and Individual Studies

As stated, the purpose of this dissertation is to examine the L2 abilities of immersion learners and factors that these abilities. This topic is addressed in the studies presented in the following three chapters that collectively investigate the following three overarching research questions.

1. How do the receptive L2 abilities of L1-Mandarin adolescent immersion learners compare to the English abilities of L1-Mandarin community-L2 learners and/or L1-English speakers?

2. Across the two L2 learning contexts, how do out-of-school L2 input and experience differ between the immersion and community-L2 contexts, and how are these between-context differences associated with differences in L2 ability?
3. Within the immersion context, how do differences between individual immersion learners in their out-of-school L2 input and experiences impact individual differences in L2 ability?

1.8.1. Study One

The main goal of the study presented in Chapter Two was to compare the morphosyntactic ability of immersion learners with community-L2 learners. Existing research supports the expectation that L2 exposure in immersion classrooms results in morphosyntactic abilities that are less target-like than those of community-L2 speakers. However, morphosyntactic ability can vary in either context, and little is known about how these groups compare directly, or how the impact of individual learner factors on L2 morphosyntactic acquisition compares across contexts. To address these gaps, the study presented in Chapter Two compared the ability of early learners of L2 English across these learning contexts. Three groups of adolescent English speakers were compared. These groups were 1) L2 learners who had acquired English as a foreign language in immersion classrooms; 2) L2 learners who had acquired English in community-L2 contexts; and 3) L1-English monolinguals as a control group. Both L2 groups were L1 speakers of Mandarin. Morphosyntactic ability was examined using a grammaticality judgment task examining English grammatical morphemes, which have been shown to be challenging for speakers of isolating L1s (Blom, Paradis, & Sorenson Duncan, 2012; McDonald, 2000). In addition, this study examined the potential impact of learner factors. The study also examined these grammatical morphemes within syntactic constructions that were expected to differ in relative difficulty.

This study included two research questions. The first research question examined whether there are differences between each participant group on the experimental task, and whether groups were similarly or differently impacted by stimuli at varying levels of difficulty. Based on prior literature, it was expected that all groups would demonstrate receptive ability indicated as accurate responses on the grammaticality judgment task, but that the control group would be the most accurate, the community-L2 group the next most accurate, and the immersion participants the least accurate. All groups were expected to be more accurate on the easier stimulus items. The second research question examined the impact of both learner-internal and learner-external factors on task accuracy for the two L2 groups. It was expected that learner factors would predict differences within each group. However, it was not assumed that these would be the same for each L2 group, or for both levels of relative construction difficulty.

1.8.2. Study Two

The main goal of the study presented in Chapter Three was to investigate comprehension of English plural-singular marking by a group of L2-English immersion learners whose L1 is Mandarin. The study compared adolescent immersion learners with age-matched participants who spoke English as their L1. This study employed a visual-world eye-tracking task measuring online comprehension combined with a picture decision task reflecting accuracy of sentence interpretation.

The data from this task were used to examine two research questions. The first research question examined whether the adolescent immersion learners were similar to the age-matched L1-English group in offline interpretations of the stimuli as measured by

picture-decision task responses. The expectation based on prior research was that the immersion group's responses would be less accurate than those of the L1-English group. The second research question examined online comprehension of plural-singular marking as measured by eye-tracking. Eye-tracking allowed for the proportion of looks to target images to be compared across groups. Like the first research question, it was expected that immersion learners would show less sensitivity to plural-singular marking as demonstrated by fewer and later looks to target images.

1.8.3. Study Three

Chapter Four reports on a study that examined how the out-of-school L2 learning environments of the adolescent immersion learners differ from those of a group of community-L2 learners, and how this impacts L2 receptive vocabulary ability. It has long been considered typical of immersion contexts that learners' L2 exposure is limited to the classroom (Swain & Johnson, 1997). However, it is unclear how this applies to immersion learners in Taiwan or whether widespread access to the Internet provides immersion learners with new and beneficial L2 experiences. In addition, little research has probed how out-of-school L2 experience potentially impacts immersion learners' L2 abilities.

This study had three research aims. The first was to examine how L2 experiences outside of school compare for adolescent L2 learners in either a community-L2 context or an immersion context. The second aim was to see if the L2 receptive vocabulary ability of participants differed across the two contexts. The third aim was to connect the L2 experiences of learners within each context to L2 receptive vocabulary ability. This was done by modelling learner-external factors related to learners' L2 learning environments

as potential predictors of individual differences in L2 receptive vocabulary. Data collected learner-external sources of out-of-school L2 experiences were grouped depending on whether they could be categorized as opportunities for additional L2 input, or whether they were more balanced in opportunities for both L2 input and L2 output (*i.e.,* reflected interactional L2 use). This distinction was motivated given that immersion learners' in-school L2 experience has been shown to be predominantly L2 input with few opportunities to produce L2 output (Allen, Swain, Harley, & Cummins, 1990; Swain, 1988). Given that the immersion learners are foreign language learners, the groups were expected to be different in that the immersion learners would have fewer opportunities for either L2 input or interactional use, and that these diminished opportunities would also be associated with lower receptive vocabulary scores. However, it was also expected that greater out-of-school L2 exposure would predict higher vocabulary scores for participants within either group.

2 Comparing the L2 morphosyntactic ability of L1-Mandarin, early L2 English speakers from immersion programs in Taiwan and mainstream English schools in Canada: Does learning context make a difference in outcomes?

2.1. Introduction

The present study examines potential differences in second language (L2) morphosyntactic ability between two highly proficient, but distinct groups of adolescent L2 learners. It is commonly observed that learners who begin an L2 in early childhood eventually demonstrate target language linguistic ability comparable to individuals who have spoken the language continuously from birth (*i.e.*, L1 speakers, DeKeyser, 2012). Nevertheless, the appearance of L1-like proficiency is not characteristic of all child L2 learners, but rather appears mainly limited to minority-L1 children living where the L2 is the majority language of both the community and mainstream schooling. In contrast to these community-L2 learners, this degree of target language convergence is generally not observed for L2 learners in foreign-language immersion classrooms. Immersion learners typically receive thousands of hours of L2 input spanning many years as they use the L2 to learn non-language content such as math or science (Turnbull, Lapkin, Hart, & Swain, 1998). In contrast to community-L2 learners, some non-target ability with L2 morphosyntax is considered characteristic of immersion learners (Kowal & Swain, 1997; Lyster, 2007),

despite their early start and the amount of L2 input they receive. However, little is known about how the morphosyntactic abilities of immersion and community-L2 learners compare directly. In addition, little is known about sources of potential individual differences in immersion learners' morphosyntactic ability and how these compare to the better-documented sources of individual differences in community-L2 learners.

Accordingly, the present study compared the morphosyntactic ability of three adolescent participant groups. Groups included two L1-Mandarin L2-English groups and a group of L1-English monolinguals. Participants of both L2 groups began English exposure in preschool but differed by whether their English exposure was in a foreign-language immersion context or in a community-L2 context. Morphosyntactic ability was gauged using a grammaticality judgment task. In order to better understand factors that predict individual differences in morphosyntactic ability in each L2 context, the impact of learner-external and learner-internal factors was also investigated.

2.1.1. Learning Context and Child L2 Morphosyntactic Acquisition

For children and adolescents learning in either the immersion or community-L2 context, opportunities to use the L2 should be expected to differ according to context. It is also expected that such differences, *e.g.*, frequency of language use, or how and with whom a language is used, can impact what individuals know of their languages (see Grosjean, 2016). Thus, if the L2 morphosyntactic ability that learners develop differs between immersion and community-L2 contexts, then the explanation should lie in disparities in L2 experience between the contexts. As such, L2 learning context is essentially a macro-variable that impacts a range of learner-external factors that, in turn, impact L2 ability.

However, as cross-context comparisons are under researched, it is unclear which of the elements that comprise an L2 learning context impact L2 morphosyntactic ability.

Despite broad agreement that children are more likely than adults to ultimately mirror the target language ability of L1 speakers (DeKeyser, 2012; Muñoz & Singleton, 2011), it is also recognized that this only generalizes to children with 'naturalistic' L2 input (Muñoz, 2014). The assumed benefits of an early start and naturalistic input were key to the theoretical motivation behind early immersion programs in Canada and the US (Fortune, Tedick, & Walker, 2008; Genesee, 1984). When these programs were begun, the expectation was that children in immersion could rely on their natural incidental language learning mechanisms to develop L1-like proficiency in an L2 as they learned content through authentic communicative language use. Thus, L2 outcomes for immersion were expected to be similar to outcomes for community-L2 exposure with learners in both contexts achieving target language ability that parallels L1 speakers. Subsequent research has shown that this assumption was not borne out, as, despite strong communicative proficiency, uniformly L1-like L2 abilities do not result from the immersion context regardless of learners' length of immersion exposure or starting age (Genesee & Lindholm-Leary, 2013). In Canadian French immersion where children begin at five years old in a kindergarten to grade 12 system, there are significant gaps in the grammatical accuracy of learners' L2 production with non-target L2 use remaining evident into sixth grade (the seventh year of immersion; Cummins, 2009). Similar gaps can be found in immersion programs with languages such as Irish (Ó Duibhir, 2011) or Spanish (Montrul & Potowski, 2007).

Though the L2 morphosyntactic ability of immersion learners can be noticeably non-target (Kowal & Swain, 1997; Lyster, 2007), studies of community-L2 learners in adulthood have also shown that detectable but subtle differences between L1 speakers and early L2 speakers can last into adulthood (Flege, Yeni-Komshian, & Liu, 1999; McDonald, 2000). However, when not scrutinized by experimental tasks, differences in morphosyntactic ability for community-L2 learners do not seem to be as apparent as those for immersion learners, though there is little research to indicate how these groups compare. Where differences exist, these should result from the L2 experience learners have within each context. However, precisely how these differences in L2 experiences impact L2 morphosyntactic ability remains unclear.

2.1.2. Predictors of Individual Differences in Child L2 Morphosyntactic Ability

Based on the research discussed in the previous section, it could be expected that, when compared at the group level, the L2 morphosyntactic ability of immersion learners would differ from that of community-L2 learners. While differences according to L2 learning context may be expected, learning context as a macro-variable would not explain them fully, as context only indexes a range of more proximal micro-variables that impact L2 development, such as opportunities to use the L2 outside school, or the degree to which learners need to rely on cognitive abilities for L2 learning. For example, it may be expected that community-L2 learners use the L2 more often at home as they are more likely to have family members who are also proficient in the majority language of the community in which they all live. However, it cannot be assumed that all community-L2 learners use the L2 more at home more than all immersion learners who may also have parents and siblings

proficient in the L2. As a result, the differences in L2 exposure between contexts may not always be categorical. Therefore, it is important to also consider context from a more granular perspective by examining the individual experiences or properties of learners within their respective contexts. By examining these factors across and within contexts, we can understand both which factors impact child-L2 development and how these factors are characteristic of each L2 context.

2.1.2.1. Learner-External Predictors of Morphosyntactic Ability across Child L2 Contexts

Learning context has a transparent connection to the learner-external factors of L2 experience that individual L2 learners have. One of the most salient differences between contexts can be in the quantity of L2 exposure that learners receive, and as noted, there are likely to be differences in this regard between community-L2 learners and immersion learners. In the early stages of L2 acquisition, a greater length of L2 exposure is consistently found to result in better L2 proficiency in community-L2 contexts (Blom, Paradis, & Sorenson Duncan, 2012; Chondrogianni & Marinis, 2011; Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017) and in immersion ones (Kalia, Lane, & Wilbourn, 2018; Nicolay & Poncelet, 2013; Swain & Lapkin, 1982). This additional input is likely an important factor in the much greater L2 communicative proficiency that immersion learners achieve when compared to peers in traditional L2 classrooms where the L2 is the topic of instruction (Genesee & Lindholm-Leary, 2013; Swain & Lapkin, 1982). While instructed-L2 learners may only have a few hours of L2 contact a week, in immersion programs, the L2 is used for content instruction for anywhere between half to all of the instructional time which can result in children amassing approximately 6,000 hours of

classroom L2 exposure by the end of eighth grade (Turnbull et al., 1998). Nevertheless, learners within either the immersion or community-L2 context should be expected to differ in their cumulative L2 exposure, defined as the proportion of waking hours that an individual is exposed to the L2 (Unsworth, 2013). In comparison to monolinguals, Scheidnes (2020) estimated that learners in Canadian French immersion receive approximately 16% of the exposure to French that L1 French speakers receive over the same time span. Similarly, it should be expected that immersion learners receive less cumulative L2 input over time as community-L2 learners, though community-L2 speakers will not have a quantity of L2 exposure equivalent to a monolingual. Despite the benefit of L2 exposure, its impact appears to eventually diminish over time with findings for both community-L2 learners (Jia & Fuse, 2007; Paradis, Tulpar, & Arppe, 2016; Soto-Corominas, Paradis, Rusk, Marinova-Todd, & Zhang, 2020) and learners in immersion (Fortune & Tedick, 2015; Hart, Lapkin, & Swain, 1991) indicating that beyond a certain threshold, the benefit of more L2 input begins to plateau.

A key difference between these L2 contexts is in the access learners have to sources of L2 input outside of school. For community-L2 learners, research has shown that the richness of the L2 environment, often characterized by things such as the extent to which a child uses the L2 to participate in extracurricular activities, read or listen to other media, or interact with peers, also impacts L2 morphosyntactic knowledge (Jia & Fuse, 2007; Paradis, 2011; Paradis, Soto-Corominas, Chen, & Gottardo, 2020; Rojas et al., 2016). In addition, community-L2 learners may have more opportunity to use the L2 at home as they are more likely to have family members proficient in the L2. However, this type of exposure may not benefit L2 development of morphosyntax (Hoff & Ribot, 2017; Place & Hoff, 2016), though

this may depend on the L2 proficiency of the family members with whom the child interacts (Sorenson Duncan & Paradis, 2020; Unsworth, Brouwer, Bree, & Verhagen, 2019). For example, Sorenson Duncan & Paradis (2020) found that more L2 interaction with older siblings was associated with stronger L2 morphosyntactic skills, but this effect was not found for more L2 interaction with mothers (see also Paradis et al., 2020).

For immersion learners, little research exists for their experiences with the L2 outside of school, with a specific lack of research examining the impact of these factors directly on morphosyntactic ability. Nevertheless, existing evidence indicates that immersion learners' general L2 proficiency and/or comprehension benefits from more L2 use outside the classroom (Genesee, 1978; Hamayan, Genesee, & Tucker, 1977). A potential reason that learner-external factors are under researched is that it is considered characteristic of immersion contexts that L2 input and use are limited to the immersion classroom (Swain & Johnson, 1997). However, this description of immersion pre-dates widespread access to the Internet which should vastly increase a learner's access to a wide range of L2 input. Due to this increasing access to an L2, the benefit of informal sources of L2 input, where learners get L2 exposure outside of instructional settings, has been of increasing interest for non-immersion foreign L2 learners (Dressman & Sadler, 2020). However, input factors such as these that may have an impact on immersion learners have remained under studied.

2.1.2.2. Learner-Internal Predictors of Morphosyntactic Ability across Child L2 Contexts

Despite the more transparent connection between learner-external factors and learning context, interactions between learner-internal differences and L2 input and experience

have been demonstrated by studies in instructed-L2 contexts. Specifically, studies show that the type of instructional intervention mediates the contribution of cognitive differences that impact language acquisition (Erlam, 2005; Hwu & Sun, 2012; Li, 2015; Yilmaz & Granena, 2019). In an overview of this issue, Li, Ellis, & Zhu (2019) argue that in contexts where there is less explicit L2 support, the impact of differences in cognitive ability may become more evident. Li et al specifically identified language analytic ability, defined as the ability to extrapolate L2 rule information, as being important. This language aptitude ability has been proposed to have a common underlying factor with general reasoning abilities (Sasaki, 1996) and thus, may be associated with nonverbal reasoning. Though immersion contexts were originally expected to encourage implicit L2 learning making explicit instruction of L2 elements such as grammatical rules unnecessary (Lyster & Tedick, 2014), current immersion pedagogy widely recognizes the necessity of instruction that systematically draws attention to L2 forms alongside teaching classroom content (Lyster, 2007; Lyster & Tedick, 2014). This suggests that differences in nonverbal reasoning may be important for immersion learners and is supported by research showing that the language aptitude measure of language analytic ability, along with memory abilities, correlates with vocabulary knowledge and listening comprehension for early-start immersion learners in their twelfth year of study (Harley & Hart, 1997). Similarly, the general L2 proficiency or comprehension of students in their second year of early immersion may be impacted by differences in nonverbal reasoning ability (Genesee & Hamayan, 1980).

Learner-internal factors also impact the rate of L2 morphosyntactic development for community-L2 learners. Impacting factors include differences in cognitive abilities such as

nonverbal reasoning (Paradis, 2011; Paradis et al., 2017, 2020) and verbal memory (Paradis et al., 2016; Verhagen & Leseman, 2016; Verhagen, Leseman, & Messer, 2015). In addition, linguistic knowledge the child already has can also have an impact. For example, whether an L2 learner's L1 is isolating, or inflecting has been shown to affect the rate at which child community-L2 learners develop particular morphosyntactic constructions, with speakers of inflecting L1s acquiring L2 inflection more rapidly than those whose L1 is an isolating language (Blom et al., 2012; Paradis, 2011). Greater accumulated L2 vocabulary knowledge can also impact, as children with larger receptive vocabularies have been found to have better L2 morphosyntactic ability (Blom et al., 2012; Paradis et al., 2016; Rezzonico, Goldberg, Milburn, Belletti, & Girolametto, 2017). These findings indicate that language knowledge is integrated across different language subsystems.

2.1.3. Morphosyntactic Acquisition in Immersion

In the case of Canadian French immersion, programs were initiated both by and in response to parents who were dissatisfied with the L2 outcomes of traditional instructed-L2 programs (Turnbull et al., 1998). The initial motivation for immersion programs was the belief that meaningful classroom interaction would create a sufficiently naturalistic L2 environment that would allow L2 acquisition to take place in the same manner as L1 acquisition (Genesee, 1984). Essentially, these programs were designed in order to get the L2 results observed for community-L2 learners via classroom exposure. In comparison to instructed-L2 learners, the L2 abilities of immersion learners rapidly outpace instructed-L2 peers (Genesee & Lindholm-Leary, 2013; Swain & Lapkin, 1982), and thus immersion has largely been successful in improving L2 outcomes. However, despite the strong

communicative ability immersion learners develop, research has shown that L2 ability of immersion learners does not eventually parallel L1 speakers (Genesee & Lindholm-Leary, 2013), particularly in morphosyntactic ability (Harley, Cummins, Swain, & Allen, 1990; Lyster, 2007), and it is unclear how they compare directly to community-L2 learners.

Specifically for French immersion, one of several notable areas of low morphosyntactic accuracy is with verb forms (Lyster, 2007). For some verb forms, an insufficient quantity of input may explain non-target usage. Research evidence shows that French immersion teachers from grade three to six produce a verb distribution heavily skewed toward present and imperative forms with past, future, and conditional being much less frequent (Harley, Allen, Cummins, & Swain, 1987; Izquierdo, 2007) with the bias in exposure suggested to impact accuracy (Harley & Swain, 1984). Swain (1988) pointed out that it is likely that immersion learners are never exposed to many verbs in less frequent forms like the conditional even into the later stages of immersion. Thus, she proposed 'restricted input' as a partial explanation for inaccurate verb marking with less frequent forms.

Naturally, if immersion learners are never or very rarely exposed to particular morphosyntactic constructions, they are unlikely to acquire them. However, non-target ability is also observed in French immersion for L2 constructions that are highly frequent in the L2 input (Allen, Swain, Harley, & Cummins, 1990). For learners in their seventh year of immersion, examples include incorrect grammatical gender marking on articles, adjectives or pronouns; omission, addition, or incorrect number marking for articles; and errors with more frequently used main and auxiliary verbs (Harley et al., 1987). It has been found that in their seventh year of French immersion, learners produce accurate verb

forms for only around 60% of verbs while age-matched L1-French speakers do so well over 90% of the time (Harley et al., 1990). These difficulties with verb form accuracy can last into the later immersion grades, as well (Hart et al., 1991), with non-target ability persisting into the final years of immersion education for morphosyntax in general (*i.e.*, grades 11 and 12, Wesche, 1993; Harley & Hart, 1997; Turnbull et al., 1998).

Currently, there is little research that directly compares the morphosyntactic ability of immersion learners with community-L2 peers. In one study that compared the sentence repetition accuracy of early French immersion learners with similar-aged simultaneous and sequential community-L2 learners of French, the immersion learners had still not achieved ability on par with their community-L2 learning peers (Scheidnes, 2020). However, these learners were only mid-way through their second year of French immersion schooling, and thus, it remains unknown how they would compare after more years of exposure. Though it should be expected that immersion learners continue to improve with more exposure, the benefit of length of exposure is subject to diminishing returns beyond the first few years (Genesee, 2006), with morphosyntactic development potentially reaching a plateau by the sixth year of immersion (Fortune & Tedick, 2015).

The research discussed here relies heavily on Canadian French immersion, a context for which extensive research exists (Cummins, 2009). Less research exists for other immersion contexts including English immersion contexts for L1-Mandarin speakers. However, existing research with these learners shows they also develop stronger L2 proficiency than instructed-L2 peers (Cheng, Li, Kirby, Qiang, & Wade-Woolley, 2010; Knell, Siegel, & Lin, 2007) and are significantly more confident in their ability to use English to communicate (Knell & Chi, 2012). Still, little is known about their development of English

morphosyntax. In contrast, more research for English morphosyntactic development exists for L1-Mandarin English learners in community-L2 contexts.

2.1.4. Morphosyntactic Acquisition in Community-L2 Contexts

Young community-L2 learners are commonly observed to closely parallel the target language abilities of their peers who speak the target language as an L1. However, studies of community-L2 learners in adulthood have shown that how closely L2 morphosyntactic ability parallels L1 speakers can depend on language learning aptitude (Abrahamsson & Hyltenstam, 2009); the learner's L1 (McDonald, 2000); and/or external language exposure factors (Flege et al., 1999). In one study of child community-L2 learners in adulthood, McDonald (2000) examined the L2 English morphosyntactic ability of L1-Spanish and L1-Vietnamese learners of L2 English. The study found that while the L1-Spanish speakers were not different from an L1-English group for grammaticality judgment accuracy, the L1-Vietnamese group were significantly different for four of the 12 stimulus types. All four of these stimulus types were grammatical morphemes, including past tense, plurals, third person verbs, and articles - constructions for which Vietnamese, an isolating language, differs from English. Similarly, research for children learning English as a community L2 has shown that there can be a lag in development of grammatical morphemes depending on whether a child's L1 is an inflecting language (*e.g.*, Arabic, Punjabi, Spanish, Urdu) or is an isolating one (*e.g.*, Vietnamese, Cantonese, Mandarin). Blom et al. (2012) and Paradis (2011) found that children with inflecting L1s had higher accuracy with verb morphology than their peers with isolating L1s within their first few years of L2-English exposure.

These isolating-L1 community-English learners showed protracted developmental trajectories, with non-target ability apparent even after three years of exposure to the L2.

In a study of L1-Chinese community-L2 learners with longer L2 exposure, Paradis et al. (2016) found that the lag in verb morphology development at earlier stages persisted. Many of the participants of this study fell short of verb form accuracy criterion scores for which monolingual English-speaking children typically perform at ceiling for by six to seven years of age. Despite participants being ten years and five months of age with six years and four months of L2-English exposure on average, 11 of the 18 participants still fell short of a criterion score for at least one of the sub-tests. The results of this study show that there can be variability in the attainment of early community-L2 learners even after many years of exposure. Similar results were found in a study of L1-Mandarin learners' development of English morphosyntax. Learners tracked over five years varied in whether they were able to supply English grammatical morphemes in more than 80% of obligatory contexts (Jia & Fuse, 2007). By the end of the five-year study, only one of the seven grammatical morphemes studied had been mastered by all participants (the bound *-ing* affix). Other bound tense/agreement morphemes, along with plural *-s*, showed variable mastery, with no participant reaching criteria for regular past tense marking. Further, participants showed plateauing development, suggesting that many of the participants were unlikely to reach the 80% criterion in the future. Thus, research evidence shows that community-L2 learners often perform similarly to L1 speakers on experimental tasks, but this can be impacted by L1 (Blom et al., 2012; McDonald, 2000), and individual variation should be expected, even after long-term community-L2 exposure (Jia & Fuse, 2007; Paradis et al., 2016).

2.1.5. The Present Study

Currently, the impact of learning an L2 in an immersion context as compared to a community-L2 context is under studied. Existing research supports the expectation that language exposure in immersion classrooms results in strong communicative L2 proficiency, but also that morphosyntactic abilities differ from those of L1 speakers. Existing research also supports the expectation that community-L2 learners can become indistinguishable from their L1 peers for many aspects of the L2, though subtle differences can still persist. However, very little is known about how these L2 groups compare directly to each other or how either group compares to L1 speakers after many years of exposure.

In order to better understand how the L2 morphosyntactic ability of early L2-English learners develops in each of these two contexts, the present study compared three groups of adolescent English speakers. These groups were 1) L2 learners who had acquired English as a foreign language in immersion classrooms; 2) L2 learners who had acquired English in community-L2 contexts; and 3) L1-English monolinguals as a control group. Both L2 groups were L1-speakers of Mandarin. A grammaticality judgment task (henceforth just judgment task) was used in order to measure differences in English morphosyntactic ability between participant groups and capture individual differences within groups. The task was designed specifically to probe L2 ability across a range of grammatical morpheme constructions, including those in the noun phrase (articles, plural marking), and in the verb phrase (past tense and subject-verb agreement marking with BE, DO, and present third-person singular '*-s*'). These constructions reflect differences between Mandarin and English in that Mandarin uses neither bound inflectional morphemes for tense or number, nor free morphemes to mark the definite/indefinite distinction on nouns (Li & Thompson, 1989).

Early L2 learners have been shown to be highly accurate on judgment tasks (*e.g.*, Birdsong & Molis, 2001; Johnson & Newport, 1989). Given that the participants of all groups were expected to have high English proficiency, it was important that the task was challenging enough to prevent ceiling effects, which, when present, can obscure both group differences, and individual differences in ability within groups. Therefore, in order to provide a range of stimulus difficulty on the task, stimuli where grammatical morphemes were presented within sentences with nonadjacent agreement were also included. Previous studies with L2 speakers have shown nonadjacent agreement increases task difficulty (*e.g.*, Song, 2015; Keating, 2009), possibly due to less reliance on morphosyntactic structure than L1 speakers during processing (Clahsen & Felser, 2018). Thus, two levels of stimulus difficulty were created to capture a range of L2 ability, potentially including some morphosyntactic ability still under development. The specific research questions and corresponding predictions for each are as follows:

1. Are there differences between each participant group in accuracy on the judgment task? Does the pattern of group differences change for morphosyntactic constructions with an additional level of difficulty?

   Prediction: Though all groups are expected to have high English proficiency, given previous research that shows a) accuracy with L2 English grammatical morphology can take an extended period to develop for speakers of isolating languages, and b) immersion learners generally demonstrate low morphosyntactic accuracy, it is expected that the L1-English control participants will have the highest accuracy, particularly for the more difficult constructions, while the immersion participants will have the lowest accuracy, with the community-L2s in between the other two groups. In addition, all groups are expected to be less accurate for morphosyntactic constructions with an additional level of difficulty.
2. How do differences in individual learner-internal and learner-external characteristics influence the L2 learners' accuracy on the judgment task? Are there differences in how these characteristics shape accuracy on the task between the immersion and community-L2 groups? Are there differences in how these characteristics impact accuracy for morphosyntactic constructions with an additional level of difficulty?

Prediction: Given prior research with community-L2 learners and immersion L2 learners, it is expected that multiple factors will predict individual variation in task performance in both groups. However, it is anticipated that the factors could diverge in how they predict individual differences across learning context. In addition, the two levels of stimulus difficulty may reflect different levels of L2 development, and therefore, different factors may be predictive for stimuli across levels of difficulty.

2.2. Method

2.2.1. Participants

The three participant groups for this study were comprised of middle-school-aged adolescents studying in either grade seven, eight, or nine of a K to 12 system (mean age = 13;9, *SD* = 9 months). The first group was the immersion-L2 group. Participants for this group were Taiwanese L1-Mandarin L1, L2-English immersion students (*n* = 29) studying at a private English-language high school in Northern Taiwan. The additional two groups were comprised of adolescents living in Canada and attending Canadian schools in large English-majority cities. These groups were the community-L2 group (*n* = 25), and the English L1 control group (*n* = 29). For inclusion in the community-L2 group, participants were required to have Mandarin as their L1, and parents who were Mandarin-speaking L2 speakers of English. Though community-L2s also reported that their current language use at home included English to varying degrees, these participants were considered to be child L2 learners as opposed to simultaneous English-Mandarin bilinguals. The L1-controls spoke only English at home and at school. Participants were recruited through their schools with some L1 participants recruited by word of mouth.

Participants in each of the L2 groups (immersion-L2 and community-L2) had begun their education in English at or before six years of age in their respective educational contexts (*i.e.*, either exposure in English-medium classrooms in an English-majority

community or in English-language immersion in a Mandarin-majority community). Though parents and participants of both L2 groups were Mandarin-speakers, three participants from each L2 group also reported exposure from birth to an additional Chinese language. None of these participants were outliers in terms of their performance on the judgment task or vocabulary measure, and therefore were not excluded from the study.

The immersion-L2 participants had been exposed to English in an immersion environment taught by foreign teachers from an English-speaking country (*e.g.,* the UK, US, Canada, etc.) since preschool (mean age of English education onset = 4;4, *SD* = 1;3). After preschool, the students attended a bilingual elementary school where half their instruction was in Mandarin from a local teacher, and half in English from a foreign teacher. On average, they had attended an immersion program for eight years and six months (*SD* = 17 months). From the beginning of grade seven, all immersion-L2s studied entirely in English. The community-L2 participants had a mean age of English onset of 3;8 (*SD* = 0;11) and a mean length of English exposure of 10 years and two months (*SD* = 13 months). For the community-L2s, the mean age of arrival in Canada was 3;2 (*SD* = 1;7), though many were born in Canada.

2.2.2. Procedures and Materials

The experimental task for the present study was a timed judgment task, the results of which were also analyzed against learner-internal (*i.e.*, cognitive abilities or language knowledge) and learner-external (*i.e.*, language exposure) factors. Measures and data collection procedures are described below.

Judgment Task. A timed auditory judgment task was created and presented using E-Prime 2.0 (Psychology Software Tools, 2012). Stimulus sentences were recorded as spoken by a male speaker of North American English. The task was presented to participants as a timed task since doing so limits participants from making explicit reflections on morphosyntactic structure (Ellis, 2005; Godfroid et al., 2015). Participants were instructed to respond whether the form of the sentences sounded 'good' or 'bad', and to respond as quickly as they could.

Stimulus items consisted of 42 grammatical/ungrammatical stimulus pairs counter-balanced across lists. In addition, there were 12 fillers; all of which were grammatical. With the fillers, each list had 33 grammatical items and 21 ungrammatical items. Altogether, the task included six separate grammatical morpheme constructions. As an additional measure to prevent participants from attempting to use explicit knowledge of grammatical rules, distractor stimuli involving odd adverb placement were also included. The adverb items were not analyzed. As mentioned previously, the grammatical morpheme constructions included articles, regular past tense, plural marking, and three stimulus sentence types that focused on subject-verb agreement: copular or auxiliary BE, auxiliary DO in constructions with 'not' (*i.e.*, *do not* or *does not*) and third-person marking. Each grammatical morpheme construction had six grammatical/ungrammatical pairs in the set of stimuli.

This study included participants who were expected to be highly proficient English speakers. Thus, in order to prevent ceiling effects that could mask between-group differences, stimuli for the subject-verb agreement morpheme constructions (BE, auxiliary DO and third-person marking) were presented in sentences with nonadjacent agreement. Nonadjacent agreement has been found to be challenging for both L2 learners (Keating,

2009; Song, 2015) and L1 speakers (Bock & Miller, 1991). For nonadjacent subject-verb agreement stimuli, the finite verb was adjacent to a noun other than the one with which it should agree (*e.g.*, 'The problems with the **computer** *were* fixed by the technician', where the auxiliary *were* agrees with *problems* rather than *computer*). The article, past tense, and plural marking stimulus items used were similar to those used by Johnson & Newport (1989) and expected to be easier than the subject-verb agreement stimuli. A complete list of the stimuli for this study are presented in Appendix A.

The six grammatical morpheme types were grouped into two conditions according to anticipated difficulty level. These were labelled the 'Easy' and 'Difficult' conditions, the latter of which included nonadjacent subject-verb agreement. Grouping stimuli was done to increase statistical power which is sensitive to the number of observations per condition (Brysbaert, 2019). Thus, pooling observations increased the probability of detecting a true effect. Because data collection took place at school during participants' class time, increasing the number of observations by increasing the total number of items was unfeasible given the time constraints.

Though the study employed stimuli with grammatical morphemes and nonadjacent agreement in order to make the task more challenging, based on findings from prior research, how the properties of the stimuli contributed to the difficulty of the task was beyond the scope of the study.

English Vocabulary Size. The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) was used to assess receptive vocabulary. For this test, participants hear a word and then indicate the appropriate one of four pictures presented. A typical monolingual age-standardized score is 100, with a standard deviation (*SD*) range of 85-115.

Verbal Memory. The Comprehensive Test of Phonological Processing – Non-word repetition sub-test (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 2013) was used to measure phonological short-term memory abilities. The test requires participants to repeat nonsense words based on English phonotactics that progressively increase in syllable length and complexity. The age-standardized mean for this test is 10 with an *SD* range of seven to 13.

Nonverbal Reasoning. The Kaufman Brief Intelligence Test, 2nd Edition, Matrices test (KBIT; Kaufman & Kaufman, 2004) was used to assess nonverbal, analytic reasoning. For this test, participants are asked to select one picture from an array of five or six pictures that best completes a relationship represented within a grid matrix. Responses for each item were coded as correct or incorrect and then an age-referenced standard score was calculated for individual participants. The age-standardized mean for this test is 100 points with an *SD* range of 85-115.

Alberta Language Environment Questionnaire - 3 (ALEQ-3; Soto-Corominas et al., 2020) All participants completed the ALEQ-3, which consisted of two parts. The first was given to parents along with the consent forms. Parents filled out this form independently, providing information about each parent's educational background, as well as where and when their child began education in an English context. Age of beginning English education is reported for all groups but represents a different variable for the L1 and L2 groups. For L2 groups, age of beginning English education is used as the age at which participants began English exposure. However, the L1 controls would already be fluent speakers of English by the time they entered preschool. Data from parents was used to calculate length of English exposure. For the community-L2 group, the variable was calculated by

subtracting age of English education from age at testing. For the immersion-L2 group, length of exposure represents years in English immersion classes. Parental education was blocked on a three-point scale from 0 to 1 with 0 for no post-secondary education, .5 for some post-secondary and 1 for an undergraduate degree or more education. Parental education, which served as a proxy for socioeconomic status (Hoff & Tian, 2005), was high for all groups, but there was little variability for the L2 groups and thus this variable was not analyzed as originally intended.

Participants were interviewed directly for the second part of the ALEQ-3 with the researcher asking questions and filling out the questionnaire. Participants in Canada were asked about the language that they were first exposed to at home (*i.e.*, their L1) and divided based on this response into the L1-English control group or community-L2 group. Participants in Taiwan were placed in the immersion-L2 group. For L2 participants, a range of questions probed current and cumulative language exposure factors. Participants were asked about whether they tended to use their L1 or English more with their parents and siblings and this information was converted into a variable reflecting home language use. A similar variable was created for language use with friends/peers. These variables were reported on a 1-5 scale (1 = English never and L1 always (90% or more), 2 = English seldom and L1 usually (30%/70%), 3 = English 50% and L1 50%, 4 = English usually L1 seldom (70%/30%), 5 = English almost always (90% or more) and L1 almost never). In addition, they were asked about whether they felt more proficient in Mandarin or English, and this information was used to create a relative language proficiency variable, also on a 1-5 scale with 1 indicating greater Mandarin proficiency, 5 indicating greater English proficiency, and 3 indicating balanced proficiency. Participants were also asked to rate

their L1 proficiency independent of their L2. This was also on a 1-5 scale where 1 would indicate full L1 attrition and 5 full L1 proficiency. In addition, L2 participants reported activities in English outside of school. These were collected into a single variable that included time spent using English for watching or listening to media, reading, using social media, playing video games with language content, or participating in extracurricular activities. This was reported in hours per week (English hours).

2.3. Results

Analyses to address the research questions were done using *R* Statistical Software (version 3.6.0; R Core Team, 2018). Data gathered about the participant groups are presented in Table 2.1. The table provides both descriptive statistics and statistical group comparisons. Comparisons show several statistically significant differences. For each of these, an advantage was found for the community-L2 compared to the immersion-L2 group. The community-L2s had longer exposure to English, used English more often with peers and at home, had larger receptive vocabularies, higher verbal memory scores and rated their English proficiency as higher than their L1 Mandarin. The immersion-L2 group rated their Mandarin fluency higher than did the community-L2s.

For analysis of each research question, mixed-effects regression models were run on participant accuracy for the judgment task using functions from the *lme4* package (version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015). Correct responses on the judgment task included correct acceptance of grammatical stimuli and correct rejection of ungrammatical stimuli, while incorrect responses were acceptance of an ungrammatical stimulus, rejection of a grammatical stimulus, or failure to enter a response before advancing to the next trial.

Trials where no response was given accounted for 0.86%, 0.44%, and 0.19% of the total responses for the immersion-L2, community-L2, and L1 control groups, respectively. For modelling, all categorical predictor variables for were sum coded as opposed to *R*'s default treatment coding. All numeric predictor variables were centred and standardized as *z*-scores. Data from one immersion-L2 and one community-L2 participant were not included as an examination of the data indicated that more than half of these participants' responses were made within 100 milliseconds of the stimulus audio onset. Thus, it was unlikely that the participants were responding to the experimental stimuli. These participants are not included in the group counts. Fast responses did not appear to be an issue for any other participants.

Table 2.1 Descriptive Statistics for Participant Groups

	Immersion-L2	Community-L2	L1 Controls	Group Comparisons
n	29	25	29	
n = female	21	10	10	$\chi^2(2) = 9.87$, $p = .007$ Com. & L1 < Imm.
parental education	.83 (.33)	.84 (.28)	.67 (.31)	$\chi^2(2) = 7.00$, $p = .03$ All corrected tests N.S.
age at testing	13;9 (0;9)	13;10 (0;6)	13;9 (0;10)	$F(2, 80) = 0.21$, $p = .81$
age of English education	4;4 (1;3)	3;8 (0;11)	3;10 (0;10)	$\chi^2(2) = 3.82$, $p = .148$
Learner-External Factors				
length of Eng. exposure	101.79 (16.55)	121.8 (12.93)	164.79 (10.15)	$\chi^2(2) = 57.11$, $p < .001$ Imm. & Com. < L1
peer language use	2.17 (0.89)	4.48 (1.05)	-	$t(47.45) = 8.66$, $p < .001$
home language use	1.3 (0.48)	2.77 (1.28)	-	$t(29.89) = 5.42$, $p < .001$
English hours	41.83 (26.93)	51.63 (21.68)	-	$t(51.78) = 1.48$, $p = .145$
Learner-Internal Factors				
receptive vocabulary	77.52 (16.09)	107.28 (17.74)	111.41 (15.73)	$F(2, 80) = 35.94$, $p < .001$ Imm. < Com. & L1
relative proficiency	2.87 (0.68)	3.82 (0.74)	-	$t(49.03) = 4.87$, $p < .001$
L1 fluency	4.38 (0.49)	3.68 (1.28)	-	$t(30.11) = 2.57$, $p = .015$
nonverbal reasoning	112.21 (10.55)	116.04 (12.24)	100.48 (12.9)	$F(2, 80) = 12.79$, $p < .001$ L1 < Com. & Imm.
verbal memory	8.03 (1.7)	11.54 (1.79)	9.59 (2.21)	$\chi^2(2) = 31.42$, $p < .001$ Imm. < L1 < Com.

Note. Summary data are presented as mean (*SD*). Age data are given in year and month (yy;mm) with SDs in months. L1 fluency, home language use, peer language use, and relative language proficiency are on scales from 1 to 5, with 5 equaling the highest L1 fluency, the most English use, and highest self-rated English proficiency, respectively. For number of female participants per group, statistical comparison is based on a logistic regression. For other comparisons of three groups, ANOVA (*F)* with a Tukey HSD post-hoc test was used unless at least one group had a non-normal data distribution or groups had unequal variances in which case a Kruskal-Wallis (χ^2) test with a Bonferroni corrected Dunn's post-hoc test. Results of statistically significant post-hoc pairwise comparisons

adjusted for three comparisons are shown with a < indicating the significantly lower group(s). Welch's *t*-tests (*t*) were used for comparisons of two groups.

2.3.1. Research Question 1: Group Comparisons for Morphosyntactic Abilities

The purpose of the first research question was to examine group differences in morphosyntactic ability for three participant groups: the immersion-L2s, community-L2s, and L1 controls. In order to compare group accuracy, individual participants' responses were grouped into the two stimulus conditions (Easy or Difficult). These were based on the presence of an additional level of difficulty introduced by including nonadjacent agreement, as discussed previously. For these two conditions, a *d*-prime score was calculated for each participant. In signal detection theory, *d*-prime scores account for differences between ability to detect stimuli with errors versus ability to accept stimuli without errors (Macmillian & Creelman, 2005). For the present data, *d*-prime scores can account for biases to either accept or reject a stimulus without regard to its acceptability. Accuracy and response time data are provided in Table 2.2, though response times were not further analyzed.

Table 2.2 Judgment Task Accuracy and Response Times

	Type		Immersion-L2	Community-L2	L1 Controls
Accuracy	Easy	*M* (*SD*)	0.31 (0.68)	1.45 (0.83)	1.89 (0.63)
		Range	-0.75-2.54	-0.24-2.98	0.00-2.98
	Difficult	M (*SD*)	-0.03 (0.56)	0.21 (0.53)	0.70 (0.67)
		Range	-1.06-1.59	-1.06-1.06	-0.51-2.23
Response Time	Easy	M (*SD*)	4244 (1157)	4109 (980)	3793 (868)
		Range	1089-7573	2035-8495	1565-6949
	Difficult	M (*SD*)	4532 (1240)	4580 (1283)	4053 (985)
		Range	2232-8220	2179-9539	2083-7225

Note. Accuracy scores and response times are given for each group and for each stimulus condition. Data are given as means (*SD*) with ranges below. A *d*-prime score of 0 indicates

chance performance with higher scores indicating greater accuracy. Response times are based only on correct responses to ungrammatical stimuli (*i.e.*, response times to detected errors) given in milliseconds. Further, in order to give means unaffected by outliers, responses time data were trimmed for outliers that were 2.5 *SD* from the mean within each group and stimulus condition. For accuracy, all responses were included.

A linear mixed effects model was used to analyzed participants' *d*-prime scores for each stimulus condition (Easy or Difficult) in interaction with participant group (Group). Random intercepts for individual participants (Subject) were also included. Results for this model are presented in Table 2.3. Functions from the `emmeans` package (version 1.4.6; Lenth, 2019) were used to perform post-hoc pairwise comparisons Tukey-corrected for three comparisons. These comparisons are presented in Table 2.4 and Figure 2.1. Pseudo-R^2 values for this model were provided by the `MuMIn` package (version 1.43.17; Bartoń, 2020). The fixed effects for the model had an R^2 of .53 and inclusion of the random effect for Subject gave an R^2 of .70.

Table 2.3 Model Results: Judgment Task Accuracy

	estimate	std.error	statistic	df	*p*-value
(Intercept)	0.75	0.06	12.65	80.00	<.001
Group 1 (Immer.-L2)	-0.62	0.08	-7.42	80.00	<.001
Group 2 (Comm.-L2)	0.08	0.09	0.88	80.00	.382
Stimulus Condition 1 (Easy)	0.46	0.04	11.27	80.00	<.001
Group 1: Stimulus Condition 1	-0.29	0.06	-5.08	80.00	<.001
Group 2: Stimulus Condition 1	0.15	0.06	2.59	80.00	.011

Note. This table presents mixed-effects regression results for judgment task response accuracy *d*-prime scores. Group and stimulus difficulty level are sum coded. Group comparisons for each stimulus condition are shown in Table 2.4. The estimate column gives standardized estimates of regression coefficients.

Table 2.4 Group Comparisons by Stimulus Difficulty Level

comparison	estimate	std. err	df	*t*-ratio	*p*-value	Cohen's *d*
Easy Stimuli						
Immer.-L2 – Comm.-L2	-1.1	0.2	141.9	-6.3	<.001	1.51
Immer.-L2 - L1 Controls	-1.6	0.2	141.9	-9.2	<.001	2.42
Comm.-L2 - L1 Controls	-0.4	0.2	141.9	-2.5	.036	0.61
Difficult Stimuli						
Immer.-L2 – Comm.-L2	-0.2	0.2	141.9	-1.4	.352	
Immer.-L2 - L1 Controls	-0.7	0.2	141.9	-4.2	<.001	1.18
Comm.-L2 - L1 Controls	-0.5	0.2	141.9	-2.7	.022	0.79

Note. These pairwise group comparisons were produced using the `emmeans` package (Lenth, 2019) based on a linear model with *d*-prime scores. Each *p*-value is Tukey corrected for a series of three group comparisons. For Cohen's *d*, effects between 0.5 and 0.8 are 'medium' and those above 0.8 are considered large. These comparisons are also shown in Figure 2.1.

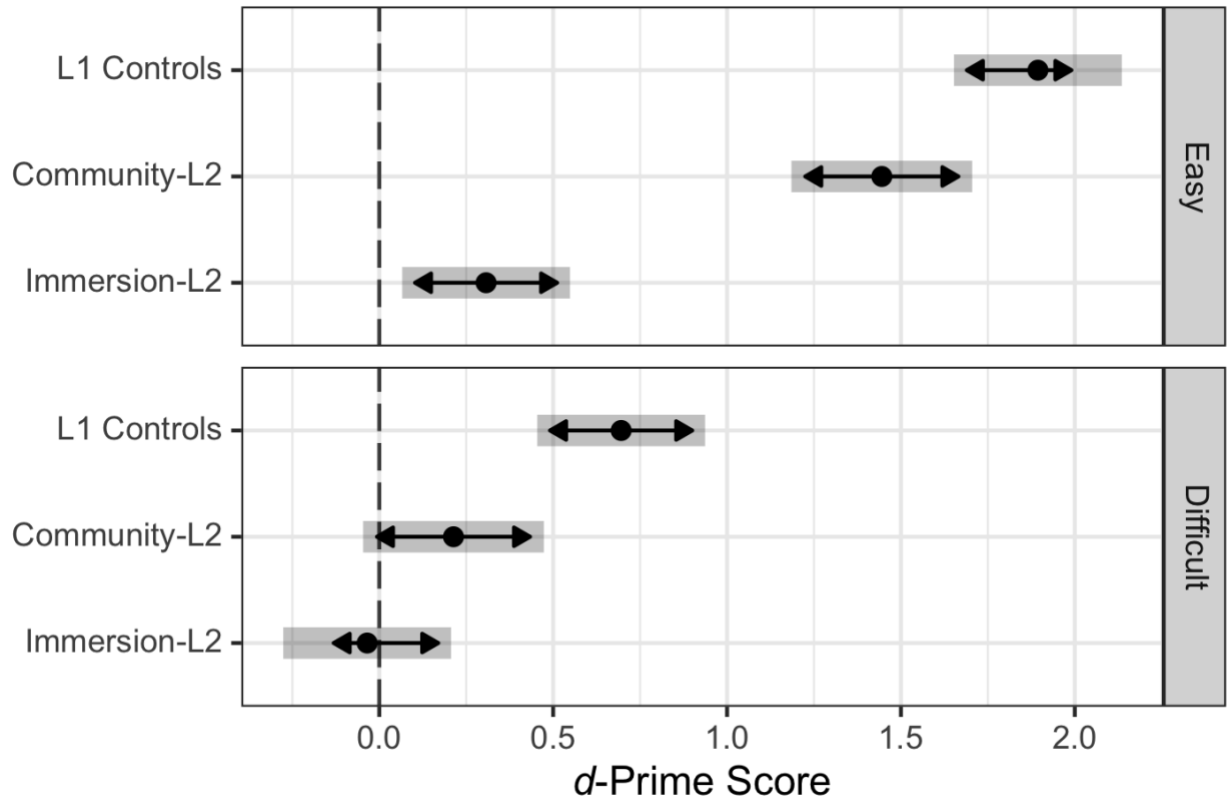

Figure 2.1. This plot shows the results of post-hoc Tukey-corrected pairwise comparisons between the three participant groups split for the two stimulus conditions Easy and Difficult. Shaded bars represent 95% confidence intervals. Mean d-prime scores for each group for each

condition are shown by the point on the double-arrow line. Where arrow ranges overlap, comparisons are not significantly different. Higher d-prime scores indicate greater accuracy while a score of zero (shown by the dashed vertical line) indicates chance performance.

Results of these comparisons (Table 2.3, Table 2.4 and Figure 2.1) show that immersion-L2s had consistently lower accuracy than L1 controls for both stimulus conditions, while they were only significantly below the community-L2s for the Easy stimuli. Differences between the community-L2 group and the L1 control group were significant for both stimulus conditions. However, Figure 2.1 shows that the community-L2s were much closer in accuracy for the Easy stimuli to the L1 controls than to the immersion participants, which is also reflected in the smaller effect size. The interactions presented in Table 2.3 show that each L2 group was significantly more accurate with the Easy stimulus items than Difficult ones.

2.3.2. Research Question 2: Sources of Individual Differences in L2 Morphosyntactic Ability

To address the second research question, mixed-effects logistic regression models were used to identify factors that predicted individual differences in accuracy in each L2 context. Logistic regression is commonly used with the binary choice data of judgment tasks (Schütze & Sprouse, 2014; Spinner & Gass, 2019) and allows for modelling each response separately making it possible to include the repeated measures of each stimulus item (Item) as a random effect in addition to the Subject random effect. The addition of random slopes for Item was also attempted for subject-level predictors but resulted in singular fits for the immersion group model. Added random slopes are noted below for the community-L2 group.

Variables were placed alternately into a model through a process of forward-fitting, retaining ones that were statistically significant. The models were then compared to a reduced model without the variable using a likelihood ratio test to see if the variable improved the model. In addition, concordance indices are reported. Concordance indices above .8 are considered to indicate the model has a good fit to the data (Chatterjee & Hadi, 2006). Correlations between variables are presented in Appendix B.1, and B.2, for the immersion-L2, and community-L2 groups, respectively. No two variables with $r > .7$ were included in models.

The aim of the second research question was to determine factors that predict individual differences in morphosyntactic ability in either context. Though the L2 groups have been matched for L1 and age of beginning English education, Table 2.1 shows that there is a significant difference between participant groups for factors including length of English exposure, amounts of peer and home language use, receptive vocabulary, relative L1-L2 proficiency, L1 fluency, and verbal memory. In addition, it should not be assumed that the present study collected background data for all relevant factors by which these groups may differ. Thus, the data were split by L2 group and models for predictors run separately, rather than relaxing the requirement of random group assignment and analyzing the data as a between-subjects study.

Models included the learner-external factors of length of English exposure, home language use, peer language use, and English hours per week. Learner-internal factors were nonverbal reasoning and verbal memory, and the language knowledge variables of English receptive vocabulary, self-rated relative language proficiency, and L1 proficiency. Age of beginning English education could also be included as a learner-internal factor. However,

this variable and the learner-external variable of length of English exposure were strongly correlated for the L2 groups ($r(52) = -.85$, $p < .001$) and thus, only the latter was included as it specified more variation between the participants.

For the immersion-L2s, the model showed that a larger receptive vocabulary (PPVT), better nonverbal reasoning (KBIT), and using English more frequently with peers (peer language use) predicted greater accuracy for both levels of difficulty. Likelihood ratio tests indicated that this was the optimal model and had a concordance index of .84. No interactions with construction difficulty were found. Fixed effects for the model had an R^2 of .02 and inclusion of the random effects for Subject and Item gave an R^2 of .39.

For the community-L2s, the model showed that being relatively more proficient in English predicted greater accuracy for both levels of difficulty, but an interaction with stimulus difficulty showed that more English at home only predicted higher accuracy on Easy stimulus items. The optimal model was selected using likelihood ratio tests, and included a random slope for Item and relative language proficiency, and an uncorrelated random slope for Item and home language. The concordance index for the optimal community-L2 model was .86. Results for these models are shown in Table 2.5. The fixed effects for the model had an R^2 of .15, with an R^2 of .40 for addition of the random effects for Subject and Item. It should be noted, however, that when stimulus difficulty is removed from the model the R^2 is reduced to .04.

Table 2.5 Model Results: Predictors of Accuracy for L2 Groups

	estimate	std. err	*z*-value	*p*-value
Immersion-L2 Accuracy				
(Intercept)	0.181	0.182	0.990	.322
Receptive vocabulary	0.192	0.087	2.191	.028
Nonverbal reasoning	0.158	0.080	1.968	.049
Peer language use	0.186	0.085	2.177	.03
	Concordance Index			.84
Community-L2 Accuracy				
(Intercept)	0.933	0.167	5.604	<.001
Relative Language Proficiency	0.369	0.124	2.969	.003
Home Language Use	0.207	0.123	1.679	.093
Stimulus Condition 1 (Easy)	0.680	0.154	4.417	<.001
Home Lang. Use: Stimulus Condition 1	0.313	0.102	3.077	.002
	Concordance Index			.86

Note. Models reporting factors found to be significant predictors of accuracy for the immersion-L2 and community-L2 groups. All numeric predictor variables were centred and standardized as *z*-scores. Stimulus condition is sum coded. The estimate column gives standardized regression estimates.

2.4. Discussion

The present study examined the morphosyntactic ability of three groups of adolescent English speakers measured by accuracy on a judgment task. As part of the first research question, the study compared three participant groups: 1) L2 learners who had acquired English in immersion classrooms (immersion-L2), 2) L2 learners who had acquired English in an English majority community (community-L2), and 3) L1-English speakers (L1 controls). These group comparisons also included analysis of different levels of construction difficulty finding that the immersion-L2 group participants had significantly lower task accuracy than the L1 control participants for stimuli of both difficulty levels, but

were only significantly below the community-L2 group for the Easy stimuli. The control group participants were significantly more accurate than the L2 groups for both stimuli types.

The second research question examined the impact of individual difference variables on the L2 groups, and how these potentially differ in importance between contexts and interact with different levels of construction difficulty. For the immersion-L2 participants, findings for this research question were that a larger receptive vocabulary, more L2 use with friends, and higher nonverbal reasoning were all significant predictors of accurate responses. For the community-L2 participants, a self-rating of greater English proficiency relative to Mandarin and more English use at home were both significant predictors of accurate responses. However, the latter predictor was only significant for responses to Easy stimuli.

2.4.1. Group Differences in Morphosyntactic Ability

The prediction that the immersion-L2 participants would have lower task accuracy was upheld for both levels of stimuli difficulty for the L1 control to immersion-L2 comparison. However, for the comparison between the community-L2 and immersion-L2 groups, findings were only consistent with this prediction for the Easy stimuli. For the Difficult stimuli, no difference was found between the two L2 groups. While the community-L2 group was different from both other groups for Easy stimuli, the difference in the estimated regression coefficient was greatest between the community-L2s and immersion-L2 participants, indicating that the community-L2 group was more similar to the L1 control group (also see Figure 2.1). Further, comparison of the *d*-prime scores in Table 2.2 shows

that many community-L2 participants performed within the one *SD* range for the L1 control participants. This indicates that while there was a significant difference between the groups for Easy stimuli, there were individuals in the community-L2 group that performed on par with their monolingual peers. This is similar to the findings of Jia & Fuse (2007), Paradis et al. (2016), and Soto-Corominas et al. (2020), where some - but not all - of the community-L2 learners in the respective studies were less accurate than expectations based on L1 criteria.

As noted, researchers often observe that the L2 morphosyntactic ability of immersion learners is conspicuously non-target (Harley et al., 1990; Kowal & Swain, 1997). However, this is widely considered to have a greater impact on L2 production than L2 receptive ability (Lyster, 2007; Swain & Lapkin, 1982; Tedick & Wesely, 2015). The results of this task show that there can be differences in receptive morphosyntactic ability, as well. The non-target L2 morphosyntactic ability of immersion learners has been suggested to result from greater reliance on semantic processing rather than deeper syntactic processing (Swain, 2005). Similar proposals have been made to explain the L2 acquisition of adult-L2 learners (*i.e.*, those who begin the L2 at or after adolescence) with the implication that child-L2 learners should process the L2 in a similar manner to L1 learners (*e.g.*, Clahsen & Felser, 2006). However, the findings of the present study echo those of other immersion studies that demonstrate L2 learning children are not always unlike adult L2 learners who may develop strong L2 communicative proficiency while some L2 morphosyntactic abilities remain divergent from those of L1 speakers.

Despite the overall low accuracy of the immersion learners, both L2 groups had difficulty determining grammatical from ungrammatical stimuli in the Difficult condition.

The lack of significant difference between the two L2 groups suggests that not all morphosyntactic differences are attributable to the immersion context. English grammatical morphemes were chosen for this study because they should be highly frequent in the L2 input, yet they can also be difficult for speakers of isolating L1s. Nevertheless, all participant groups including the L1 controls appeared to have more difficulty with the nonadjacent agreement stimuli. Number-marked verbs are highly frequent in spoken English even despite its impoverished agreement system (Bock & Middleton, 2011). However, this does not mean that participants frequently encounter constructions with nonadjacent agreement. English users should be considered more likely to hear sentences like 'The **computers** *were* fixed by the technician' as opposed to 'The **problems** with the computer *were* fixed by the technician'. If lower accuracy for the difficult stimuli is related to frequency, then this would be expected to impact all participant groups. However, the immersion learners' low accuracy with Easy stimuli is unlikely to result from lack of exposure as eight years of communicative L2 exposure should provide numerous instances of English articles, plural-marking, and past tense verbs. Overall, the comparison of the immersion-L2 group to the L1 control and community-L2 groups supports prior conclusions that immersion learners find L2 morphosyntax challenging.

2.4.2. Individual Predictors of Accuracy across L2 Contexts

The purpose of the second research question was to understand how learner-external and learner-internal factors influenced the L2 groups' accuracy on the judgment task. As expected, learner factors had an impact for both L2 groups, but these were different for

each group, and to some extent, for construction difficulty. Nevertheless, R^2 values for both models were low (.02 and .15 for the fixed effects of the immersion-L2 and community-L2 models, respectively. Note that the R^2 value for the community-L2 model is higher almost entirely due to the inclusion of stimulus difficulty), indicating that these factors did not explain a large portion of the variance between participants. However, this should not be unexpected if L2 morphosyntactic development has begun to slow or plateau. The factors examined in the present study were included because they have been found to predict L2 ability for community-L2 learners with less L2 exposure than the participants of the present study. However, factors predictive of individual differences in morphosyntactic development at earlier stages of L2 exposure could lose predictive power if development has slowed down.

Of the variance that the models did explain, comparison of the estimated model coefficients for the immersion-L2 learners shows that a larger receptive vocabulary was the strongest predictor of accuracy. Previous research on immersion learners has shown that listening comprehension is predicted by both an immersion learners' L1 and L2 receptive vocabulary in the first year of grade-four-start immersion (Vandergrift & Baker, 2015). Receptive vocabulary was measured by age-standardized PPVT score which is not a measure of absolute vocabulary, but vocabulary knowledge relative to same-aged L1 speakers. For an L1 learner, a large receptive vocabulary typically indexes that the child has either richer language exposure (Hart & Risley, 1995), greater cognitive ability to learn vocabulary (Gathercole, 2006), or a combination of both. For immersion learners, if language exposure is mainly limited to classrooms (Swain & Johnson, 1997), then, variation

in PPVT score is more likely to reflect an underlying ability to learn vocabulary (*i.e.*, a cognitive ability) rather than differences in L2 input.

The learner-internal factor of nonverbal reasoning also significantly predicted accuracy for the immersion-L2 learners. This variable has been associated with first grade immersion learners' general L2 proficiency (Genesee & Hamayan, 1980). For community-L2 learners in their first few years of L2 exposure, it is associated with inflectional morphology ability (Paradis, 2011; Paradis et al., 2020), and use of complex syntax in production (Paradis et al., 2017). In the present study, however, nonverbal reasoning was not a significant predictor for the community-L2 group. As noted, if differences in L2 exposure between the two contexts result in the immersion context being less supportive of morphosyntactic development, it may be that, like the contribution of language analytic ability in instructed-L2 contexts (Erlam, 2005; Li et al., 2019), immersion learners with better nonverbal reasoning ability are better able to compensate.

For the immersion-L2 participants, the learner-external factor of peer language use was the second strongest predictor after receptive vocabulary. This result parallels other immersion findings for general L2 proficiency, where greater L2 use outside the classroom has a positive impact for learners in their first years of immersion (Genesee, 1978; Hamayan et al., 1977). Based on prior community-L2 research with younger learners, L2 use with peers should also be expected to make an impact, as it contributes to a richer L2 environment, which in turn, impacts morphosyntactic acquisition (Jia & Fuse, 2007; Paradis et al., 2017, 2020; Rojas et al., 2016). However, no impact was found here for the community-L2 group. The two L2 groups had a statistically significant difference in amounts of peer L2 use, with almost all community-L2 participants reporting they always

use English with peers, whereas there was a greater range of L1-L2 use for the immersion-L2 participants. Thus, the lack of impact for this factor may result from the community-L2s all being at ceiling for this variable. Despite the lack of impact, amount of peer language use appears to be a characteristic difference between the learner contexts at the macro level.

For the community-L2 participants, a self-rating of being relatively more proficient with English than Mandarin was the strongest predictor of task accuracy for both the Easy and Difficult stimulus items. Greater relative proficiency is often interpreted as being dominant in a particular language, and dominance can be predictive of morphosyntactic accuracy for simultaneous bilingual children (Gathercole, Laporte, & Thomas, 2005; Paradis, 2010). Relative language proficiency may also simply reflect a participant's general sense that they have strong English skills. However, that their awareness of their own proficiency correlates with better grammatical morphology ability is noteworthy as grammatical morphology can be an area of difficulty for community-L2 English learners whose L1 is an isolating language (Blom et al., 2012; Jia & Fuse, 2007; McDonald, 2000; Paradis et al., 2016). Results of the present study indicate that these L2 learners understand, whether consciously or not, that grammatical morphology is a key component of English knowledge.

In contrast to the community-L2 participants, few of the immersion-L2 participants reported being more proficient in English, and of these, only three reported feeling they were much stronger in English (*i.e.*, rated themselves a four or above on a five-point scale). In addition, relative proficiency can depend on the social domain in which the language is used (Grosjean, 2008). Thus, immersion-L2 participant responses may have only reflected their relative proficiency in a school setting (which was also the immediate location in

which they were tested), and thus, it may be advantageous to ask about relative language proficiency for a range of social and educational settings in future research with immersion populations.

Greater use of L2 at home also had a positive impact on the community-L2 group's accuracy for Easy stimulus items, while there was no impact for the immersion-L2 group. Like peer language use, there was also a significant difference between the groups for L2 use at home, with very few immersion-L2s reporting any English use at home while the community-L2s had a greater range.

Length of exposure was not a significant predictor for either group in the models run for the second research question. L2 proficiency of early learners in immersion contexts is argued to plateau in many linguistic domains before learners reach the ends of their programs (Fortune & Tedick, 2015; Hart et al., 1991). Given that a similar situation may exist for community-L2 learners (Jia & Fuse, 2007; Paradis et al., 2016), and that detectable differences between L1 speakers and early L2 speakers can last into adulthood (Flege et al., 1999; McDonald, 2000), it is possible that no large shifts in morphosyntactic ability would occur for either L2 group. However, longitudinal research would provide a clearer timeline of L2 development in either context.

2.5. Conclusions

The present study shows that the morphosyntactic ability of adolescent L2 learners with more than eight years of L2 exposure can differ based on the context in which they learn their L2. Nevertheless, a difference in context is not the only factor that impacts as both community-L2 learners and immersion-L2 learners showed differences from L1 controls

and were similar to each other for constructions that were relatively more difficult. Learner-external and learner-internal factors were found to predict ability for the immersion-L2 group, but only learner-external factors predicted ability for the community-L2 group. However, these learner factors explained very little variance in ability within each group. This finding combined with the lack of impact of length of L2 exposure suggests that L2 development has slowed for the L2 learners of both of these high proficiency L2 groups.

3 Comprehension of English plural-singular marking by Mandarin-L1, early L2-immersion learners

3.1. Introduction

Immersion classrooms have been demonstrated to be a particularly effective way to establish second language (L2) communicative ability (Genesee, 1987, 2006; Lyster & Genesee, 2019). Unlike a traditional L2 classroom, where the new language is the main topic of instruction, in immersion, the L2 is used for presentation of classroom content such as mathematics and science, and for interaction in general. Despite the effectiveness of immersion classrooms for the development of functional proficiency in the L2, children who participate in these programs do not uniformly attain L2 competence identical to monolingual speakers of the target language (Lyster & Genesee, 2019).

Specifically for Canadian French immersion programs, it is often noted that immersion learners typically attain native-like or near-native receptive L2 skills, but at the same time, their L2 production skills differ from the sociolinguistic and grammatical norms of first language (L1) French speakers (Cummins, 1998; Kowal & Swain, 1997; Lyster, 2007; Lyster & Genesee, 2019; Swain & Lapkin, 1982). For L2 production, differences have been found for morphosyntactic features such as, but not limited to, learners' use of verb forms (Harley, 1993; Harley & Swain, 1978), and grammatical gender with nouns (Warden,

1997). Studies of immersion programs involving other languages have come to parallel conclusions about morphosyntactic production, with differences between immersion learners and L1-speakers being documented in Cherokee (Peter, Sly, & Hirata-Edds, 2011), Irish (Ó Duibhir, 2011), and Spanish (Montrul & Potowski, 2007) programs. In contrast to language production, immersion learners consistently demonstrate strong language comprehension ability on holistic tasks of receptive language (Genesee, 1987; Knell, Siegel, & Lin, 2007; Lindholm-Leary, 2011; Swain & Lapkin, 1982; Turnbull, Lapkin, & Hart, 2001). However, such assessments of comprehension do not provide evidence about how these learners integrate specific morphosyntactic information, or whether they interpret this information differently from L1-speaking peers.

To address the gap between investigations of morphosyntactic production versus comprehension skills in immersion-L2 learners, the present study examines the comprehension abilities of middle-school-aged L2 learners who have been learning English as an L2 in a foreign language immersion classroom for more than 8 years on average. The study examined the online (*i.e.*, moment-to-moment) comprehension of English inflectional plural-singular marking, a feature for which the participants' L1 (Mandarin) and L2 display key differences (see Li & Thompson, 1989), and for which speakers of Chinese languages have displayed differences in both early L2 acquisition (Jia, 2003) and late L2 acquisition (Jiang, 2007). The experiment combined a visual-world eye-tracking task that measures online comprehension and a picture decision task that measures final interpretations. In so doing, this study offers a fine-grained look into the under-studied area of immersion learners' comprehension of L2 morphosyntax.

3.1.1. The Immersion-L2 Context

The key objectives of immersion education are for learners to successfully learn required curriculum through their L2 and, in so doing, obtain much greater proficiency in their L2 than peers who receive traditional L2 instruction (Genesee, 2006; Lyster & Genesee, 2019). When it comes to L2 grammar, immersion education was initially based on the premise that for early start programs (*i.e.*, at or before first grade) language learning could occur naturalistically in much the same way that L1 acquisition occurs (Genesee, 1984), thus obviating the need for explicit language instruction, such as teaching of grammatical rules (Lyster & Tedick, 2014). Currently, however, the limitations of immersion environments are better understood, in that they do not perfectly mimic community exposure for things like the frequency of morphosyntactic constructions (Lyster, 2007; Peter, Hirata-Edds, & Montgomery-Anderson, 2008), or range of sociolinguistic variation (Harley, Allen, Cummins, & Swain, 1990; Mougeon, Nadasdi, & Rehner, 2010; Swain & Lapkin, 1990). Importantly, these variations in language exposure are inherently part of the immersion learning context and are believed to underlie differences in learners' L2 attainment. Factors such as frequency of language use, and how and with whom a language is used impact what individuals know of their languages (Grosjean, 2008), and these factors will be largely governed by learning context. Therefore, contextual factors need to be considered in explaining potential differences in morphosyntactic ability.

At a basic level, the immersion-L2 context can be situated somewhere between exposure to an L2 within a community where the L2 is the majority language (henceforth, the community-L2 context) or within a traditional L2 classroom where the L2 is the subject of instruction (henceforth, the instructed-L2 context). The differences between the

community-L2 and instructed-L2 contexts entail immense differences in the amount and quality of L2 input learners receive (Muñoz, 2014). To date, studies of early learners' acquisition of L2 morphosyntax have largely been done in community-L2 contexts (*e.g.*, Paradis, Tulpar, & Arppe, 2016; Chondrogianni & Marinis, 2012; Jia & Fuse, 2007). Like these community-L2 learners, immersion learners typically begin young, and have many hours a week of communicative L2 exposure from native speakers (or highly proficient speakers) of the L2. However, like instructed-L2 learners, the exposure that immersion learners receive is largely limited to the classroom in communities where the L2 is a foreign (or minority) language, making them similar to the participants of studies of instructed-L2 classrooms (*e.g.*, Sun, Steinkrauss, Tendeiro, & de Bot, 2016; Butler & Le, 2018; Muñoz, Cadierno, & Casas, 2018). Though immersion contexts are not likely to reach the same quantity of L2 exposure as community-L2 contexts, learners in these programs can amass around 6000 hours of L2 exposure by their middle-school years (Turnbull, Lapkin, Hart, & Swain, 1998), which far exceeds that of instructed-L2 contexts for which L2 exposure is limited to only a few hours a week and where total classroom exposure may only be around 800 hours by the end of compulsory education though this will vary widely by program (Muñoz, 2008). Though children may begin learning an L2 at similar ages, the L2 exposure that they receive will differ depending on whether they are instructed, immersion, or community-L2 learners. Because immersion is a distinct L2 learning context, L2 outcomes for immersion learners should be expected to differ from those of both instructed-L2 and community-L2 learners.

3.1.2. Acquisition of Morphosyntax in Language Immersion

When examining the morphosyntactic accuracy of immersion learners, many of the factors that potentially impact late learners (*i.e.*, L2 learners who begin at or after puberty), such as cross-linguistic influence, and lack of sensitivity to features with low saliency, have also been proposed to impact learners in immersion classrooms (Harley, 1993; Lyster, 2007). Within Canadian French immersion programs, learners tend to show non-standard use of the verbs ÊTRE and AVOIR (both highly frequent), as well as, second-person pronouns, *passé composé*, and verbs with derivational morphemes (Harley, 1993; Harley & Swain, 1978). Similar to adult L2 learners, grammatical gender can be non-target-like even up to the final years of high school for learners who began immersion in first grade, despite the feature being one French-speaking monolinguals master early (Warden, 1997).

These differences extend beyond French immersion. In Spanish immersion in the United States, majority-L1 learners' usage of past-tense forms (Tedick & Young, 2016), and gender (Montrul & Potowski, 2007) may have inaccuracies into fifth grade. For contexts where the language of immersion is a heritage language, there is also evidence of variable grammatical accuracy. For Irish immersion, by the fifth or sixth grade, production of genitive case can still be variable (Ní Dhiorbháin & Ó Duibhir, 2017; Ó Duibhir, Ní Dhiorbháin, & Cosgrove, 2016), and usage of the Irish copula and substantive verb, as well as syntax with verbal nouns more closely reflect correct English usage rather than Irish (Ó Duibhir, 2011). Similarly, knowledge of the complex verb system of Cherokee appears to be difficult for children in early immersion (Peter et al., 2008), but benefits from instruction that focuses on these morphological forms (Peter et al., 2011). Though little of this research specifically addresses ultimate attainment, it has been proposed that many of these non-

target features can be persistent and therefore require pedagogical intervention in order for learners to continue to progress (Harley, 1992; Kowal & Swain, 1997; Lyster, 2007).

In contrast to production, it is frequently noted that immersion learners are more native-like in their L2 comprehension (*e.g.*, Kowal & Swain, 1997; Cummins, 1998; Lyster, 2007; Swain & Lapkin, 1982; Tedick & Wesely, 2015). However, this conclusion is based on the absence of conspicuous differences in global comprehension rather than close examination of these learners' receptive abilities. Nevertheless, the claim that immersion learners typically have strong L2 comprehension is well supported as they reliably demonstrate L2 skills sufficient for completing appropriate grade-level content (Alanís, 2000; Cazabon, Nicoladis, & Lambert, 1998; Cheng, Li, Kirby, Qiang, & Wade-Woolley, 2010; Genesee, 1987; Harley, Hart, & Lapkin, 1986; Knell et al., 2007; Lindholm-Leary, 2011; Swain & Lapkin, 1982; Turnbull et al., 2001). Assessments that measure L2 proficiency by tracking immersion learners' holistic listening and reading comprehension (*e.g.*, Turnbull et al., 1998; Genesee, 1981; Harley & Hart, 1997; Siegel et al., 2010; Swain & Lapkin, 1982) are certain to have more validity in predicting academic success than a psycholinguistic experiment, yet such assessments offer little evidence about the precise nature of immersion learners' comprehension of particular linguistic features. As a result, how these populations interpret specific components of L2 morphosyntactic structures is a relatively unresearched topic. In contrast to holistic comprehension measures, production is often evaluated using targeted tasks such as elicitation, sentence repetition, or detailed error analysis from oral interview transcriptions (*e.g.*, Turnbull et al., 1998; Bild & Swain, 1989; Day & Shapson, 1987; Harley et al., 1990; Harley & Hart, 1997; Thomas, Williams, Jones,

Davies, & Binks, 2014). In sum, the comprehension and production abilities of immersion learners are not evaluated at comparable levels of scrutiny.

Though the trend is to use holistic assessment for receptive L2 skills, a notable exception to this is Lew-Williams (2017). In this study, eye-tracking was used to examine the receptive ability of English-L1 learners learning L2 Spanish in two-way immersion classrooms. These participants' performance was compared with that of their Spanish-L1 classmates. Prior eye-tracking research has shown that L1 speakers of languages with grammatical gender can use gender-marked determiners to identify noun referents more rapidly, but that late L2 learners typically do not use gender information this way (Hopp, 2013; Lew-Williams & Fernald, 2010), suggesting that the sometimes inconsistent usage of grammatical gender of even highly proficient L2 speakers is not only reflective of productive abilities, but comprehension as well (Grüter, Lew-Williams, & Fernald, 2012). In Lew-Williams (2017), 3 separate eye-tracking experiments measured participants ability to identify upcoming nouns using Spanish determiners with abstract grammatical gender, grammatical gender that coincided with the notional gender of the noun (*e.g.*, the Spanish translations of 'the man' or 'the woman'), or number information. The English-L1 participants in fourth or fifth grade, like Spanish-L1s peers, were able to use notional gender and number-marked determiners to identify referents. However, unlike their Spanish-L1 peers, they could not use grammatical gender this way, suggesting that the acquisition of these immersion learners was similar to late L2 learners. Lew-Williams concluded that the lack of semantic transparency of features like grammatical gender may be an obstacle to acquisition in L2 contexts that have more limited input than community-L2 contexts, but that it remained unclear whether the advantage for interpretation of

notional gender and number was semantic transparency or the result of transferred L1 processing skills.

3.1.3. Number Marking in English and Mandarin

Mandarin and English differ substantially in how and when number information on nouns is conveyed, and this has been proposed to impact Mandarin speakers' L2 acquisition of English (*e.g.*, Jiang, 2007; Luk & Shirai, 2009). While nouns can be quantified in both languages, a key difference between them is that number marking is a non-optional feature of English nouns (Corbett, 2000). Nouns in Mandarin have what can be referred to as *general number* (Corbett, 2000; Rullmann & You, 2006), meaning that a noun can be expressed without reference to number. As a result, bare Mandarin nouns, unmarked for number, are very common in typical usage (Cheng & Sybesma, 1999; Li & Thompson, 1989; Li, 1999). They are, therefore, ambiguous in this regard when used outside of a discourse context.

The attention that English plural marking has received in previous L2 literature for L1 speakers of Mandarin, (*e.g.,* Jia, 2003; Jiang, 2007; Luk & Shirai, 2009; Wen, Miyao, Takeda, Chu, & Schwartz, 2010; Xu Rattanasone, Davies, Schembri, Andronos, & Demuth, 2016) suggests the feature is a good candidate for one that is less likely to benefit from positive language transfer, and therefore needs to be learned anew through L2 exposure. This difference between Mandarin and English also applies to Cantonese (Cheng & Sybesma, 2005), Japanese (Nakanishi & Tomioka, 2004), and Korean (Kwon & Zribi-Hertz, 2004), and thus, relevant research involving these languages is also discussed. These

languages all contrast with English in that, for each, plural-singular marking is optional, depending on discourse context, while in English it is an obligatory grammatical feature.

3.1.4. Acquisition of English Plural-Singular Marking

In the L1 acquisition of English, plural '*-s*' is one of the earliest grammatical morphemes produced consistently by monolingual English-speaking children (Brown, 1973; de Villiers & de Villiers, 1973). Data from overregularization errors indicate that English-L1 children generally have productive knowledge of plural '*-s*' by the age of 4 (Matthews & Theakston, 2006), though these children may not master it to adult levels until the age of 7 or later (Berko, 1958; Graves & Koziol, 1971). Around 28 months, young English speakers are likely aware that '*cats*' deconstructs into '*cat*' and '*-s*' (Zapf & Smith, 2007). For semantic understanding, children around 24 months demonstrate knowledge of the plural-singular distinction (Barner, Thalwitz, Wood, Yang, & Carey, 2007; Li, Ogura, Barner, Yang, & Carey, 2009), though, at this early stage, they may not have successfully mapped it to every phonological marker (*i.e.,* to verb inflections, quantifiers, or all allomorphs of '*-s*'; Kouider, Halberda, Wood, & Carey, 2006; Davies, Xu Rattanasone, & Demuth, 2017; Wood, Kouider, & Carey, 2009).

For young L2-English learners in community-L2 contexts, Dulay and Burt (1974) initially proposed that the acquisition of English morphemes was not impacted by L1. However, their own data show a very large discrepancy between the Spanish-L1 and Chinese-L1 language groups for plural marking, with the Spanish L1s marking the feature approximately 90% of the time in elicited speech and the Chinese speakers doing so only about 50% of the time. More recent evidence from measurement of receptive language

ability indicates that in early preschool exposure to English, speakers of Chinese L1s may acquire English plural marking more slowly than monolinguals and L2 peers who speak other L1s. In a study of preschool-aged children, Xu Rattanasone et al. (2016) used a pointing task with touchscreen tablet computers to measure 3- and 4-year-old children's ability to correctly interpret singular or plural marked nouns. In the experiment, English-L1 children performed above chance for both plural and singular marking with nonce nouns. Speakers of non-Chinese L1s were only above chance for plural items, and those with a Chinese L1 were only above chance for singular items. When the singular and plural conditions were compared, the English L1s and non-Chinese L1s were both significantly better with plural than singular. However, there was no difference between the types for the Chinese L1s. Xu Rattanasone et al. (2016) concluded that the Chinese L1s showed greater difficulties acquiring inflectional morphology at that time point in development.

For young Korean- and Japanese-L1 children in community-L2 contexts, production of plural '*-s*' appears to come later in L2 English than in L1 English relative to other grammatical morphemes. In specific contrast to L1 acquisition, Hakuta (1976), and Shin and Milroy (1999) found that the phonologically identical possessive '*-s*' was acquired before plural marking for the children they studied. The authors of both studies ascribe this L1-L2 English disparity to the lack of an obligatory plural marking morpheme in the children's prior learned L1s, Japanese and Korean, respectively.

In a 5-year longitudinal study of Mandarin speakers, Jia (2003) studied participants' production of English inflectional plural marking in picture descriptions and spontaneous speech. Her participants were 10 children and youths whose English development in a community-L2 context began between 5 and 16 years of age. Despite the duration of

exposure, 3 of the participants fell short of the 80% criterion score for consistent plural production. Participants' most common error was to use an unmarked noun that should be marked plural with either regular or irregular inflection. Jia noted that each participant across all rounds of measurement tended to use the same nouns correctly marked at times and erroneously unmarked at others, even within the same testing session. Age of English onset (AoO) and language environment explained some but not all of the variation between learners.

With the exception of Xu Rattanasone et al. (2016), studies of early community-L2 acquisition relied on participant children's ability to produce plural marking. In contrast, L2 research with adult late-L2 participants has used self-paced reading experiments with response times measuring receptive sensitivity to errors with plural-singular marking. Jiang and colleagues (Jiang, 2007; Jiang, Novokshanova, Masuda, & Wang, 2011) used this paradigm to provide evidence that native-like acquisition of plural marking was extremely difficult or impossible for L2 learners without an equivalent morphological feature in their L1. In Jiang (2007) and Jiang et al. (2011), participants were presented with lists of stimuli that probed sensitivity to error stimuli for plural marking in sentences like **The teacher noticed that* ***three of his students*** *were always missing*, and for verb subcategorization, as in **Joe's father didn't* ***show him drive*** *the car that day*. In both studies, the optional-number-L1 participants (Chinese and Japanese respectively) were only sensitive to the verb subcategorization errors as measured by slower reading times for stimulus sentence regions subsequent to the error. In Jiang et al. (2011), a comparison group of Russian-L1, English-L2 speakers was also included. The prediction was that the Russian speakers would show sensitivity to both verb subcategorization errors *and* plural errors, because

Russian, like English, obligatorily marks plural. This prediction was borne out as the speakers of the L1 more like English demonstrated sensitivity to plural errors.

The two L1 participant groups in Jiang et al. (2011) were matched for performance on a cloze task used to gauge proficiency and were similar in other aspects of English experience such as AoO, and years of formal instruction, though the Japanese L1 group began earlier and had more years of formal instruction than the Russian L1s. Nevertheless, two subsequent studies found that Chinese and Japanese (Wen et al., 2010), and Korean speakers (Song, 2015) can show evidence of monolingual-like knowledge of English plural inflection. Both studies were largely replications of the Jiang studies, but included the two key modifications of separating the L2 learners into intermediate and advanced proficiency groups and reducing the syntactic complexity of experimental stimuli. Wen et al. (2010) pointed out the stimuli of the previous studies were grammatical and ungrammatical partitive constructions (*e.g., …three of his students…*), requiring tracking agreement across multiple hierarchical nodes. This syntactic complexity has been demonstrated to influence L2 sensitivity to grammatical gender agreement (Keating, 2009). Thus, it was unclear whether L2 participants' difficulties were with plural marking, or syntactic complexity. To address this, the stimuli used in Wen et al. consisted of phrases with an adjective separating the number-marked determiner and the noun (*e.g., these nice houses*), while Song (2015) included stimuli of both types. Both studies found evidence of native-like knowledge for advanced, but not intermediate learners.

In sum, the acquisition of plural-singular marking in L2 English has been examined for both Mandarin-L1 children residing in community-L2 contexts and for late-L2 Mandarin L1 adults who, where reported, began in instructed contexts. To date, no studies

have examined the comprehension of this morphosyntactic feature in early immersion learners of English.

3.1.5. The Present Study

The main goal of this study was to investigate comprehension of a single semantically transparent morphosyntactic feature, English inflectional plural-singular marking, by a group of L2-English immersion learners whose L1 is Mandarin. The study employs an experimental task and compares adolescent early L2-immersion learners in Northern Taiwan with age-matched Canadian monolinguals who speak English as their L1. The immersion learners in this study have had long-term and extensive exposure to their L2 from preschool. Though the appropriateness of monolingual-bilingual comparisons is often questioned, an empirical comparison is a better alternative than a priori assumptions about how adolescent language users interpret morphological features on experimental tasks. If specific learning contexts (whether L1 or L2) impact acquisition, then group differences should be expected. Nevertheless, comparisons can be informative without the expectation of identical performance.

Though the L2-immersion group has extensive exposure to English, quantifying this exposure in relation to the monolingual group is difficult. Given that greater language exposure is associated with L2 vocabulary size, receptive vocabulary has previously been used as an estimate reflecting general proficiency (Unsworth et al., 2014). In addition, larger vocabulary sizes have been shown to predict community-L2 learners' accuracy with inflectional morphemes (Paradis et al., 2016) and to decrease both L1 and L2 processing

times (Brysbaert, Lagrou, & Stevens, 2017). For these reasons, receptive vocabulary size was included in our analysis.

In order to compare comprehension by the two study groups, this study employed a visual-world eye-tracking task measuring online comprehension combined with a picture decision task reflecting accuracy of sentence interpretation. Data from these tasks were used to examine the following research questions:

1. Do adolescent L2-immersion learners differ from age-matched monolinguals in offline interpretation of English plural-singular marking?

   Hypothesis One: This research question is addressed by accuracy and response times on the picture decision task. Two predictions are formed based on previous research for Mandarin-L1 participants in both community-L2 and late instructed-L2 contexts, and for immersion contexts in general. First, the immersion participants of this study are predicted to be less accurate in picture decisions. Second, a larger receptive vocabulary size is expected to predict greater accuracy and faster button response times. How the immersion participants will behave in comparison to the monolinguals for response times in general is less clear and therefore, no specific prediction is made.

2. Do adolescent L2-immersion learners differ from age-matched monolinguals in online comprehension of English plural-singular marking?

   Hypothesis Two: This research question is addressed by the eye-tracking task, which tracks the proportions of looks to targets over time. It is expected that the immersion participants will be less sensitive to plural-singular marking as compared to the monolingual participants, and this will be demonstrated by delays in looks to the target images and lower overall proportions of looks to target images. Again, higher receptive vocabulary scores are expected to predict less delay in looking to targets and a greater proportion of looks to targets.

3.2. Method

3.2.1. Participants

This study reports data from 33 Taiwanese middle-school participants recruited from a single private high school in northern Taiwan, and studying in grade seven, eight, or nine (mean age = 13;8, *SD* = 9 months). Data from these participants (henceforth the L2 group) were compared to that from 21 monolingual Canadian middle-school participants (mean

age = 13;6, *SD* = 13 months). In order to provide an independent baseline of proficiency, receptive vocabulary was assessed using the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007), a test normed against monolinguals and which provides a score standardized for age. The standard score for the PPVT is 100, with a one *SD* range of 85-115. The L2 participants of this study fell below the one *SD* range with a mean score of 78.42 (*SD* = 15.65). This is not unexpected as similar disparities have been reported for French immersion learners (Hermanto, Moreno, & Bialystok, 2012). Data were collected from 37 L2 participants. However, the data of four participants were excluded from analysis. One excluded participant was an outlier for receptive vocabulary, scoring approximately three *SD*s below the mean for the immersion learners and one *SD* below the next lowest participant. In addition, two participants who had less than five years of immersion were also excluded. Finally, one participant reported having English exposure from birth from a native English-speaking father, and thus would be better classified as a heritage-language speaker. All monolingual participants were included. The remaining L2 participants had been learning in an immersion context taught by a foreign teacher from an English-speaking country (*e.g.,* the UK, US, Canada, etc.) typically since preschool (mean AoO = 4;7, *SD* = 17 months). After preschool, the learners attended a bilingual elementary school where half their content instruction was in Mandarin from a local teacher, and half in English. On average, they had attended an immersion program for eight years and five months (*SD* = 20 months). All study participants were in good academic standing (*i.e.*, they were passing their classes). Summary data are given for the L2 participants and the monolinguals in Table 3.1.

Table 3.1 Age, Vocabulary, and Exposure Data for Participant Groups

	L2s				Monolinguals			
	M	*SD*	min	max	*M*	*SD*	min	max
Receptive vocabulary	78.42	15.65	50	116	110.86	16	87	154
Age	13;8	0;9	12;8	15;5	13;6	1;1	11;10	15;5
Age of immersion	4;7	1;5	2;0	8;0	-	-	-	-
Years of immersion	8.42	1.66	5	13	-	-	-	-

Note. Statistical means (*M*) are given with *SD*s and both the minimum and maximum values for each variable. Receptive vocabulary is the standardized score on the PPVT. The mean standard score for the PPVT is 100 with the one *SD* range 15 points above and below. The L2 participants fall below the one *SD* range on this monolingual-normed test. Age is the participants' age at testing and age of immersion is the age beginning English in an immersion classroom. Both age variables are given in years and months (yy;mm).

3.2.2. Materials

Experimental stimuli for this study consisted of 16 sentence pairs. The head noun for each sentence appeared in either its singular or plural forms creating 32 experimental items, which were counterbalanced across two stimulus lists containing equal numbers of singular and plural stimuli. The stimuli are provided in Appendix C. The 16 subject nouns all appear in a particular syntactic context where the noun target (*e.g.,* either *pencil* or *pencils* in Example 3.1 or 3.2) is preceded by the definite article *the* and appears before a verb which is not inflected for number. No stimulus noun appeared immediately before a word that began with either /s/ or /ʃ/ so that the plural allomorph, if present, should have been distinct from adjacent phonetic segments. The stimulus nouns were chosen to be highly imageable, frequent, and to be roughly balanced for the three potential plural allomorphs /s/, /z/, and /əz/ when in their plural form. According to the Corpus of Contemporary American English (Davies, 2010), the median frequency for the nouns was

12,818 tokens per million words (*IQR* = 27,223 to 7,591). An example noun stimulus pair is given below.

Example Stimulus Pair

3.1. The **pencil** can be used for writing the test.
3.2. The **pencils** can be used for writing the test.

In order to prevent participants from immediately focusing on inflectionally marked number differences, stimuli also included 40 filler items that differed on a non-morphological semantic contrast (*e.g.*, whether a shoe could be considered a woman's or a man's, or whether a flag was a Canadian flag or not). Within the experiment, there were also 24 stimulus items that differed in verb number agreement morphology. However, these are not analyzed here. Stimulus sentences were spoken by a female speaker of North American English and recorded digitally in a soundproof booth using a Korg MR-2000S studio recorder with a 44.1 kHz sampling rate and through a Countryman E6 headset microphone. The speaker was instructed to speak at a natural pace for casual speech, and the resulting mean articulation rate for the stimuli was 4.6 (0.71) syllables per second, calculated using Praat software (6.0.36; Boersma & Weenink, 2017) running a script created by de Jong and Wempe (2009).

For each trial, participants saw four pictures on a computer screen. The four pictures were juxtaposed so that within each trial there were two potential target-competitor sets. For example, the four-picture stimulus display that accompanied either audio stimulus Example 3.1 or 3.2 presented pictures of a single pencil, multiple pencils, and two pictures of single light bulbs that had a salient difference in type. This stimulus display is shown in Figure 3.1. For stimulus Example 3.1, the picture of the single pencil was the target, the multiple pencils the competitor, and the pictures of the light bulbs were distractors. Target

images could appear in any of the four image locations, but for each counter-balanced stimulus pair the relative target and competitor locations were always the same (*i.e.*, when the single-pencil image was the target, it appeared in the upper right corner of the display, and when the multiple-pencil image was the target the image also appeared in the upper right corner). In each trial, it was possible for picture pairs to both differ in number of items, a non-morphological semantic contrast (for fillers), or to have one of each distinction, as in Figure 3.1. This was done in order to prevent participants from predicting the type of distinction in the audio stimulus.

Figure 3.1. The picture set for this trial includes two noun pairs: two pencil images, and two light bulb images. The pencil images differ in number, while the light bulbs have a salient difference in type. Participants were instructed to keep their index fingers on left and right gamepad buttons and that target images would always differ by left or right side and never by top or bottom. This

trial corresponds to the example audio stimulus given in Example 3.1. where the correct button would be the right button for the image of a single pencil.

3.3. Procedure

For each of the trials, participants listened to a stimulus sentence while viewing each four-picture set on a computer monitor. The audio stimulus played 1000 milliseconds (msec) after the pictures appeared. Picture sets differed for each trial. While listening to the audio stimulus through a pair of headphones, participants were tasked with deciding which picture was being referred to in the audio stimulus. Participants entered their response on a gamepad controller and were required to decide whether the picture was on the left or right side of the monitor, and to respond as quickly as possible. Participants had to respond within 3500 msec of the audio offset, after which the images on the display would disappear and no response would be recorded. A target image could appear in any of the four grid locations, with competitor images potentially appearing in either the top or bottom position on the alternate side of the screen. All potential combinations were equally probable. Participants held the gamepad in both hands and responded with their index fingers which were held on either the left or right button. Stimulus trials were randomized for each participant with the condition that a single trial type (either plural-singular, verb, or filler) could not appear in more than three consecutive trials. For each trial, eye movements over the visual array were recorded using an SR-Research Eyelink Portable Duo set to a 500 Hz sampling rate and recording the right eye. Participants were positioned 50 cm from a 17-inch computer display with a resolution of 1920 by 1080 and a 60 Hz refresh rate. A head mount was not used, and participants could move their head freely. The Eyelink Portable Duo allows the distance between the participant and display to be

continuously monitored throughout the experiment. If the participant shifted outside of the optimal 42 to 62 cm range they were asked to reposition themselves. The eye-tracker was calibrated for each participant, and between each trial a fixation point appeared at the centre of the screen. When the participant's gaze was on the fixation point a drift correction was performed, and the experimenter advanced the experiment to the next trial. Before beginning the experiment, participants completed three practice trials.

3.4. Results

3.4.1. Picture Decision Task: Response Accuracy and Response Time

Data from both tasks were analyzed using *R* Statistical Software (version 3.6.0; R Core Team, 2018). The first hypothesis tested potential group differences in final interpretation reflected in picture decision task performance. Mixed effects models were run on the accuracy and response time of button responses. These models used functions from the *lme4* package (version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015). Trials where no button response was recorded (2.4% of trials; 2.8% for L2s and 1.8% for monolinguals) were counted as incorrect. Figure 3.2 presents a plot for response accuracy with responses divided by group and whether the correct response was the singular or plural image. When averaged across individual participants, the L2 group's responses were 69% accurate with an *SD* of 14%, while the monolinguals were 92% accurate, with an *SD* of 12%.

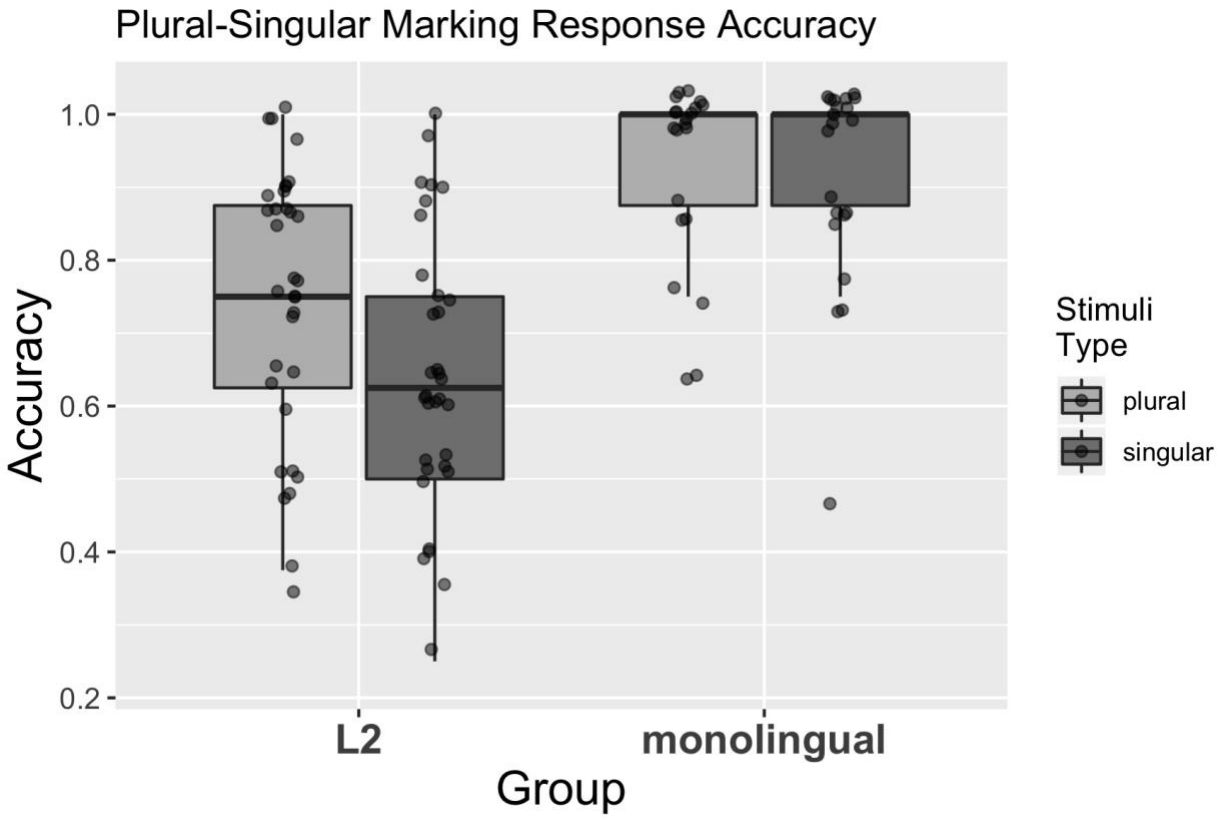

Figure 3.2. Participant accuracy by participant group for both plural and singular stimulus types. Area within the boxes shows scores within the 2nd and 3rd quartile, with horizontal bars showing median accuracy. Whiskers cover data within 1.5 of the upper and lower quartile ranges. Points show individual participant accuracy for each stimulus type.

To compare accuracy of the 2 study groups, a binomial generalized linear mixed effects model was run with the button response for each experimental trial as the dependent variable. In addition to group, the model was run with fixed effects for stimulus number (either plural or singular), and each participant's PPVT score. Inclusion of these variables was motivated by previous research showing learners can differ in responses to either plural or singular (Arias-Trejo, Cantrell, Smith, & Canto, Elda, 2014; Xu Rattanasone et al., 2016), and that language proficiency may also impact whether participants respond to plural-singular marking (Song, 2015; Wen et al., 2010). Group and stimulus number were categorical variables coded using effects coding rather than *R*'s default treatment

coding, while PPVT score was a centred continuous variable. For Group, monolinguals and L2s were coded as -0.5 and 0.5, respectively, and for stimulus number, singular and plural items were coded as -0.5 and 0.5, respectively. For random effects, the model included by-participant (Subject) and by-stimulus item (Item) random intercepts, as well as an uncorrelated random slope for PPVT score for Item. Following Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017), an attempt was made to add an uncorrelated random slope for stimulus number to Subject, and one for Group to Item, but the `lme4` package indicated that the addition of these slopes overfit the data. Including the correlation between PPVT and Item caused the model to fail to converge. A likelihood ratio test indicated that inclusion of the random slope for PPVT was warranted as it lowered AIC ($\chi^2(1) = 14.561$, $p < .001$). In the model with the three fixed effects, only Group and PPVT score made a significant contribution to accuracy. The model was also run with interactions between Group and either of the variables for PPVT or stimulus number, but neither interaction was significant. Full results for this model are given in Table 3.2. Because Group was also highly predictive of PPVT score, a point-biserial correlation for comparing dichotomous Group with continuous PPVT score was run finding a moderately strong significant correlation ($r_{pb}(52) = .71$, $p < .001$). Therefore, variance inflation factors (VIF) were calculated for the model using the `vif` function from the `car` package (version 3.0.3; Fox & Weisberg, 2019). The results did not indicate that collinearity was an issue as the VIF scores were 1.36 and 1.36 for Group and PPVT, respectively. These VIF scores are well below the 5 to 10 range where collinearity may be considered an issue (Levshina, 2015).

Table 3.2 Model Results: Picture Decision Accuracy

	Estimate	Std. Error	*z*-value	*p*-value	
Intercept	1.85	0.189	9.79	<.001	***
Group	1.05	0.372	2.82	.005	**
Receptive Vocabulary	0.577	0.215	2.69	.007	**
Stimulus Number	-0.471	0.285	-1.65	.099	

Note. Receptive vocabulary is the centred standardized score for the PPVT (Dunn & Dunn, 2007). *p<0.05; **p<0.01; ***p<0.001

Because the participants were under time pressure to respond, response times for the picture decision task were analyzed to see if this may have impacted accuracy. To compare response times, a linear mixed model was run with response time for every individual trial answered correctly as the dependent variable. Response time was calculated as the time duration in msec between the plural-singular marking cue and button press. The plural-singular cue was the end of the noun lemma for both plural and singular. For plural, this would be between the noun lemma and the beginning of the '*-s*' (… *pencil*[**cue**]*s can* …), and for singular, between the lemma and the onset of the next word in the stimulus (… *pencil*[**cue**] *can*…). Mean response times (in msec) for correctly answered trials was 3285 (*SD* = 678) for the L2 participants, and 2996 (*SD* = 619) for the monolinguals. A Shapiro-Wilk test indicated that response times were not normally distributed (W = 0.99, p < .001). Inspection with quantile plots and the *boxcox* function of the `MASS` package (version 7.3.51.4; Venables & Ripley, 2002) did not indicate analysis would benefit from transformation of the data. Participant group (either L2 or monolingual), stimulus number (plural or singular), and PPVT score were fixed effects. The same approach to random effects was taken as that for the accuracy model. Random slopes for stimulus number on Subject, and Group on Item overfit the data. Further, a likelihood

ratio test indicated that the random slope for PPVT on Item did not lower AIC, and was subsequently removed leaving just random intercepts for Subject and Item. From this model, the dataset was trimmed to just responses that fell within 2.5 *SD*s of the mean response time (Baayen, 2008). This trimming resulted in eliminating 1.5% of the total observations. The `lmerTest` package (version 3.1.0; Kuznetsova, Brockhoff, & Christiansen, 2017) provided *p*-values for models as `lme4` does not provide these. This model found that only receptive vocabulary and neither participant group nor stimulus number made a significant contribution to the model either on their own or in interaction with other variables. Likelihood ratio tests comparing the full model to one with each fixed effect alternately removed also indicated that only receptive vocabulary contributed to a significantly better model ($\chi^2(1) = 14.724$, $p < .001$). Results for the full model are given in Table 3.3.

Table 3.3 Model Results: Picture Decision Response Times

	Estimate	Std. Error	df	*t*-value	*p*-value	
Intercept	3120.0	58.5	65.3	53.4	<.001	***
Group	112.0	134.0	49.8	0.836	.407	
Stimulus Number	32.1	76.8	28.9	0.418	.679	
Receptive Vocabulary	-264.0	66.1	52.0	-4.00	<.001	***

Note. Receptive vocabulary is the centred standardized score for the PPVT (Dunn & Dunn, 2007). *p<0.05; **p<0.01; ***p<0.001

3.4.2. Eye-Tracking: Proportions of Looks over Time

For the second hypothesis, eye-tracking data were analyzed for differences between the participant groups in the proportions of looks after the plural-singular cue. However, because proportions are inherently bounded, proportions were transformed to empirical logits. This transformation requires aggregating over multiple observations (Barr, 2008).

First, the proportion of looks a participant made to each of the four stimulus images across trials was calculated for every 20 msec time bin. At the sampling rate of 500 Hz, this resulted in 10 samples per bin. These steps were performed using the R package `VWPre` (version 1.1.0; Porretta, Kyröläinen, van Rij, & Järvikivi, 2018). From these transformations, a *target advantage score* was calculated for the difference in looks to the target versus the competitor image for each 20 msec time bin, where a positive score indicated more looks to the target, zero indicated equal looks to the target and competitor, and a negative score indicated more looks to the competitor. Unlike for the accuracy and response time models where each trial represented one data point, each trial in the eye-tracking data represents a time series with one observation for target advantage score for every 20 msec time bin. Target advantage score served as the dependent variable for modelling.

In order to model the looks for the two participant groups separately for plural and singular stimuli, an additional variable, *Condition*, was created, which combined the two variables. The four levels of Condition are shown in Figure 3.3 where the four figure panels show performance of the two participant groups (left and right) for the two stimulus types of plural (top) and singular (bottom). Figure 3.3 uses locally weighted smoothing (LOESS) lines to show the proportion of looks to the four interest area images averaged for each trial. The creation of Condition allowed the model to analyze the dependent variable, target advantage score, independently over time for the four different conditions of L2 responses to plural stimuli; L2 responses to singular stimuli; monolingual responses to plural stimuli; and monolingual responses to singular stimuli. These four independent time courses could then be compared for difference. This approach to the creation of Condition as a predictor

variable capturing interactions between categorical variables follows van Rij, Hollebrandse, and Hendriks (2016).

To analyze differences in target advantage score for the time series in each trial, the eye-tracking data were analyzed with generalized additive mixed models (GAMMs; Baayen, Vasishth, Kliegl, & Bates, 2017; Wood, 2006). Because GAMMs allow the comparison of non-linear effects over time, they are well suited for use with visual-world data. GAMMs were run using the `mgcv` package (version 1.8.28; Wood, 2006).

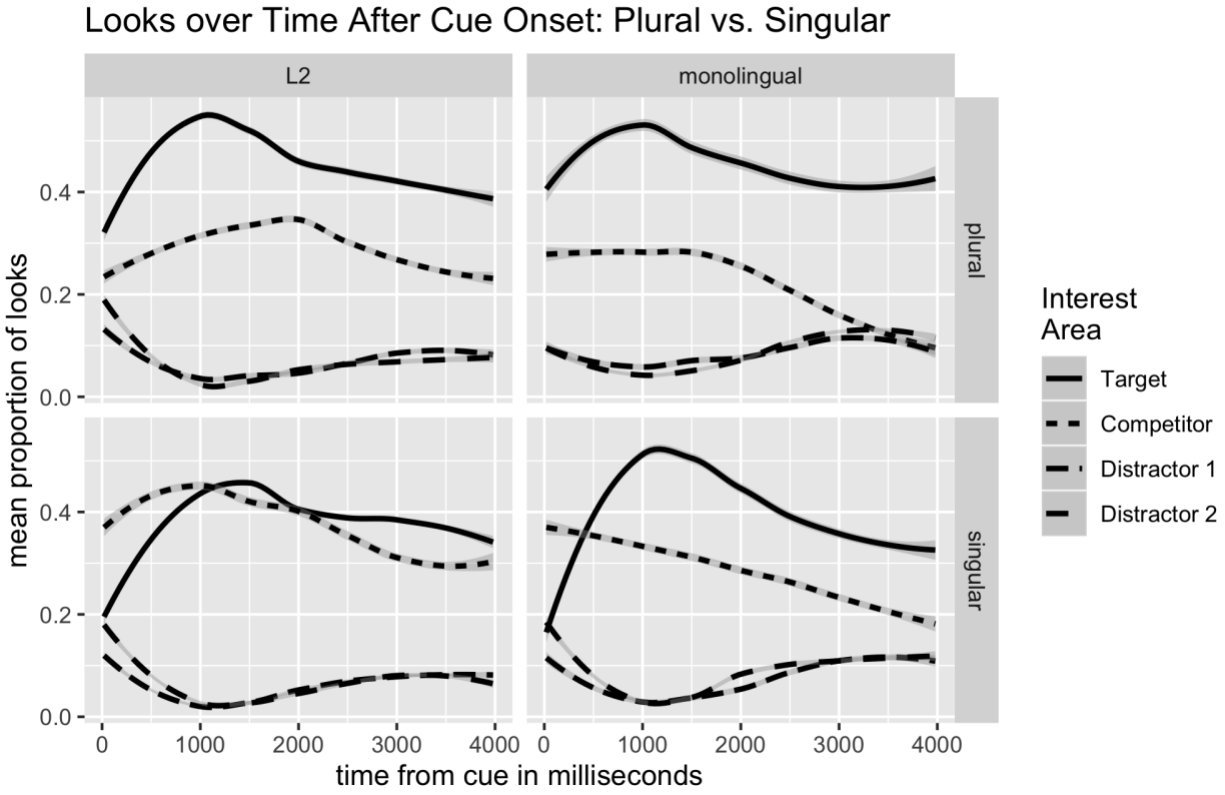

Figure 3.3. The figure shows LOESS smooth lines for mean participant looks by participant group for plural and singular stimulus types over a four-second time interval beginning at the offset of the target noun lemma, i.e., at either the onset of the plural '*-s*' or its absence. At the 0 point, looks to the target and competitor have already begun to diverge from the distractor images as looks to the potentially appropriate images begin shortly after the noun onset. This is

expected and shows that both groups rapidly comprehend the target noun but have a bias toward looking at plural images.

An initial model was run with the four levels of Condition and PPVT score as fixed effects. Condition was modeled as a smooth term to capture the non-linear performance reflected in Figure 3.3, and as a parametric predictor in order to account for potential intercept differences between levels of factors. PPVT was also modeled as a smooth term to capture potential non-linear influence. Rather than model the entire time course for each trial, a time window of 2000 msec beginning 200 msec post cue was selected. The 2200 msec end point was used as inspection of Figure 3.3 shows that looks to the target image generally peak around 1000 msec, then decrease until 2000 msec, and then plateau (with a slight continued decrease) around 2000 msec when looks to distractors are also beginning to increase. This indicates the impact of plural-singular marking on looks has largely faded beyond this approximate point meaning looks beyond this point are less likely to be reliably associated with the experimental task.

Like the generalized linear mixed effects models, the GAMM included random effects for Subject, and Item, as well as one for each experimental trial (Event). Each unique time-series for Subject and Item was given its own random smooth, while Event was given a random intercept. The functions `acf_resid` and `compareML` from the `itsadug` package (version 1.1.0; van Rij, Wieling, Baayen, & van Rijn, 2017) were used to calculate a *rho* value of .006 to correct for autocorrelation within the model, and to do model comparisons, respectively. Comparisons indicated that PPVT score did not improve the fit of the model and was therefore removed. Inclusion of Condition provided a better fit to the data than a

model with just the random effects (freml difference = 70.14, $p = < .001$). The parametric and smooth term coefficients for this model are given in Table 3.4.

Table 3.4 GAMM Output for the Eye-tracking Model

A. parametric coefficients	Estimate	Std. Error	*t*-value	*p-value*
Intercept	0.88	0.39	2.27	.023*
Condition: L2 Singular	-1.32	0.44	-3.01	.003**
Condition: Mono Singular	0.37	0.40	0.92	.36
Condition: Mono Plural	-0.24	0.62	-0.39	.70
B. smooth terms	edf	Ref.df	*F*-value	*p*-value
s(TimeIndex): Condition L2 Plural	4.58	5.10	1.42	.229
s(TimeIndex): Condition L2 Singular	1.00	1.00	0.81	.368
s(TimeIndex): Condition Mono Plural	5.06	5.65	1.76	.094
s(TimeIndex): Condition Mono Singular	6.84	7.56	4.31	<.001***
s(TimeIndex, Subject)	397.90	484.00	38.01	<.001***
s(TimeIndex, Item)	239.34	286.00	48.98	<.001***
s(Event)	727.45	860.00	9.95	<.001***

Note. The table presents the parametric and smooth coefficients for the optimal model with the difference in looks to the target minus looks to the competitor image for both stimulus types, and both participant groups. The results of this model are visualized in Figure 3.4. *p<0.05; **p<0.01; ***p<0.001

As indicated, Condition was an informative variable in explaining differences in target advantage score (*i.e.*, the proportion of looks to either the target or competitor image). Interpretation of GAMM results requires visual inspection of the model output which is given in Figure 3.4. The top row of Figure 3.4 shows comparisons of the differences in target advantage score between the two groups and for the two levels of stimulus number, singular (left) or plural (right). The bottom row shows comparisons for the two stimulus types within each group (L2s on the left and monolinguals on the right).

Points where the error bar shading does not cover the *x*-axis represent time windows with significant differences in proportions of looks. For the between group comparison for singular stimulus items (Figure 3.4, top left), the monolinguals have significantly more looks to the target beginning at 463 until 887 msec, and then again from 1190 to 1635 msec, indicating that the monolinguals look to singular targets more rapidly, and more in general throughout the time window. In contrast, the comparison for plural stimulus items (top right) showed no significant difference between the groups over the two-second time window. For the two levels of stimulus number, the L2s already had significantly greater looks to the target for plural stimuli by the 200 msec intercept, which lasted until 1493 (bottom left), indicating that the L2s have a greater proportion of looks to plural targets than singulars. For the monolinguals, there was a significant difference between the two stimulus types at the 200 msec intercept, but it disappeared at 281 msec (bottom right). This small significant difference was likely the result of an early bias to look at images with multiple items.

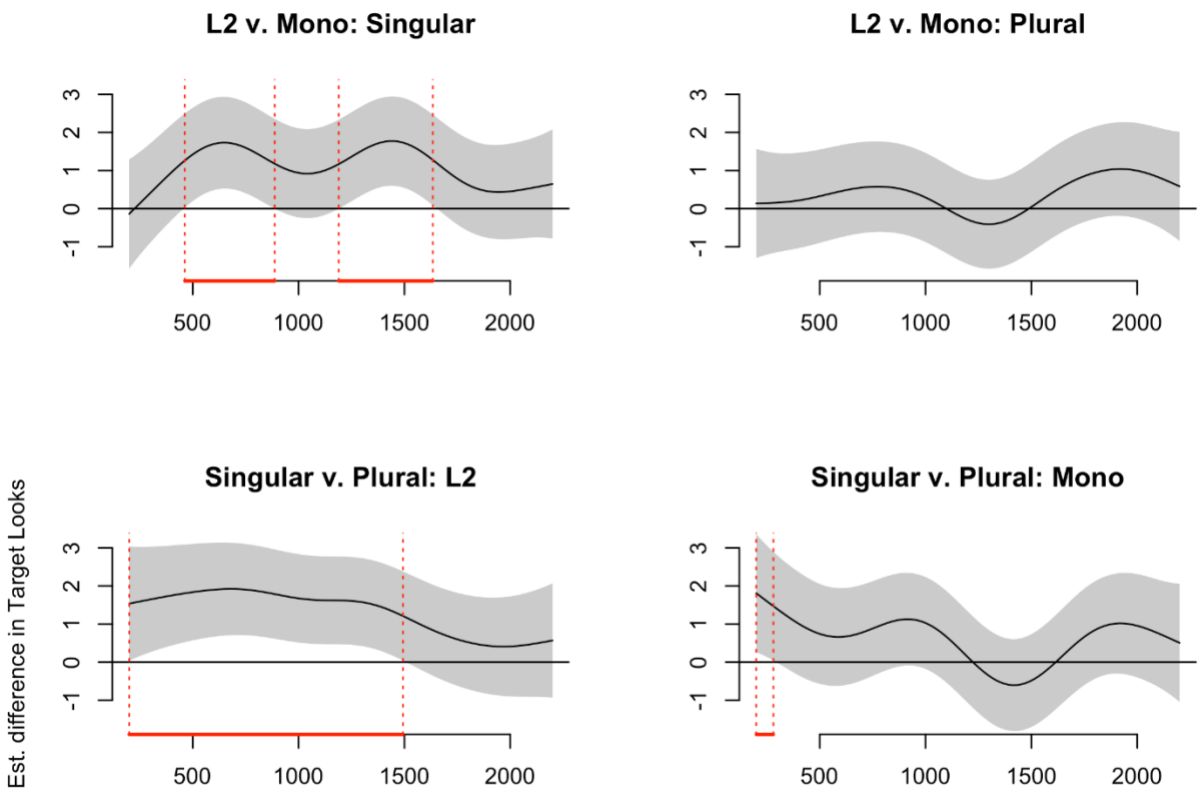

Figure 3.4. This plot shows the output of the *plot_diff()* function from the *itsadug* package for the optimal GAMM model. The plot shows comparisons of the four levels of Condition over time in msec. In each case the first term in the plot label serves as the *x*-axis and the second term is represented as the smooth line in relation. Points where the error bar shading does not cover the *x*-axis represent significant differences and are delineated by vertical dashed lines.

3.4.3. Post-Hoc Analysis

The purpose of the present study was to test comprehension of plural-singular marking, specifically with the expectation that the L2 group is likely to differ in their sensitivity to this feature in relation to their general English comprehension. The evidence presented here demonstrates that this population of L2 English learners use the information provided by inflectional plural-singular marking less consistently, but also that proficiency, in the form of greater receptive vocabulary (*i.e.*, PPVT score), makes a difference for accurately

interpreting plural-singular marking. This raises the question of whether the lower performance for the L2 participants of the present study is due to a general lack of proficiency. Here, additional data from the picture decisions for filler trials and looking behaviour during experimental trials are presented to aid interpretation of this study's findings.

In completing the experiment, participants made button responses to all trials including 40 fillers that required responding to a non-morphological semantic contrast. For example, one such filler trial presented the audio stimulus *The shovel is stuck into the sand* along with a target image of a toy shovel stuck upright into a sandy beach alongside a competitor image of a similar shovel lying flat. For these items, the accuracy of the two groups was more comparable with monolingual accuracy at 93% (*SD* = 4%), and the L2s at 89% (*SD* = 4%). To further compare the group results for filler trials, a mixed effect model predicting accuracy was run with the Group variable in interactions with PPVT score and trial type (either plural-singular or semantic filler coded as -0.5 and 0.5, respectively). Due to overfitting, random effects were reduced to only a random intercept for Item with a random slope for PPVT score. The impact of each of these factors was investigated using estimated marginal means from the `emmeans` package (version 1.3.4; Lenth, 2019). Results of the model are presented in Table 3.5 and show that the group interactions were significant. The significant interaction between Group and trial type is unpacked in pairwise comparisons shown in Table 3.6. These comparisons show that there was no statistically significant difference between the two trial types for the monolingual group while the L2 group was significantly lower for plural-singular trials as opposed to the filler trials. After accounting for receptive vocabulary, a significant difference still existed

between the groups for the semantic fillers. The impact of PPVT score on accuracy is illustrated in Figure 3.5 where comparison of the two panels shows that the monolinguals did not vary greatly among themselves or between the two trial types of plural-singular marking (left panel) and semantic fillers (right panel). In contrast, the L2s showed a much greater range in accuracy in their interpretations of the experimental plural-singular stimuli. While PPVT score significantly impacted the L2s' accuracy for both trial types, the influence was greater for plural-singular trials.

Table 3.5 Model Results: Picture Decision Accuracy

	Estimate	Std. Error	*z*-value	*p*-value	
(Intercept)	2.57	0.17	15.1	<.001	***
Group	1.01	0.23	4.41	<.001	***
Receptive Vocabulary	0.20	0.13	1.53	.125	
Trial Type	1.16	0.29	3.95	<.001	***
Group: Receptive Vocabulary	-0.44	0.22	-2.01	<.045	*
Group: Trial Type	-0.95	0.34	-2.79	.005	**

Note. Receptive vocabulary is the centred standardized score for the PPVT (Dunn & Dunn, 2007). *p<0.05; **p<0.01; ***p<0.001

Table 3.6 Post-hoc Pairwise Comparisons for Group and Trial Type

contrast	type	odds ratio	std. err	*z*-ratio	*p*-value	
L2 - mono	Plural-Singular	0.23	0.07	-4.82	<.001	***
	Semantic Filler	0.58	0.15	-2.06	.04	*

contrast	group	odds ratio	std. err	*z*-ratio	*p*-value	
Plural-Singular – Semantic Filler	L2	0.20	0.06	-5.72	<.001	***
	mono	0.50	0.19	-1.78	.07	

Note. Pairwise comparisons of the Group - trial type interaction in the post-hoc picture decision accuracy model. Plural-singular trials are the experimental stimuli trials. These comparisons show that while the L2 group made significantly less accurate picture

decisions than the monolinguals for both trial types, only the L2 group had a significant difference between the two trial types. *p<0.05; **p<0.01; ***p<0.001

An additional post-hoc test on the eye-tracking data was run with the aim of showing whether L2 participants understood the stimulus nouns. This examined whether upon hearing a stimulus noun such as 'pencils', L2 participants were looking at either of the pencil images, or at distractor images. To test this, an additional GAMM was run as before, but with looks to the distractor images subtracted from looks to the lemma-appropriate (either target or competitor) images as the dependent variable. An initial model was run over the original 2000 msec time window. However, a significant difference emerged late, and therefore the time window was expanded to encompass this entire difference window. (When the data points beyond 2200 msec are put into the model the beginning of this difference shifts beyond 2200 msec.) With this model, both groups show comparable relative looks to either the lemma-appropriate images or distractors until 2382 msec where a significant difference emerges showing that the L2 participants have more relative looks to the appropriate images. This lasts until 3968 msec. Results for this model are reported in Table 3.7. For looks that were recorded but were not in any interest area the proportions were similar across groups at 8% (SD = 3%) and 11% (SD = 5%), for L2s and monolinguals, respectively. This post-hoc analysis of the eye-tracking data indicates that both groups were similar in their ability to comprehend the noun lemmas.

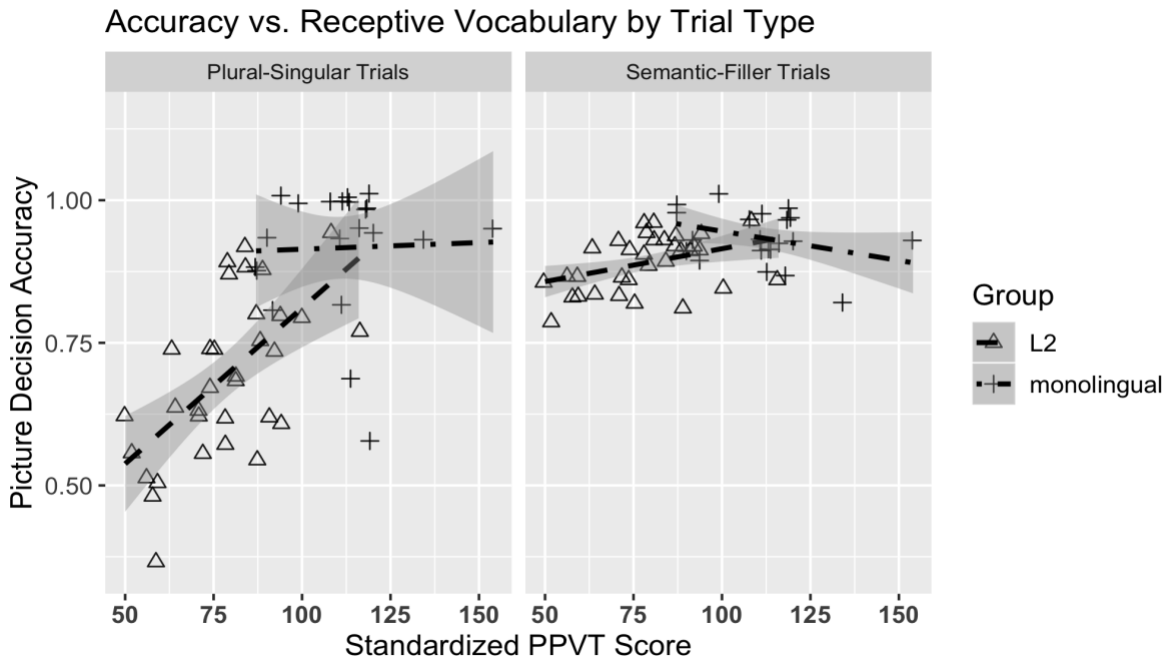

Figure 3.5. Participant accuracy by participant group is plotted against PPVT score for plural-singular trials and semantic-filler trials. Points show individual participant accuracy for each stimulus type.

Table 3.7 GAMM Output for Looks to Lemma-Appropriate vs. Distractor Images

A. parametric coefficients	Estimate	Std. Error	*t*-value	*p*-value	
(Intercept)	4.41	0.34	12.83	<.001	***
Group: Monolingual	-0.90	0.39	-2.32	.02	*
B. smooth terms	**edf**	**Ref.df**	***F*-value**	***p*-value**	
s(TimeIndex): Group - L2	8.46	8.56	37.96	<.001	***
s(TimeIndex): Group: - Mono	7.91	8.12	12.98	<.001	***
s(TimeIndex, Subject)	417.01	475.00	866.23	<.001	***
s(TimeIndex, Item)	253.30	287.00	856.88	<.001	***
s(Event)	757.14	846.00	38.64	<.001	***

Note. This table presents the parametric and smooth coefficients for a model with the difference in combined looks to the target and competitor images minus the combined looks to the 2 distractor images compared across the 2 participant groups. The results of this model showed that the L2 group was statistically significantly more likely to be looking

at the target and competitor images within a late window from 2382 msec to 3968 msec. *p<0.05; **p<0.01; ***p<0.001

3.5. Discussion

The purpose of this study was to investigate the comprehension of English plural-singular marking, a highly frequent and semantically transparent morphosyntactic feature, by middle-school-aged, Mandarin-L1 speakers who have learned their L2-English in immersion classrooms since preschool. An age-matched monolingual control group was included to aid interpretation of the immersion learners' task performance. The study employed a visual-world eye-tracking task, combined with a task that involved making a picture decision based on the presence or absence of the plural '*-s*' morpheme. Findings indicated that there were differences between these English immersion learners and the monolinguals in their online comprehension of plural-singular marking and in their final interpretations. These differences emerged despite the L2-immersion learners typically having begun English immersion in preschool and having more than 8 years of L2 exposure.

3.5.1. L2 Plural-Singular Marking Comprehension: Offline Interpretation and Online Sensitivity

The first research question asked if the L2 group would differ in their offline interpretation of plural-singular marking, and it was hypothesized based on research for community-L2 learners and late instructed-L2 contexts that they would be less accurate in responding on the picture decision task. Results for the picture decision task revealed that L2 participants were significantly less accurate than the monolinguals in their ultimate interpretations of the audio stimuli. The task used here was not designed to be difficult, and as expected, the

monolinguals made highly accurate picture decisions (92% accurate) with 9 of the 21 monolinguals making no errors. In contrast, no L2 participant was 100% accurate, and only 2 of the 33 scored above 90%, with group performance at an average of 69%. Also as hypothesized, receptive vocabulary size, used here as a proxy for English proficiency, was a predictor of accuracy. However, after the variance explained by receptive vocabulary was accounted for, the influence of participant group remained. This evidence shows that the L2 participants were less likely to arrive at an accurate ultimate interpretation of plural-singular marking as compared to the monolinguals, though participants with larger receptive vocabularies were more likely to arrive at a picture decision consistent with the marked stimulus noun.

For picture decision response times, the only significant predictor was receptive vocabulary. This means that participants of either group were similarly fast when making a correct response to a trial and without regard to whether the stimulus was plural or singular. When L2 participants correctly interpreted the audio stimulus, they did so as quickly as the monolinguals. The difference is that these correct interpretations happened less often for the L2 participants. As participants were given a limited time to respond, the L2 participants may have been impacted by a speed-accuracy trade off, which does not appear to be the case for the monolinguals, though additional research is required to determine if, given more time, the L2s' response accuracy would increase.

The second research question asked about online interpretation of plural-singular marking. In line with what was hypothesized, participant group predicted significant differences in looking behaviour. However, this was limited to singular stimulus trials, with plural trials showing no difference from monolinguals (see Figure 3.4). For singular trials,

an early difference arose between the L2 participants' and monolinguals' looks to target images indicating that the L2 participants typically had a delayed response to singular nouns as compared to monolinguals. In contrast, on plural stimulus trials, the L2 participants were able to mirror the timing of monolinguals' increasing looks to plural targets.

When the two stimulus types are compared directly, the L2 participants also showed an advantage for looks to plurals as compared to looks to singular targets. The L2 participants' looks were significantly more likely to correspond to the appropriate image on plural trials as opposed to singular trials and this difference lasted from cue onset until about 1500 msec (see the lower left panel of Figure 3.4). Though it appears that the L2 participants had an advantage for comprehension of plurals, this did not translate into greater ultimate accuracy on the picture decision task, where there was not a significant difference between accuracy for the plural and singular items. In sum, the hypothesis that the L2 group's looks to target images would be delayed and be lower in proportion only held for singular stimulus items. In addition, it was hypothesized that receptive vocabulary would have an impact on speed and proportion of looks. However, no statistically significant impact of receptive vocabulary was found.

In order to aid in interpretation of these results, 2 post-hoc tests were run. A potential explanation for the accuracy difference found for the picture decision task is that the immersion learners simply had low proficiency and that findings would not be unique to interpretation of plural-singular marking. However, a post-hoc test for accuracy with the filler items showed that the L2 group had significantly lower accuracy in their interpretation of the plural-singular stimuli than for the semantic filler items. In contrast,

this was not true for the monolingual participants who did not differ in their accuracy for either trial type. Despite the much smaller range in accuracy for filler trials, receptive vocabulary was still a predictor of accuracy for the L2 group. This test shows that, like research for immersion learners in other contexts, these L2 participants have good general L2 comprehension, but can exhibit differences when specific morphosyntactic features are scrutinized.

In addition, a post-hoc GAMM was run on the difference between looks to the experimental images (either the target or competitor images) versus the distractors to see if L2 participants had online difficulty interpreting the stimulus nouns. This showed that the L2 participants were as likely to be looking at lemma-appropriate images as the monolinguals within the 200 to 2200 msec post-cue time window, and more likely to be looking at the experimental images after the experimental time window. However, this later 'advantage' in looks to the target and competitor came at a timepoint when looks to these images were decreasing for both groups (see Figure 3.3) and therefore represents a slower decrease rather than an increased response. That the L2 participants tend to look at the experimental images for longer likely reflects their lesser certainty in determining the target image based solely on plural-singular marking.

A seemingly incongruous finding of the present study was that the L2 participants appeared to interpret plural marked nouns as consistently as monolinguals in online comprehension, but this did not translate into either monolingual-like accuracy on the picture-decision task or even significantly greater accuracy than for singular stimuli. However, it is not uncommon for participants to differ in their performance during an online task and when completing a subsequent explicit offline task similar to that of the

present study. For young children, eye-movements have demonstrated understanding of others' false beliefs (Clements & Perner, 1994; Oniski & Baillargeon, 2005), or verb number-agreement marking in L1 German (Brandt-Kobele & Höhle, 2010) even when this was not reflected in children's explicit answers. Thus, online tasks can show understanding earlier in development than explicit tasks. This should be predicted given that for tasks like eye-tracking, even a partial representation of a feature can be beneficial (Chang, Dell, & Bock, 2006), whereas, accomplishing an interpretive task requires integrating that information into a full utterance, a task likely to require a more robust linguistic representation.

Overall, this study supports research for child community-L2 learners (Jia, 2003) and late-L2 learners (Jiang, 2007; Jiang et al., 2011) that shows speakers of languages with optional number marking can have differing attainment for English plural-singular marking. Findings also support studies of language immersion that show differing attainment for some features of L2 morphosyntax (*e.g.*, Ó Duibhir, 2011; Harley, 1993; Warden, 1997). This study joins Lew-Williams (2017) in showing that it is not only the morphosyntactic production skills of immersion learners that can be divergent from those who speak the language as an L1. In addition, this study also shows that differences in immersion learners' comprehension are not limited to abstract features like grammatical gender.

3.5.2. Acquiring L2 Morphosyntactic Constructions in Immersion

Like Lew-Williams (2017), this study finds that early immersion learners can have differences in L2 comprehension. In the Lew-Williams study, the author found that unlike

Spanish-L1 peers, English-L1 classmates could not identify upcoming nouns by using abstract grammatical gender but were able to use the more semantically transparent features of number and notional gender. However, it remained unclear whether it was semantic transparency or that English has both number-marked determiners and morphological forms that reflect notional gender (*i.e.*, pronouns) for which processing skills could be transferred. In the present study, semantically transparent bound plural-singular morphology was examined. However, unlike Spanish and English, Mandarin and English differ with regard to whether number marking is obligatory. Despite having 3 fewer years of exposure to their L2 than the L2 participants of the current study, the English-L1 participants of Lew-Williams (2017) were similarly responsive to Spanish number marking on a task that reflected rapid comprehension as compared to their Spanish-L1 peers. In the present study, the differences between the participant groups for both the eye-tracking and the picture decision tasks maintain after more than 8 years of immersion indicating that for plural-singular marking, semantic transparency is not enough to ensure that L2-immersion learners eventually make use of this contrastive L2 feature in the same manner as those who have learned the language as an L1.

Despite the lower task accuracy for L2s, given a binomial distribution, most participants of the L2 group were above chance in their responses for both singular and plural stimulus items indicating that they have established receptive knowledge of plural-singular inflection. Thus, the present study's findings may suggest these participants' plural-singular marking knowledge is at an intermediate stage of development. It is important to note that in the Lew-Williams study, participants responded to number marked determiners rather than bound morphemes. For L1 English, research indicates that

the bound morpheme on the noun may be a later acquired number marker as compared to features (or the combination of features) such as determiners and verb agreement (Kouider et al., 2006; Wood et al., 2009) and thus, the differences in findings between the two studies may be the result of construction difficulty. If this is the case it is possible that the participants of the Lew-Williams study would have had lower accuracy if limited to bound morphology and that the participants of this study would do better if number were also marked on determiners (the latter having a strong likelihood given that in Mandarin plural and singular can also be marked this way when appropriate to the discourse context). As a result, the findings of the present study may not be the result of differences between the L1 and L2.

In addition, prior L1 developmental research has found that young children developing comprehension of plural-singular marking can be more sensitive to plural forms than singular (see Arias-Trejo et al., 2014 for L1 Spanish; Xu Rattanasone et al., 2016 for L1 English and L2-English speakers without a Chinese L1). Thus, the present study's findings that the L2 participants had low accuracy for both plural and singular, but more looks to plural stimulus targets may also suggest these participants' plural-singular marking knowledge has been assessed at an intermediate stage of development. However, Xu Rattanasone et al. (2016) found that L1-Chinese participants were more accurate with singular stimuli and thus did not follow this pattern. Though this study cannot rule out the possibility that the L2 participants are still developing this knowledge, it is unclear how well the L1 developmental patterns of children 2 to 4 years of age generalize to adolescent L2 learners with more than 8 years of L2 exposure.

3.5.3. Connections to Late L2 Learners

Findings for late learners also support that knowledge of plural-singular marking can be late emerging for speakers of an optional-number L1. Studies that have probed receptive sensitivity to English number marking for these participants using self-paced reading (*e.g.*, Jiang et al., 2011; Jiang, 2007; Song, 2015; Wen et al., 2010) have differed as to whether their respective participants were as sensitive to written grammatical errors as native-speaking controls. Wen et al. (2010) and Song (2015) found that advanced L2-English speakers were sensitive to the marking, but that intermediate speakers were not, indicating that this knowledge does not appear until a very high level of English proficiency is attained. In contrast, this delay does not appear to apply to English learners whose L1 also has obligatory number marking as Jiang et al. (2011) compared proficiency-matched L1 groups and found only the optional number-marking-L1 group (Japanese) were less sensitive to plural-singular marking errors, while both groups performed similarly in detecting verb subcategorization errors.

An additional concern within the late-L2 literature was stimulus complexity that required noticing agreement errors beyond multiple phrase boundaries as in **several of the board member* (Jiang et al., 2011). In the present study, participants were only required to interpret the noun in its correct plural or singular morphological form within a stimulus sentence. Using these stimuli, both the picture decision task and eye-tracking data indicated that the participants differed from the monolingual participants. Generalizing between participant groups that differ in age, type of L2 exposure, and across experiment types should be done with caution. However, evidence that proficient child L2 learners can

also have non-target acquisition with features of an L2 indicates that the issue of L1-L2 differences is not (only) rooted in maturation, and that other factors can play a role.

3.5.4. Ultimate Attainment in Immersion

Despite the years of exposure that the L2 group had received, a limitation of this study is that it does not provide conclusive evidence that the participants had reached their ultimate L2 attainment. The results of the present study show that the L2 participants were above chance in their interpretation of plural-singular marking indicating they had acquired knowledge of the L2 feature, though it is unclear why they use this marking with a consistency that differs from the monolinguals. Though the current study's results show parallels with developmental patterns in L1 acquisition, explaining these results as the product of either continuing or incomplete attainment may not accurately capture usage within this context. Based on English majority language contexts, plural-singular marking can be considered a frequent and reliable morphosyntactic feature. However, this assumption may not generalize to immersion classrooms (or other foreign language learning contexts). A potential advantage in studying immersion learners within a single school is that the participant group is both a study sample, and a sample of the individuals with whom each participant is likely to interact. The findings of this study indicate that the interaction these L2 learners have with classmates is likely to involve seemingly less reliable use of plural-singular inflection and given that a feature of immersion contexts is that interaction is largely limited to the classroom (Swain & Johnson, 1997), this may account for a substantial proportion of their L2 interaction. A potential alternative explanation for the results found here is that the L2 participants' usage of plural-singular

marking is neither continuing in development toward a monolingual standard, nor in a 'fossilized' state short of that standard, but rather at a stable stage of development influenced by and potentially consistent with L1 pragmatic usage.

In studying Mandarin-L1 children and adolescents, Jia (2003) observed that a participant might provide obligatory plural marking on a particular noun, but then later produce the noun without the obligatory marking, even within the same testing session. Importantly, neither Jia nor the current study can provide evidence that these differences in language use were unsystematic errors, and not a consistent and accurate reflection of pragmatic usage within the respective L2 users' linguistic context. Importantly, in a later study of the same participants, Jia and Fuse (2007) found that whether an individual's linguistic environment favoured the L1 or L2 impacted the likelihood of standard use of inflectional morphemes. Thus, it is possible that participants whose linguistic environments were more Mandarin dominant produced plural marking influenced by Mandarin pragmatics. Potentially, both Jia's participants and the L2 participants of the present study were more likely to interpret and produce plural-singular marking when assumed to be contextually relevant, a possibility that can be probed in future research. In addition, given the possibility these immersion learners rely more on other quantifiers such as number-marked determiners, future research should aim to link L2 participants' preferences for marking number in production with their language comprehension ability (*i.e.*, Do L2 learners who produce more quantifiers rely more on quantifiers in comprehension?).

3.6. Conclusion

The goal of this study was to examine immersion learners' comprehension of L2 morphosyntax. This study shows that, unlike global measures of language comprehension, more fine-grained online measures and more targeted explicit tasks demonstrate that differences exist between monolinguals and immersion learners with an early L2 start and extensive exposure to the L2. The present study provides evidence that these immersion learners do not interpret English plural-singular marking identically to monolingual English speakers. However, it remains unclear why these differences should exist for a semantically transparent, highly frequent morphosyntactic feature.

4 How out-of-school L2 input and L2 use impact adolescent L2 vocabulary across the classroom immersion and community-L2 contexts: A richer L2 environment is key

4.1. Introduction

The specifics of a child's second language (L2) learning environment can impact their L2 development (Paradis et al., 2020). One aspect that has numerous implications for a child's L2 environment is whether the L2 is a language widely spoken in the community or not. For example, some children speak a minority first language (L1) at home and then begin learning the community majority language as an L2 when they enter school. These community-L2 learners will have a large advantage in terms of opportunities for L2 experience over children who are learning an L2 not widely used in the community. Nevertheless, building L2 proficiency in either context is possible. That learning a non-community L2 is possible is especially illustrated by language immersion learners, who are exposed to the L2 in classrooms in which they are expected to learn mainstream educational curricula through content instruction in the L2 (*i.e.*, they are taught about math or science rather than explicitly about the L2). When compared to peers in traditional instructed-L2 classrooms who receive explicit L2 instruction for only a few hours a week, immersion learners rapidly demonstrate far superior L2 ability that can eventually be on

par with L1 speakers on tests of listening comprehension (Genesee & Lindholm-Leary, 2013; Swain & Lapkin, 1982).

Both community-L2 and immersion learners are similar in that they receive extensive and frequent exposure to the L2 through content lessons in school. However, the L2 experiences of community-L2 and immersion learners outside of school are likely to be very different. These differences in exposure should be expected to have important implications for L2 development, but currently little is known about how the out-of-school L2 experiences of these two types of L2 learners compare, or how they may impact L2 development. In order to investigate this topic, the present study examined differences between the out-of-school L2 environments of learners in the two contexts. In addition, to understand how these differences may impact L2 ability, their potential impact on L2 receptive vocabulary knowledge was also examined. Both participant groups were adolescent L1-Mandarin speakers learning L2-English.

4.1.1. L2 Outcomes in the Immersion and Community-L2 Contexts

For many L2 learning children, the L2 is learned as the majority language of a community after they have spent several years acquiring an L1 at home. In addition to L2 exposure in the wider community, these community-L2 learners receive L2 exposure through mainstream schooling alongside L1 speakers of the majority language and it is commonly observed that they attain L2 abilities that closely match their L1 speaking peers (DeKeyser, 2012; Muñoz & Singleton, 2011). That beginning L2 learning in early childhood results in more L1-like ultimate attainment is a consistent finding for L2 learners in this context (*e.g.*, Abrahamsson & Hyltenstam, 2009; Flege et al., 1999; McDonald, 2000), and is consistent

with the experience of countless adult immigrants whose children attain a level of proficiency in the language of their new community that they generally do not.

In order to capitalize on this apparent advantage, early immersion programs in Canada and the US were founded on the assumption children could rely on L2 input and their natural incidental language learning mechanisms to develop L1-like proficiency in an L2 (Fortune et al., 2008; Genesee, 1984). Essentially, it was hoped that the success of community-L2 learners could be reproduced within immersion classrooms. However, children in immersion contexts do not reliably attain L2 abilities that parallel L1 speakers of the target language, though they still achieve high levels of L2 communicative proficiency (Lyster & Genesee, 2019).

Gaps in L2 ability are observed particularly for immersion learners' more limited range of vocabulary, and in their morphosyntactic accuracy (Cummins, 2009), despite years of L2 exposure and an early start. Specifically for their vocabulary, performance on a monolingual-normed assessment of receptive vocabulary shows that French immersion learners in grades two and five (*i.e.*, their third or sixth years of immersion) fall more than one standard deviation below monolingual French speakers on average (Hermanto et al., 2012). Presumably to compensate for this lag in L2 vocabulary, immersion learners throughout their programs are observed to stretch their vocabulary resources, particularly for verbs, by resorting to words with more general meanings where L1 speakers would use a more exact term (Harley, 1992). However, this compensation strategy is also observed for community-L2 learners in their first few years of L2 exposure (Golberg et al., 2008).

Despite similar early use of this compensation strategy, studies of community-L2 learners' receptive vocabulary show that these learners steadily catch up to their

monolingual peers (Farnia & Geva, 2011; Golberg et al., 2008), with individuals potentially meeting or exceeding standard scores on monolingual-normed assessments by four to six years of L2 exposure (Paradis & Jia, 2017). When compared to peers within the same schools, community-L2 learners with more than seven years of L2 exposure can attain target language vocabulary knowledge on par with monolingual controls (Soto-Corominas et al., 2020). Therefore, in contrast to immersion learners, it appears that the gap between community-L2 learners and monolinguals is eventually greatly diminished, if not disappearing altogether. However, precisely how community-L2 and immersion learners' vocabulary knowledge compare is unclear as these groups have never been compared directly. In addition, it is also unclear which factors associated with these two L2 learning contexts contribute to potential differences in vocabulary knowledge.

4.1.2. Sources of Individual Differences in Early L2 Acquisition

Currently, research on child L2 acquisition in community-L2 contexts has increasingly examined factors that underlie differences in the rate of development between individuals. This research is motivated by the position that an understanding of how these factors can impact an individual's L2 development is informative about the processes and mechanisms that contribute to early L2 acquisition (Chondrogianni, 2018; Paradis, 2019, 2011). Research examining sources of individual differences in learners' L2 abilities may investigate *learner-internal* factors that include differences in cognitive abilities like memory or analytical skills, or in relationships between different linguistic subsystems (*e.g.*, morphosyntactic and vocabulary knowledge; or, L1 and L2 knowledge). In addition, this research also the impact of *learner-external* factors that relate to the individual's

experiences with the L2. These learner-external factors are expected to be significantly impacted by learning context.

4.1.3. Learner-External Sources of Individual Differences in L2 Vocabulary Ability

As a source of individual differences in L2 ability, differences in learning context, such as those between the immersion and community-L2 context, serve as a learner-external macro-variable indexing a range of more proximal micro-variables that impact L2 development. Micro-variables include sources of L2 experience like opportunities to use the L2 with peers or at home, as well as access to forms of media in the L2, all of which would correlate with whether or not learners reside in a community where the L2 is widely used. For community-L2 learners, these learner-external factors have received increasing attention as they have been demonstrated to impact L2 development (*e.g.*, Jia & Fuse, 2007; Paradis et al., 2016, 2020; Rojas et al., 2016). By contrast, the L2 experiences of individual immersion learners have remained under studied. The immersion context has been characterized as one where participants' L2 exposure is confined to classrooms (Swain & Johnson, 1997). If this is the case, then the potential for individual variation in learner-external factors would be limited. However, this description of the immersion context pre-dates widespread access to the Internet. Currently, youth around the world have access to additional sources of L2 media and communication, and thus, the extent to which immersion learners' L2 experience remains limited to the classroom is currently unclear and worth re-evaluating.

For community-L2 learners, one learner-external variable showing a consistent association with gains in L2 vocabulary ability is the length of time learners are exposed to

the L2 (Carhill et al., 2008; Chondrogianni & Marinis, 2011; Cobo-Lewis et al., 2002; Hammer et al., 2012; Paradis & Jia, 2017). However, the rapid vocabulary expansion of L2 vocabulary in the early years eventually appears to slow. Farnia and Geva (2011) found that L2 vocabulary knowledge increased for learners who were followed longitudinally from first to sixth grade, but that the degree of impact was reduced as cumulative exposure increased over time, a finding supported in other research (Soto-Corominas et al., 2020).

For immersion learners, the benefit of more years of immersion on vocabulary knowledge is evident in the early stages of immersion (Harley & Jean, 1999; Kalia et al., 2018; Nicolay & Poncelet, 2013), but, similar to community-L2 learners, rapid early vocabulary development appears to eventually slow down. When compared to early immersion learners, late-start learners (*i.e.*, those who start in seventh grade, rather than kindergarten) begin to catch up to early-start peers (Harley & Jean, 1999). Though early-starters typically still out-perform late-starters, for L2 abilities in general, findings indicate that, after only a few years of immersion, late-starters can often attain similar proficiency as the early-starters who have had many more years of study (Turnbull et al., 1998). This indicates that the benefits of length of L2 exposure begin to diminish for immersion learners as is the case for community-L2 learners.

Beyond simply longer exposure, researchers have also looked at factors of L2 experience that may contribute to a richer L2 environment outside of school. For community-L2 learners, the richness of the out-of-school L2 environment is often examined as an aggregated variable made up of the diverse sources of L2 exposure they can potentially access in addition to their L2 exposure in school. Findings indicate that the richness of this type of exposure can impact L2 vocabulary development (Paradis, 2011;

Paradis et al., 2020; Paradis & Jia, 2017). Sources that contribute to a richer L2 environment include participation in extracurricular activities, interaction with peers, listening or watching media, and reading. Research has also shown that these specific sources can make an individual contribution. For example, language use with peers can be an important determinant of adolescent L2 learners' lexical knowledge (Carhill et al., 2008; Soto-Corominas et al., 2020). Watching educational television programming in the L2 outside of school hours has also been found to benefit preschool-aged community-L2 learners (Uchikoshi, 2006). For literacy activities, reading to pre-literate community-L2 learners benefits their L2 vocabulary knowledge (Collins, 2010; Prevoo et al., 2014). However, little research exists for the impact of out-of-school literacy activities on the vocabulary knowledge L2 learners in adolescence. Another learner-external factor that has had equivocal results is home language use. For community-L2 learners, speaking the L2 at home can be a significant predictor of greater L2 vocabulary (Cobo-Lewis et al., 2002), but this link between home L2 use and L2 vocabulary is not always found (Paradis, 2019, 2011).

An additional source of L2 exposure that is under researched for school-aged community-L2 learners is language use via mobile mediated written messaging such as SMS messages or messaging applications (*e.g.*, WhatsApp or WeChat). Existing research shows that community-L2 learners have been found to have more limited lexical diversity when messaging in their L2, with those who message more frequently in the L2 having the smallest range of vocabulary (McSweeney, 2017). This finding suggests that the benefit of using the L2 for messaging may only have limited benefit for vocabulary ability. Nevertheless, findings for monolingual children indicate that increased text messaging can

have a positive impact on vocabulary (Plester et al., 2009), but even for monolinguals, the relationship of text messaging to language skills is not fully understood (Verheijen, 2013).

For immersion learners, there is a lack of research examining the richness of L2 experiences beyond the classroom and how these may impact L2 vocabulary ability. However, evidence that greater L2 use with acquaintances outside the classroom predicts better general L2 ability for learners in their eighth year of immersion (Hamayan et al., 1977), which indicates that these learners can also benefit from L2 exposure outside of school. In contrast to immersion, there is a growing body of research looking at the impact of L2 exposure outside of classrooms on the development of a foreign language for instructed-L2 learners (see the recent volume by Dressman and Sadler (2020)). For the L2 exposure of school-aged, instructed-L2 learners outside of their instructional time, studies find that L2 vocabulary ability benefits from additional sources of L2 exposure such as watching films or television (Kuppens, 2010; Lindgren & Muñoz, 2013), listening to music (Pavia et al., 2019), playing video games (Kuppens, 2010; Peters, 2018; Peters et al., 2019), viewing social media (De Wilde et al., 2019), receiving text messages (Lu, 2008) and reading (Webb & Chang, 2015; Zahar et al., 2001).

As both immersion learners and instructed-L2 learners live in communities where the L2 is not widely spoken, they are likely to have similar limitations on opportunities for L2 exposure outside of school. However, despite the benefits to vocabulary knowledge that these additional sources of L2 exposure can have for non-immersion L2 learners, it is unclear whether these should have similar benefits for immersion learners. It has been argued that factors associated with out-of-school language experience may have a greater impact in these traditional instructed-L2 contexts (Muñoz, 2014; Sun et al., 2016). The

reason for this is that most school-aged instructed-L2 learners are likely to have only a few hours a week of L2 exposure and thus, small differences in informal L2 exposure can add relatively larger percentages of total L2 exposure (*i.e.*, for a learner who receives two hours of L2 instruction a week, watching one hour of television programs in the L2 would be a 50% increase in weekly L2 input). This is illustrated by a study of Swedish high-school content and language integrated learning (CLIL) L2-English learners. For these learners, Olsson and Sylvén (2015) found that out-of-school sources of L2 exposure did not benefit their written L2 production of academic vocabulary. Given that their in-school CLIL exposure should be a rich source of academic vocabulary, the potential benefit of L2 exposure for online gaming or watching movies is overshadowed. These L2 experiences may be valuable for learners who have limited in-school contact with the L2, but are not rich enough to impact those with richer in-school L2 experience.

4.1.4. L2 Input and L2 Usage in Immersion

L2 learners in immersion contexts can amass thousands of hours of in-school exposure, with early immersion learners estimated to have around 6,000 hours of L2 exposure by the end of eighth grade (*i.e.*, ninth year of study, Turnbull et al. (1998)). Despite the quantity of L2 exposure immersion learners receive, Swain (1985) argued that most of the exposure comes in the form of L2 input from teachers, with a relative lack of conversational exchange (*i.e.*, interactional use), a finding also supported by subsequent research (Allen et al., 1990; Swain, 1988). Contrary to approaches that ascribed the process of L2 acquisition entirely to authentic/comprehensible input that were popular throughout the 1980s (*e.g.*, Krashen, 1985), Swain indicated that the lack of opportunities for L2 *output* that are a part

of interactional L2 use could be a potential limiting factor on the L2 development of immersion learners (or any L2 learner). Subsequent research has shown that, in addition to L2 input, opportunities for L2 output are also important (Izumi et al., 1999; Kowal & Swain, 1997). Both immersion and community-L2 learners groups of L2 learners receive extensive and frequent L2 input in school. In contrast to their in-school L2 experience, there should be greater differences between the groups' respective opportunities for L2 experiences outside of school as immersion L2 learners are characteristically considered to have their L2 experience largely limited to the L2 classroom (Swain & Johnson, 1997). However, as Swain (1985) noted for L2 experience in immersion classrooms, not all L2 experience provides equal opportunities for both L2 input and L2 output. If classroom experiences tend to provide limited opportunities for interactional L2 use, then it may be that, for both types of L2 learner, out-of-school L2 exposure is a key source of opportunities for sources of L2 experience where input and output opportunities are more balanced. Thus, when examining the out-of-school sources of L2 exposure that L2 learners receive it is important to consider whether these are largely opportunities for increased L2 input or whether they are opportunities for interactional L2 use. If a lack of opportunities for interactional L2 use rather than a lack of L2 input limits L2 development in immersion, it may be that both immersion and community-L2 learners, but immersion learners in particular, benefit more from additional opportunities for interactional L2 use. As a result, the present study examines out-of-school L2 experience as either a source of greater L2 input or greater L2 interactional use.

4.1.5. The Present Study

The present study compared the out-of-school L2 learning environments of adolescent L2-English learners in two different learning contexts and examined how these differences in learning context can impact L2 receptive vocabulary ability. Learners were either acquiring L2-English in a community-L2 or immersion context. Both groups were comprised of L1-Mandarin speakers. In addition to L1, the groups were matched for age of English education onset, age at testing, and level of parental education. Participants of both groups had begun learning English before five years of age on average, and were currently in middle school (grade seven, eight, or nine in a K-to-12 system) giving them extensive exposure to English. The age and English proficiency of the two groups make them capable of accessing a wide range of L2 experiences including both traditional and social media.

The present study has three research aims. The first was to examine how L2 input and interactional L2 use outside of school compare for adolescent L2 learners in either a community-L2 context or an immersion context. It was expected that participants in each context should differ, with the community-L2 learners having greater amounts of L2 input and opportunities for interactional L2 use.

The second aim was to see if the L2 receptive vocabulary ability of participants differed across the two contexts. Given the expectation that community-L2 learners have more L2 input and use, it was also expected that the community-L2 participants would have greater L2 vocabulary ability.

Finally, in order to understand how L2 experience impacts L2 development within each context, potential predictors of individual differences in L2 receptive vocabulary were examined against the vocabulary ability of participants within each group. This included

examination of the out-of-school learner-external L2 input and interactional use factors compared between contexts. These aims are addressed in the following three research questions:

1. How do the adolescent L2 learners in the immersion versus the community-L2 context differ in their access to sources of L2 input and interactional L2 use outside school?
2. How does the receptive vocabulary knowledge of the immersion learners compare to the community-L2 learners? Do the immersion learners have equivalent L2 vocabulary sizes to the community L2 learners?
3. How do learner-external differences in L2 input and use predict individual variation in the L2 receptive vocabulary ability of immersion learners and community-L2 learners? Are additional sources of interactional L2 use more predictive of L2 vocabulary ability than additional sources of L2 input? Is there a difference in how predictive these factors are depending on the learner context?

4.2. Methods

4.2.1. Participants

The two comparison participant groups for this study were comprised of middle-school-aged adolescents studying in either grade seven, eight, or nine (mean age = 13;10, *SD* = 8 months). The first group was the immersion-L2 group. Participants for this group were Taiwanese L1-Mandarin, L2-English immersion students (*n* = 33) studying at a private English-language high school in Northern Taiwan. The community-L2 group (*n* = 31) was comprised of adolescents living in Canada and attending mainstream Canadian schools in large English-majority cities who were also L1-Mandarin speakers.

For the immersion-L2 group, data were collected from 37 participants. From this total, the data of four participants were excluded from analysis. Of these, two were participants with fewer than five years in immersion whose early English experience had come from instructed-L2 contexts. In addition, one participant had a native-English speaking parent, and would potentially be better classified as a heritage language speaker.

Finally, one participant was an outlier for receptive vocabulary, scoring approximately three standard deviations (SDs) below the mean for the immersion learners and one *SD* below the next lowest participant. The remaining participants had been enrolled in immersion classes taught by foreign teachers from English-speaking countries (*e.g.,* the UK, US, Canada, etc.) since preschool (mean age of English onset = 4;7, *SD* = 17 months). After preschool, the students attended a bilingual elementary school where half their instruction was in their L1, Mandarin, from a local teacher, and half in the L2, English, from a foreign teacher. On average, they had attended an immersion program for eight years and five months (*SD* = 19.9 months). All study participants were in good academic standing (*i.e.*, they were passing their classes).

The community-L2 group participants were recruited in Edmonton and Vancouver, both English-majority cities in Western Canada. After local public school boards reviewed the study research proposal and ethics approval, the study was referred to the administration of individual schools for participant recruitment. Recruitment also included one charter school. The community-L2 participants had only been formally educated in English. To match the immersion-L2 participants, only community-L2 participants who spoke Mandarin as an L1 and were first educated in English within the age range of the immersion participants (min age = 2; max age = 8) were included. Community-L2 participants mean age of English onset was 4;0 (*SD* = 13.9 months) and they had attended mainstream schools in an English-majority community for nine years and eleven months (*SD* = 16 months).

4.2.2. Materials and Procedures

English Vocabulary Size. The Peabody Picture Vocabulary Test [PPVT; Dunn and Dunn (2007)] was used to assess receptive vocabulary. The test requires the experimenter to say a word. Then, the participant chooses the appropriate one of four pictures presented. The age standard score is 100, with an *SD* range of 85-115.

Alberta Language Environment Questionnaire - 3 (ALEQ-3; Soto-Corominas et al. 2020) Participant data were collected using the ALEQ-3. Information was collected for variables that were expected to vary with L2 learning context, as well as for general participant background data. The questionnaire included two components; a take-home portion completed by parents, and a portion given directly to the participant in the form of an interview where the researcher asked questions and filled out the questionnaire. In the take-home portion, parents provided background information about when their child began education in English, and about parents' educational background. Parental education was blocked on a three-point scale from 0 to 1 with 0 for no post-secondary education, .5 for some post-secondary and 1 for an undergraduate degree or more education. Information for age of beginning English education was used to calculate length of English exposure for each group. For the immersion-L2s, this reflects years in English immersion classes. For the community-L2s, this variable reflects length of time since beginning education in English. These background data are presented in Table 4.1.

Information about learners' L2 environment was collected directly from participants. A range of questions probed L2 experience in terms of both input and interactional use. These were used to compare participants' L2 learning environments across the two L2 learning contexts. For interactional use, participants were asked about

whether they tended to use their L1-Mandarin or L2-English more with their parents and siblings and this information was converted into a variable reflecting home language use. A similar variable was created for language use with friends/peers. These variables were reported on a 1-5 scale [1 = English never and L1 always (90% or more), 2 = English seldom and L1 usually (30%/70%), 3 = English 50% and L1 50%, 4 = English usually L1 seldom (70%/30%), 5 = English almost always (90% or more) and L1 almost never]. Participants also reported the number of messages or emails to friends and family using either their L1 or English that participants sent and received. From these numbers, a relative proportion of messages in English was calculated. In addition, participants reported hours per week spent in various activities using English outside of school. These include time spent using English for watching or listening to media (English TV, video, & music), reading, viewing social media and/or playing video games with language content (social media & gaming), or participating in extracurricular activities (*e.g.*, sports teams or music lessons).

Variables were grouped according to whether they mainly constituted a source of L2 input or were sources of L2 input *and* L2 output. Sources of both input and output were categorized as *interactional L2 use* (or just *L2 use*). These groupings broadly reflect the anticipated balance of input and output each L2 source provides. If a source of L2 experience should involve roughly equal opportunities for L2 input and output, it was classified as L2 use. Otherwise, it was classified as *L2 input*. Nevertheless, sources of L2 experience cannot be considered entirely categorical. For example, immersion learners have some in-class opportunities for L2 output. However, immersion research shows that these opportunities are limited (Allen et al., 1990; Swain, 1985, 1988), and that immersion

classrooms are mainly sources of L2 input. In line with this classification, therefore, classroom L2 exposure would be classified as a source of L2 input.

Sources of out-of-school L2 input included reading; watching video or listening to music; participating in extracurricular activities; and engaging with social media or video games with a language component. Engagement with traditional media and reading are unambiguously sources of L2 input as opposed to L2 use given that they do not offer opportunities for L2 output. However, this not the case for either participation in extracurricular activities or engagement with social media and gaming. Nevertheless, both of these sources should offer substantially more L2 input relative to opportunities for output. For extracurricular activities, these are likely to be structured similarly to classrooms where L2 experience comes mainly from the speech of a coach, instructor, or other individual leading the activity, and as a result, this source of L2 experience was also classified as a source of L2 input.

Also, like classrooms, social media provides the opportunity for both input and output. However, where data are available, they show that a typical social media user consumes much more content posted by others as opposed to posting their own content. For example, industry data for Facebook users around the world show that users are more than twice as likely to *like* (*i.e.*, offer a reaction by selecting a clickable symbol for 'like,' 'love,' 'anger,' etc) another user's post than they are to provide a post of their own (Hootsuite, 2019). Though a *like* should be fairly concrete indicator that a post has been read, posts that users react to should only represent a fraction of total posts read, though the exact ratio is unknown. Similarly, data for Twitter users in the United States show that 80% of tweets are produced by only 10% of users, and that the average user tweets only

twice per month (Wojcik & Hughes, 2019). For these reasons, online activity including both social media and gaming in English were combined and categorized as sources of L2 input.

In sum, variables for L2 use included the proportion of English versus Mandarin used for interacting with peers; with family at home; and for written messaging with friends or family. Written messaging could be done via text messaging, chat applications (including use of messaging features of social media platforms like Facebook) or email. Though written messaging is less typical of interactional use, many participants of the present study reported sending more than 50 text messages or more per day and given that young adults typically expect a response to a text message from within a few minutes to under an hour (Forgays et al., 2014), it is more consistent with interactional L2 use as it provides a balance of both input and output. Variables reflecting sources of L2 input are given in hours of English input, while L2 use variables are presented as relative proportions of L1-Mandarin to L2-English use. Sources of each type are presented in Figure 4.1 and Figure 4.2, respectively.

4.3. Results

Analyses to address the research questions were done using *R* Statistical Software [version 3.6.0; R Core Team (2018)]. To show how the two study groups compare, descriptive statistics for each participant group, along with the results of a pairwise group comparison are presented in Table 4.1. Comparisons are shown for participants' age at the time of study, age at onset of English education, length of exposure to English, and level of parental education. Groups were matched on most of these variables, but the community-L2 group

had significantly longer exposure to English ($t(61.85) = 2.03$, $p = .046$). The results for each research question are presented below.

Table 4.1 Background Data for Participant Groups

	Community-L2	Immersion-L2	Group Comparisons
n	31	33	
age at testing	13;11 (0;6)	13;8 (0;9)	$t(57.07) = 1.3$, $p = .200$
age of onset of English education	4;0 (1;2)	4;7 (1;5)	$t(60.92) = -1.59$, $p = .117$
length of Eng. exposure (months)	118.55 (15.96)	101.09 (19.91)	$t(61.85) = 2.03$, $p = .046$
level of parental education	.82 (.28)	.85 (.32)	$t(61.59) = -0.35$, $p = .728$

Note. Summary data are presented as *mean* (*SD*). Age data are given in year and month (yy;mm) with SDs in months. Age at testing is chronological age. For immersion-L2s, age of English education and length of exposure are the age of beginning and years of immersion classes. Parental education, used here as a measure of SES, was blocked on a three-point scale from 0 to 1 with 0 for no post-secondary education, .5 for some post-secondary and 1 for an undergraduate degree or higher.

4.3.1. L2 Input and Interactional Use across L2 Contexts

The goal of the first research question was to compare the community-L2 and immersion-L2 groups for differences in their out-of-school L2 use and sources of L2 input. Altogether, groups were compared for seven variables reflecting sources of L2 experience. Four were classified as sources of L2-English input (gaming and social media engagement; reading, watching TV, movies, or video and listening to music; and participation in extracurricular activities), and three were classified as sources of L2-English interactional use (language use with peers; language use at home; and language use for written messaging). Each variable was compared using a pairwise test. With the exception of the comparison for participation in extracurricular activities in English, all pairwise tests were Welch's unequal variances *t*-test. These tests are robust enough to deal with the data types of these

variables, including the two variables scored on 5-point Likert scales (Derrick & White, 2017; Winter & Dodou, 2010). The extracurricular activities factor was compared using a Wilcoxon rank sum test because values for the immersion-L2 participants clustered around zero and were thus strongly skewed.

Reported *p*-values were Bonferroni corrected for a series of seven comparisons. Statistically significant differences were found for four of the seven factors. Significantly different sources of out-of-school L2 experience were participation in extracurricular activities in English; language use with peers; language use at home; and language use for written messaging via SMS, chat apps or email. The comparisons for reading; watching video or listening to music; and engaging with social media or video games were all non-significant. Effect sizes are given using Cohen's *d*, for which values over 0.8, 0.5, or 0.2 are respectively considered *large*, *medium*, and *small* effect sizes. For all statistically significant comparisons, effect sizes for differences were large.

Table 4.2 Statistical Comparisons across L2 Contexts

	Community-L2	Immersion-L2	Cohen's *d*	Test statistic
Eng. social media & gaming	13.86 (10.31)	8.74 (9.74)	0.51	$t(61.11) = 2.04$, $p = .321$
Eng. reading	20.52 (19.42)	16.24 (9.03)	0.24	$t(53.18) = 0.96$, $p = 1.000$
Eng. TV, video, & music	17.57 (13.28)	16.15 (14.54)	0.1	$t(61.96) = 0.41$, $p = 1.00$
Eng. extracurricular activities	6.87 (5.71)	1.09 (2.85)	1.29	$W = 929$, $p < .001$
peer lang. use	4.39 (1.09)	2.21 (0.89)	2.2	$t(58.21) = 8.72$, $p < .001$
home lang. use	2.77 (1)	1.39 (0.58)	1.71	$t(47.42) = 6.73$, $p < .001$
messaging lang. use	.66 (.25)	.43 (.17)	1.09	$t(52.7) = 4.3$, $p < .001$

Note. Summary data are presented as *mean* (*SD*). English social media & gaming, English reading, English TV, video & music, and English extracurricular activities are given in hours per week. Home language use and peer language use are on scales from 1 to 5, with 5 indicating English use always. Messaging language use is a ratio of messages sent or received in the L1 versus English with higher values reflecting more English use. Effects sizes of comparisons are given as Cohen's *d*. Groups were compared with a series of pairwise tests Bonferroni corrected for seven comparisons. Given that immersion-L2 participants clustered around zero for hours of extracurricular activities, a non-parametric Wilcoxon rank sum test was used for comparison. All other tests were Welch's unequal variances *t*-tests.

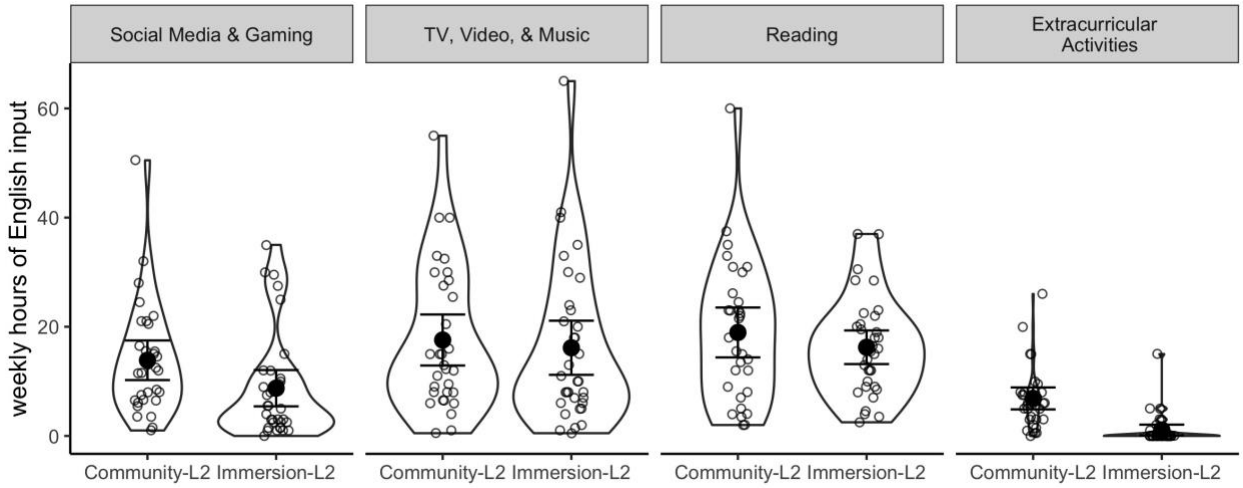

Figure 4.1. This figure shows the hours of English input across participant groups. Points represent individual participant values and are jittered to avoid overplotting. The single point on the error bar shows the group mean, and error bars show the 95% confidence interval.

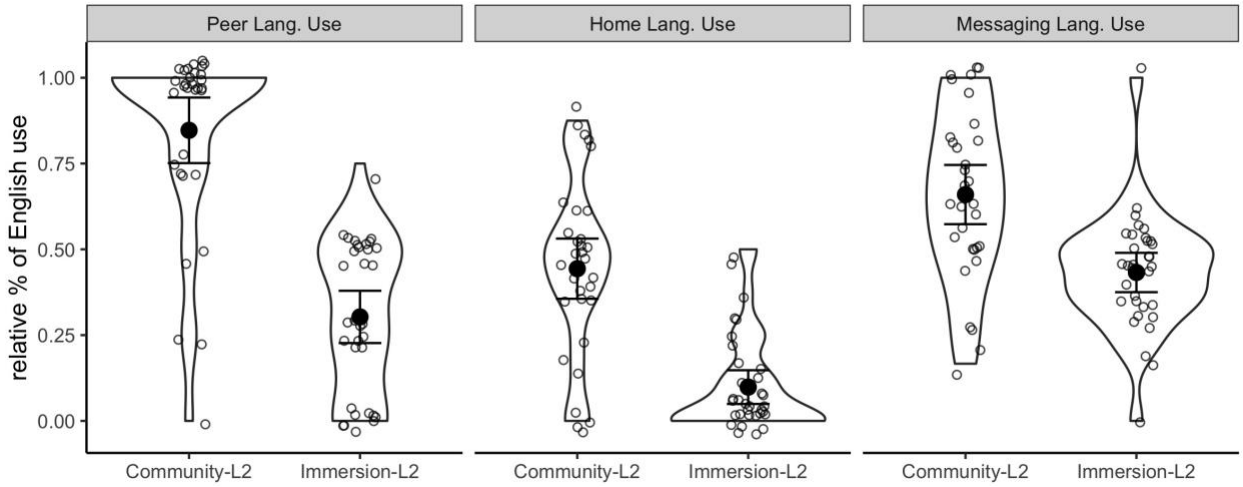

Figure 4.2. This figure shows the relative proportion of L1-Mandarin to L2-English use. Higher values indicate more relative English use. Points represent individual participant values and are jittered to avoid overplotting. The single point on the error bar shows the group mean, and error bars show the 95% confidence interval.

4.3.2. Comparison of L2 Receptive Vocabulary across L2 Contexts

The two groups' mean receptive vocabulary sizes were compared using a Welch's unequal variances *t*-test on participants' standardized PPVT score (*i.e.*, their PPVT score adjusted for age). Descriptive statistics showed that the community-L2 participants had a larger English receptive vocabulary (mean = 103.84, *SD* = 20.38, range = 51-152) than the

immersion-L2 group (mean = 78.42, *SD* = 15.65, range = 50-116). Results of the *t*-test showed this difference between the groups was statistically significant ($t(56.26) = 5.57$, $p < .001$). Cohen's *d* for this comparison was 1.4, indicating a large difference. Standardized PPVT scores indicate that the community-L2 participants were mostly within the normal range for typically developing same-aged monolinguals (standard score = 100, *SD* = 15), whereas the immersion-L2 group's mean of 78.42 fell below the one *SD* range (85 to 115) for monolinguals.

4.3.3. Predictors of L2 Receptive Vocabulary across L2 Contexts

To examine how elements of L2 environment predict L2 receptive vocabulary across learner contexts, factors were entered as predictor variables into a linear model with participants' standardized PPVT score as the dependent variable. The goal of this research question was to analyze predictors of receptive vocabulary independently across contexts, and therefore the data for participants in each context was split and analyzed separately. Analyzed variables included the seven factors from the first research question, as well as, the additional learner-external variable of length of L2 exposure. All variables were standardized as z scores in order for variable coefficients to be comparable within the models. Correlations between modelled predictors are presented in Table 4.3 and Table 4.4.

Inclusion of this number of variables based on the number of participants within each group would overfit the data. To account for this, variables were placed alternately into a model through a process of forward-fitting. Models with variables found to make a significant impact were compared to a reduced model using a likelihood ratio test to see if

the variable improved the model. For the immersion-L2 group, significant predictors of receptive vocabulary included language use for written messaging, and length of English exposure (i.e., years of immersion) showing that greater use of English during written messaging and more years of English exposure both predicted a larger English vocabulary. This model had an R^2 of .32, with an adjusted R^2 of .28 (F(2, 30) = 7.09, p = .003).

For the community-L2 group, significant predictors of receptive vocabulary included peer language use, messaging language use, English reading, and English extracurricular activities. This model had an R^2 of .75, with an adjusted R^2 of .72 ($F(4, 26) = 19.87$, $p < .001$). Greater use of English with peers and during written messaging and more hours of reading in English predicted a larger English vocabulary. However, increased hours of extracurricular activities in English had a negative impact on receptive vocabulary. With the exception of L2 use for messaging, predictors of receptive vocabulary knowledge differ for the two groups. In addition, a comparison of the adjusted R^2 for the two models indicates that predictors explained much more variance in receptive vocabulary knowledge for community-L2 learners. Model results for both participant groups are presented in Table 4.5.

Table 4.3 Correlations for Modelled Variables: Immersion Participants

	1.	2.	3.	4.	5.	6.	7.
1. Eng. social media & gaming							
2. Eng. reading	.298						
3. Eng. TV, video, & music	.535**	.270					
4. Eng. extracurricular activities	-.032	.156	.132				
5. home lang. use	.205	-.078	.200	.329			
6. peer lang. use	.190	.114	.384*	-.130	.136		
7. messaging lang. use	.214	.123	.412*	-.064	.047	.452**	
8. length of Eng. exposure	.203	.215	-.103	-.041	.010	-.087	-.354*

Note. $p < .05$ '*,' $p < .01$ '**,' $p < .001$ '***'

Table 4.4 Correlations for Modelled Variables: Community-L2 Participants

	1.	2.	3.	4.	5.	6.	7.
1. Eng. social media & gaming							
2. Eng. reading	-.049						
3. Eng. TV, video, & music	.301	-.037					
4. Eng. extracurricular activities	-.009	.614***	-.009				
5. home lang. use	-.165	.163	-.002	.201			
6. peer lang. use	-.001	.028	.196	.105	.476**		
7. messaging lang. use	.035	.126	.424*	.340	.253	.607***	
8. length of Eng. exposure	-.184	.155	-.083	.060	-.016	-.136	-.203

Note. $p < .05$ '*,' $p < .01$ '**,' $p < .001$ '***'

Table 4.5 Model Results: Predictors of Receptive Vocabulary

	estimate	std. err	*t*-value	*p*-value
Immersion participants				
(Intercept)	78.42	2.32	33.82	<.001
messaging lang. use	8.43	2.42	3.49	.002
years of immersion	5.26	2.42	2.17	.038
Community-L2 participants				
(Intercept)	103.84	1.95	53.20	<.001
Eng. Reading	11.24	2.53	4.44	<.001
peer lang. use	9.87	2.52	3.92	<.001
messaging lang. use	8.40	2.68	3.13	.004
Eng. extracurricular activities	-8.73	2.69	-3.25	.003

Note. This table presents the results for separate statistical models for each participant group. All predictors are standardized as *z* scores.

4.4. Discussion

The present study compared two groups of adolescent L2 learners for their out-of-school L2 experiences and their L2 receptive vocabulary knowledge. L2 learners in an immersion context, where the L2 is not widely spoken in the community, were compared to peers in a community-L2 context. Three research questions were investigated. The first research question examined how the immersion learners differed from the community-L2 learners in their sources of L2 input and opportunities for L2 use. Findings for this research question indicated that the groups were more likely to differ in out-of-school opportunities for L2 use than for access to L2 input. The second research question compared the L2 receptive vocabulary size of participants across the two L2 contexts finding that the community-L2 participants had larger L2 vocabularies than the immersion-L2 group. The third research question examined how learner-external differences in L2 environment,

along with length of L2 exposure, predicted participants' L2 receptive vocabulary size. For learners of both L2 learning contexts, sending and receiving more messages in English was a predictor of a larger vocabulary. However, additional predictors were different for each context. For the community-L2 group, more English use with peers, more English reading, and less participation in extracurricular activities were also significant predictors of a larger vocabulary, while the only additional significant predictor for the immersion-L2 group was length of English exposure.

4.4.1. L2 Input and Use in the Immersion and Community-L2 Contexts

Previous research for immersion learners indicates that their L2 vocabulary can lag behind monolinguals (Harley, 1992; Hermanto et al., 2012). In contrast, beyond the first few years of L2 exposure, individual community-L2 learners are increasingly likely to demonstrate L2 receptive vocabulary within the normal range on standardized tests by the end of elementary school (Paradis & Jia, 2017; Soto-Corominas et al., 2020). Because early L2 learners in both the immersion and community-L2 contexts get L2 exposure in classrooms, these groups of L2 learners are most likely to differ in their out-of-school informal L2 exposure. Thus, differences in vocabulary ability may result from differences in out-of-school L2 exposure. To identify what these differences might be, the first research question compared the two groups for sources of L2 input and opportunities for interactional L2 use.

The present study collected comparison data for seven variables, of which, four recorded hours of out-of-school L2 input and three were relative proportions of L1 and L2 use. L2 input variables were: Time spent using L2-English to access social media or playing

video games; viewing or listening to media; reading; and participating in extracurricular activities. For participation in extracurricular activities, there was a statistically significant difference between the two participant groups, but for the other L2 input variables, there were no statistically significant differences between the immersion or community-L2 groups. For variables reflecting proportion of L1 and L2 use, there were statistically significant differences for all variables. For each of these variables, Cohen's *d* indicated the differences were large. Thus, these results show that that immersion-L2 learners have a great deal of potential access to L2 input and that hours of L2 input can be comparable to community-L2 learners, but the groups differ in their respective opportunities to use the L2.

That L2 input is generally limited to the immersion classroom has been considered characteristic of the immersion context (Swain & Johnson, 1997). However, in light of increasingly widespread access to the Internet, the extent to which this would remain true was unclear. The findings of the present study indicate that the immersion-L2 learners potentially have access to amounts of out-of-school L2 input comparable to community-L2 learners for the sources studied, despite the immersion learners residing in a community where the L2 is not widely used. Though L2 input variables for reading or viewing and listening to media were not subdivided for the proportion of these activities mediated through the Internet, it is likely that Internet access has a large impact on the degree of L2 input available to participants of both groups, but especially the immersion-L2 participants.

Despite being similar to the community-L2 participants in their access to L2 input, the immersion-L2 participants continue to have limited opportunities to use the L2 in

interaction with others. This should be expected given that learners in contexts where the L2 is a non-community/foreign language will have more limited access to other speakers of the target L2. For each variable reflecting proportion of L2 use, the two groups differed significantly. The community-L2 participants were more likely to use English with peers, at home, and for messaging than were the immersion-L2 participants. Though all differences were large, comparison of effect sizes shows the largest difference for peer language use and smallest for messaging.

A richer L2 environment has been found to impact L2 vocabulary development for community-L2 learners (Paradis, 2011; Paradis et al., 2020; Paradis & Jia, 2017), and both L2 input and L2 use should contribute to this. The present study shows that the two groups differ in the richness of their respective out-of-school L2 environments. However, based on the variables studied, this difference is largely limited to opportunities for interactional L2 use rather than sources of L2 input, for which the two groups were comparable.

4.4.2. L2 Receptive Vocabulary Size across L2 Contexts

The second research question compared the receptive vocabulary scores of the two groups finding a significant difference with a large effect size indicating that the participants of the community-L2 group typically have larger L2-English vocabularies than their peers in the immersion-L2 group. Finding a difference between these groups should not be unexpected given that length of L2 exposure in school has been found to be a consistent predictor of greater L2 ability (Paradis & Jia, 2017) and the community-L2 group had 17.5 more months of exposure to English than the immersion-L2 group. In addition, the community-L2 group participants had an additional advantage over the immersion-L2 participants given the

former had studied the full day in English throughout elementary school (*i.e.*, from grades one through six), as opposed to only a half day for the immersion-L2 participants.

The community-L2 group had a mean standardized PPVT score of 103.84 (20.38). The normal range of standardized scores for this test, normed on monolinguals, is between 85 and 115 with a mean of 100. The data show that most community-L2 participants performed within this range. This aligns with previous studies of community-L2 groups with shorter lengths of L2-English exposure that show that these learners are steadily catching up with monolinguals (Farnia & Geva, 2011; Golberg et al., 2008) and potentially meet or exceed the PPVT mean standard score around four to six years of L2 exposure (Paradis & Jia, 2017).

By contrast, the immersion-L2 group had a standardized PPVT score of 78.42 (15.65), which is outside the lower bound range of the monolingual norm (*i.e.*, below 85). The immersion-L2 group's mean score aligns closely with results for other immersion learners studied by Hermanto et al. (2012). For the French version of the PPVT, the authors reported the Canadian French immersion learners in grade two and five had mean scores of 83.7 and 80.3, respectively, as compared to 78.42 for the immersion-L2s of the present study. This similarity suggests that scores of approximately 80 may be typical for immersion learners. However, research with community-L2 learners shows that the similarity between an L1 and L2 can provide an advantage for development of L2 receptive vocabulary knowledge (Blom et al., 2020; Goriot et al., 2018). In Canadian French immersion, participants are typically L1-English speakers, making the L1-L2 pairing one with far greater typological and historical connection than Mandarin and English. The degree to which the immersion context is predictive of receptive vocabulary knowledge

irrespective of L1 knowledge is unclear and would need to be investigated in future research.

4.4.3. Sources of Individual Differences in Receptive Vocabulary within L2 Contexts

The third research question examined the impact of potential predictors of individual differences on participants' L2 receptive vocabulary knowledge. Predictors included the learner-external variables associated with L2 learning context examined in the first research question, as well as the additional learner-external variable of length of L2 exposure in school.

Comparison of standardized coefficients showed that the strongest predictor of receptive vocabulary for the immersion-L2 participants was a greater proportion of English use for written messaging. This was also the only predictor shared with the model for the community-L2 group. Meta-analysis for the impact of text messaging on instructed-L2 learners ranging from elementary to graduate school shows that generally this type of language exposure is beneficial for vocabulary learning (Lin & Lin, 2019). However, this conclusion is based on quasi-experimental studies where researchers manipulate the behaviour of learners. Little previous research has looked at either immersion or community-L2 learners and how their language preferences when text messaging may impact L2 receptive vocabulary knowledge.

Results of the present study suggest that written messaging can have a beneficial impact for L2 learning groups with greater language exposure than instructed-L2 learners. However, this result could also be due to common underlying factors. For example, immersion learners who are more motivated to learn English may, as a result, have both

larger vocabulary sizes and use English for written messaging more often. For community-L2 learners, a larger relative proportion of English rather than L1 use when messaging could also be reflective of greater English dominance relative to Mandarin. Future research will be required to clarify the relationship between written messaging and development of L2 receptive vocabulary.

For the immersion-L2 learners, length of English exposure was the only other significant predictor of L2 vocabulary knowledge. For both groups length of English exposure was calculated from the time a participant entered school. The absence of other out-of-school predictors suggests that the immersion-L2 learners still depend on in-school English input for vocabulary development. This contrasts with the community-L2 learners for whom L2 receptive vocabulary was not predicted by length of English exposure. That length of English exposure was not a significant predictor of L2 receptive vocabulary knowledge is consistent with findings for other community-L2 groups whose early vocabulary growth is more rapid and begins to slow as learners accumulate more L2 experience (Farnia & Geva, 2011; Soto-Corominas et al., 2020).

That length of time in immersion and only one out-of-school L2 exposure variable were significant predictors of the immersion-L2 learners' L2 receptive vocabulary knowledge is in partial agreement with the findings of Olsson and Sylvén (2015). For Swedish L2-English CLIL learners, the authors found that out-of-school sources of L2 exposure did not benefit knowledge of academic vocabulary. These results combined with the results of the present study indicate that out-of-school L2 exposure makes a relatively smaller contribution to variations in L2 receptive vocabulary for learners in non-community-L2 contexts, but who have rich in-school experience. In contrast, findings for

instructed-L2 contexts indicate that out-of-school L2 exposure through sources such as gaming, and social media can benefit L2 knowledge (De Wilde et al., 2019; Peters, 2018). For instructed-L2 learners, small differences in additional L2 exposure may have a larger impact given that they contribute a larger relative proportion of exposure for learners who only receive a few hours a week of instruction (Sun et al., 2016). Essentially, the in-school L2 exposure of instructed-L2 learners is less rich than that of immersion (or CLIL) learners, and as a result, even limited out-of-school L2 exposure provides a benefit.

For the community-L2 participants, the strongest positive predictor of L2 receptive vocabulary was more weekly hours of reading in English, followed by a greater proportion of English use with friends, and then a greater proportion of English for messaging. This model also indicated that more hours of extracurricular activities in English had a negative relationship with L2-English receptive vocabulary.

For monolinguals, the reciprocal benefits between reading and vocabulary knowledge are well known (Adlof & Perfetti, 2013; Mol & Bus, 2011) and evidence indicates that L2 learners can also acquire new vocabulary incidentally through reading, as well (Webb, 2019). However, it is unclear why this source of L2 input was only predictive for the community-L2 participants and not the immersion-L2 learners despite both participant groups reporting similar amounts of out-of-school reading.

In addition to reading as a source of L2 input, two sources of interactional L2 use had a positive association with L2 receptive vocabulary knowledge for the community-L2s. The model showed that an increased proportion of L2 use with peers and as discussed above, L2 use when written messaging were predictive of larger L2 vocabulary sizes. L2 use with peers has previously been considered to contribute to a richer L2 environment

and has been found to benefit vocabulary development in community-L2 learners (Paradis & Jia, 2017; Soto-Corominas et al., 2020). Similarly, increased use of L2-English for written messaging can also be considered to contribute to a richer L2 environment as it increases both L2 input and opportunities for L2 output.

The negative impact of extracurricular activities on receptive vocabulary was unexpected as it was anticipated such activities would contribute to a richer L2 environment, and thus, if the variable predicted receptive vocabulary, the impact would be a positive one. Examination of the data indicated that several participants reported more than 10 hours a week of extracurricular activities, and it appears these participants tended to have average or below average vocabulary scores. The present study did not collect information about the extracurricular activities of participants, and thus, a possible explanation for this finding may lie in this missing data. One might expect that two hours a week participation in a debate club should offer much richer language exposure than 15 hours a week of competitive swim practice, and thus, in future, collecting information about the types of extracurricular activities rather than the number of hours may provide a better reflection of out-of-school L2 input.

4.4.4. Implications and Future Directions

Overall, the findings of the present study show that the L2 receptive vocabulary ability of the immersion-L2 learners lags behind the community-L2 group, falling outside the one-*SD* standardized score range for the PPVT. Further, the findings suggest that this is likely due, at least in part, to a less rich out-of-school L2 environment for the immersion-L2 learners. In examining predictors of L2 receptive vocabulary ability, the model for the community-

L2s had an adjusted R^2 of .72 and included multiple sources of out-of-school L2 experience. In contrast, the model for the immersion-L2 participants was a much lower adjusted R^2 of .28 and only included one source of out-of-school L2 use (*i.e.*, messaging language use) along with length of English exposure. Therefore, it appears that the community-L2 learners' out-of-school L2 environments had a greater impact than those of the immersion-L2s, who appear to depend more on in-school exposure. Thus, the findings of this study support conclusions of other studies of community-L2 participants that found richness of the L2 environment is important for developing L2 vocabulary (Carhill et al., 2008; Paradis, 2011; Paradis et al., 2020; Paradis & Jia, 2017). Though the present study indicates that the community-L2 participants benefited more from their out-of-school L2 exposure, this does not mean that this type of L2 exposure is not potentially important for immersion learners. If the immersion-L2 participants had access to richer out-of-school L2 experiences, then it is most likely that they would benefit from this L2 experience similarly to community-L2 participants.

Contrary to expectation, the present study found that these immersion learners had extensive access to L2 input beyond the classroom. Research has shown that L2 experience in the immersion classroom is predominantly in the form of input with far fewer opportunities to use the L2 in interaction where learners also get to produce the L2 (Allen et al., 1990; Swain, 1985, 1988) and it appears that the out-of-school L2 environment of these immersion is also dominated by opportunities for input with fewer opportunities for interactional use. Swain (1985) argued that opportunities for interactional use are key to L2 development, and as the biggest differences between the two contexts were in

opportunities to use the L2 it appears that these are key to the community-L2 participants' better L2 receptive vocabulary ability found in the present study.

Finally, as a future direction, it would be good to consider sociolinguistic context in combination with vocabulary knowledge. The participants of the present study showed a lag in vocabulary knowledge. However, it is possible that this lag would be smaller or eliminated if test words for receptive vocabulary knowledge were classified as either academic or non-academic words (*i.e.*, classified by the sociolinguistic contexts in which each word is common). In a study of bilingual children who speak one language at school but two at home, Bialystok et al. (2010) found that the bilinguals' vocabulary for words expected to be more common at home, where their input was divided between two languages, was smaller than their monolingual peers. However, for academic vocabulary the gap was smaller. Similarly, it might be expected that the gap between the immersion-L2 and community-L2 groups would be smaller if only academic vocabulary items were considered given that both groups should have rich L2 exposure in school. Future research on this topic can clarify where immersion learners' gaps in receptive vocabulary knowledge lie.

4.4.5. Limitations

This study has several limitations that should be addressed in future research. The aim of the present study was to examine a range of potential sources of out-of-school L2 experience. However, there are several ways that the conceptualization of variables may be overly broad. First, classification of sources of L2 experience as either sources of L2 input or opportunities for L2 use may erase important differences in the L2 experiences that

individual learners have. For example, while cited research indicates that, on balance, social media and gaming are sources of L2 input for the typical user, this does not reflect the range of L2 experiences for learners. It is possible that some L2 learners produce a lot of L2 output through these sources of L2 experience. However, this range of potential experience was not examined here. Similarly, the aggregation messaging language use eliminates potentially important variations pertaining to the nature of messaging (*i.e.*, how long are messages and how long between message and response) and the interlocutors involved. The present study did not collect this data and it is likely that messaging with peers or family have differing impacts. Finally, it should be noted that the variable of peer language use is not directly comparable across contexts. For immersion learners, the most easily available speakers of the L2 are classmates who typically all share an L1, while community-L2 learners have ready access to L1 speakers of the target language. As a result, the peer language use variable reflects interaction among L2 learners in the immersion context, but a greater proportion of L2 to L1 speaker interaction in the community-L2 context. It should be expected that L1-speaking peers will be more likely to introduce new words into conversations with L2 learners and are thus a greater benefit to L2 receptive vocabulary.

4.5. Conclusions

The aim of the present study was to investigate differences between the out-of-school L2 environments of the immersion and community-L2 contexts and how these might impact L2 receptive vocabulary knowledge. The two groups differed in the richness of their out-of-school L2 environments with a key difference being that the immersion learners reported

fewer opportunities for interactional L2 use as opposed to amounts of L2 input for which they were more similar to the community-L2 participants. Findings also showed that the community-L2 participants had larger L2 vocabularies than the immersion participants, with the former performing on par with age-standardized monolingual norms and the latter falling below the normal monolingual range. For the community-L2 participants, out-of-school L2 experiences were highly predictive of L2 receptive vocabulary ability, while only one source of out-of-school L2 experience was predictive of the immersion learners' L2 ability. The findings of the present study support the conclusion that opportunities for interactional L2 use are important contributors to a rich out-of-school L2 environment, and that a richer L2 environment can benefit the L2 receptive vocabulary ability of adolescent L2 learners. However, it also appears that, in general, this type of L2 environment is only available to learners in community-L2 contexts.

5 General Discussion and Conclusion

5.1. Summary of Results

This dissertation investigated the L2-English ability of a group of L1-Mandarin immersion learners, examining the following three overarching research questions.

1. How do the receptive L2 abilities of L1-Mandarin adolescent immersion learners compare to the English abilities of L1-Mandarin community-L2 learners and/or L1-English speakers?
2. Across the two L2 learning contexts, how do out-of-school L2 input and experience differ between the immersion and community-L2 contexts, and how are these between-context differences associated with differences in L2 ability?
3. Within the immersion context, how do differences between individual immersion learners in their out-of-school L2 input and experiences impact individual differences in L2 ability?

These are addressed by research presented in Chapters Two, Three, and Four. In this chapter, the results of these studies are summarized. Then, the implications of these results are discussed with regard to what these findings can contribute to our understanding of L2 acquisition in immersion. Finally, possible future directions for research with immersion learners are proposed.

5.1.1. Study One

The goal of the first study reported in Chapter Two was to compare the L2-English receptive morphosyntactic ability of immersion learners to both community-L2 learners

and L1 speakers of the target language. The study examined whether there were differences in the ability with English grammatical morphemes between the studied participant groups, and, for the L2 groups, whether either learner-internal or learner-external factors impacted this ability. Previous research with immersion learners has shown that their L2 production is often non-target (*i.e.*, they make errors) when compared to an L1 standard of morphosyntactic proficiency. However, the degree to which these differences in L2 ability can be ascribed to the immersion context or to L2 acquisition more generally is unclear given that immersion learners have never been compared directly to another group of high-proficiency early L2 learners. To compare the receptive morphosyntactic ability of the three study groups, a grammaticality judgment task was used. Stimuli examined a range of grammatical morphemes that varied in level of difficulty. In addition, learner factors which have been found to be predictive of L2 morphosyntactic ability at earlier stages of acquisition for community-L2 learners (*e.g.*, Jia & Fuse, 2007; Paradis, 2011; Paradis, Soto-Corominas, Chen, & Gottardo, 2020) were examined in order to identify factors that may be associated with L2 development in each of the two L2 learning contexts.

The results of the judgment task showed differences in accuracy between all three groups. Both L2 groups had lower accuracy than the L1-English group for stimuli classified as either easier (Easy) or more difficult (Difficult). For the participants of the community-L2 group this indicates that even after more than eight years of L2-English exposure in their schools and community, their receptive morphosyntactic ability is not identical to their L1-speaking peers. The L1-English group had very high accuracy for the Easy stimuli. Despite being less accurate than the L1 participants, the community-L2 participants were

significantly more accurate than the immersion group for stimuli classified as Easy, with the gap in accuracy between the community-L2 participants and the L1 speakers being much smaller than that between the community-L2 and immersion participants. The immersion participants' accuracy was low overall, while the accuracy of community-L2 participants was only particularly low for the Difficult stimuli, for which the two L2 groups were not significantly different from each other. Nevertheless, it was not only the L2 groups that found the Difficult stimuli challenging as the L1-English group also had lower accuracy for these stimuli.

Examination of factors that predicted individual differences in judgment-task accuracy showed that both learner-internal and learner-external factors had an impact. However, these factors predicted very little of the within-group variance in L2 ability, indicating that the examined factors were not powerful explanations of individual differences for these later-stage high-proficiency L2 learners. Further, for neither group was longer L2 exposure a predictor of morphosyntactic ability, suggesting that L2 morphosyntactic ability has slowed or plateaued by this point in development.

5.1.2. Study Two

Findings from the judgment task in Study One indicated that early L2 learners in immersion differ in their morphosyntactic ability even after many years of L2 exposure. However, how these differences in morphosyntactic ability may or may not impact L2 comprehension remains unclear. The study presented in Chapter Three assessed comprehension of English plural-singular marking, a semantically transparent and frequently used morphological distinction. This construction differs from Mandarin and

has been noted as an area of difficulty in L2 acquisition for L1-speakers of classifier languages like Mandarin (Jiang, 2007; Luk & Shirai, 2009). Comprehension was assessed online, with a visual-world eye-tracking task, and offline, with a picture-decision task. Immersion participants were compared to a group of L1-English speakers.

For the picture-decision task, the accuracy of the immersion participants was lower at 69% compared to the L1-English controls who were 92% accurate. Receptive vocabulary was also examined as a predictor of accuracy. While a larger L2 vocabulary size was a significant predictor of higher response accuracy for the immersion learners, their responses were still significantly less accurate after taking vocabulary size into account. In sum, the immersion participants were less able to use the bound plural '*-s*' morphological marker or its absence to disambiguate between pictures of either single or multiple items.

For the eye-tracking task, the immersion group's looks to the target images had both similarities and differences with the L1-English control group. For plural stimulus trials (*i.e.*, when a plural noun in the auditory stimulus was paired with a target image of multiple corresponding items), the immersion participants were as likely as the L1-English speakers to be looking at the target image. However, for singular stimulus trials, the immersion participants were less likely to be looking at the target image. Despite the immersion participants being more likely to parallel the L1-speakers for the plural trials, this did not translate into an accuracy advantage on the picture-decision task for plural stimulus items and was likely the result of an early bias exhibited by both groups to look at the pictures with multiple items. However, it remains possible that a greater proportion of looks to targets on plural trials reflects superior online comprehension that does not carry over to the explicit picture-decision task.

Together, findings from the two tasks indicate that these immersion learners do not interpret English plural-singular marking identically to L1-English speakers, even after many years of L2 exposure. However, it remains unclear why these differences should exist for a semantically transparent, highly frequent morphosyntactic construction, particularly given that plural-singular marking is early acquired in L1-English acquisition. This study shows that comprehension tasks targeting specific morphosyntactic constructions can reveal differences in comprehension that may not be apparent on tasks of holistic L2 comprehension.

5.1.3. Study Three

The final study, reported in Chapter Four, examined how the out-of-school L2 environments of immersion and community-L2 learners differ, and what implications these differences potentially have for the development of L2 ability. In order to collect comparison data for participants' out-of-school L2 experiences, a language background questionnaire was used. In addition, the groups were compared for L2-English receptive vocabulary, assessed using the PPVT. Finally, these differences in out-of-school L2 experiences and length of L2 exposure were modelled to see if they predicted receptive vocabulary development. Results suggest that the two participant groups differ in the richness of their out-of-school L2 environment and that this has implications for their L2 receptive vocabulary knowledge.

Sources of out-of-school L2 experience were categorized depending on whether they were mainly sources of L2 input, additional to what the learners of each group would receive in school, or whether they were more balanced in opportunities for both L2 input

and L2 output. This was key given that research shows immersion learners' in-school L2 experience is predominantly L2 input with few opportunities to produce L2 output (Allen, Swain, Harley, & Cummins, 1990; Swain, 1988). These categories were labelled as L2 input and interactional L2 use, respectively.

Though it has been considered typical of immersion contexts that learners' L2 experiences are limited to the classroom (Swain & Johnson, 1997), findings for the study presented in Chapter Four indicate that the immersion participants have access that is comparable to the community-L2 learners for sources of L2 input from reading, social media, and traditional media such as television, video, and music. However, large differences between the groups were found in their access to sources of L2 interactional use with peers, at home, or via text messaging. Though the findings of the study indicated that the immersion participants had richer exposure to out-of-school L2 input than previously recognized, the differences in opportunities for interactional L2 use indicate that overall, the L2 environment for the immersion participants was less rich than that of the community-L2 participants. Like findings for in-class L2 exposure, the out-of-school L2 experience of immersion learners is also predominantly input as opposed to opportunities for interactional L2 use.

Comparison of the two groups for L2 receptive vocabulary indicated a large difference between the groups. The community-L2 participants had a group mean score just above the age-standardized receptive vocabulary score for monolingual English speakers with most scores within the normal range. In contrast, the immersion group had a mean score outside the normal monolingual range. These findings indicate that the

community-L2 participants have caught up to the monolingual norm, but that the immersion learners still lag.

For both groups, L2 receptive vocabulary size was predicted by richer L2 input and richer interactional L2 use, though the models included mostly different predictors across groups. For the community-L2 participants, using English for reading, interacting with peers, and messaging all predicted larger vocabulary sizes. For the immersion participants, messaging in English and longer exposure to English both predicted a larger receptive vocabulary. In contrast to the models predicting morphosyntactic accuracy reported in Chapter Two, much more of the within-group variance was explained by the models for receptive vocabulary.

5.2. General Discussion

This dissertation investigated the L2-English ability of L1-Mandarin immersion learners, by examining how their L2 abilities compared to both community-L2, and L1-English speakers, and how L2 input and experience impact these L2 abilities. Collectively, the research presented in Chapters Two, Three, and Four addressed the dissertation research questions by examining 1) how the L2-English abilities of the adolescent immersion participants are similar or different from the English abilities of community-L2 learners and/or L1-English speakers; 2) how the out-of-school language environments of these immersion learners compare to their community-L2 learning peers; and 3) ways in which these differences in L2 learning environment may be associated with identified differences in L2 abilities between individual immersion learners.

Findings noted in the chapter summaries above indicate that in comparison to the other groups of English speakers, these adolescent immersion learners have non-target ability with L2 grammatical morphemes, and for L2 receptive vocabulary, they fall short of the mean standardized score for monolinguals and have smaller L2 receptive vocabulary sizes than community-L2 peers. This aligns with the conclusions of previous studies of immersion learners showing that when compared to a target language standard based on L1 speakers, immersion learners often have non-target use of morphosyntactic constructions (Fortune & Tedick, 2015; Kowal & Swain, 1997; Wesche, 1993) and smaller vocabularies (Harley, 1992; Hermanto, Moreno, & Bialystok, 2012) even after many years of exposure. For the second research question, immersion learners had greater access to L2 input than might be predicted based on prior research, but still used the L2 in interaction much less frequently than community-L2 learners. Finally, differences in L2 input and experience predicted individual differences in the immersion learners' L2 abilities, but the factors that impacted as well as their degree of impact differed in each study. In the following sections, reasons for these differences in L2 ability are discussed in more detail and are also interpreted in relation to findings for the other dissertation participant groups and prior research for both immersion and community-L2 learners. In addition, potential explanations for how non-target ability emerges for these immersion learners, and why these non-target abilities may persist are discussed. Findings are interpreted through research and theory within the usage-based paradigm.

5.2.1. Predictors of Differences in L2 Ability

In addition to examining how L2 ability compares between the immersion and community-L2 contexts, this dissertation also investigated how variability between learners in L2 input and experience impacted L2 ability. These learner factors were analyzed to identify elements of an L2 environment that potentially drive the development of L2 abilities. In order to focus on L2 context, the L2 participant groups were matched for L1, age at testing, age of beginning English education, and parental education. Thus, the expectation was that differences in L2 ability should be the result of differences in the L2 environment provided by each L2 learning context. In addition to these learner-external factors, in Chapter Two, some learner-internal factors were also examined for their potential impact on L2 morphosyntactic ability.

For L2 receptive vocabulary, only learner-external factors associated with L2 learning context were examined. For the community-L2 participants, larger L2 vocabulary sizes were predicted by greater amounts of L2 input through reading, as well as more L2 use with peers and for messaging were associated with larger L2 vocabulary sizes. The model for the community-L2 group had an adjusted R^2 of .72 indicating that these factors explained a large proportion of variance in L2 receptive vocabulary knowledge among the participants of this group. For the immersion learners, only two variables predicted a larger vocabulary size. Of these, the variable most predictive of a larger L2 vocabulary was more L2 use for messaging. This was also the only significant variable that was a source of out-of-school L2 exposure, as the other variable, length of L2 exposure, was defined as the length of time the participant had been in English immersion. In addition, the model for the immersion learners had an adjusted R^2 of only .28, indicating that the model explained less

variance for the immersion learners. These results indicate that a richer L2 environment can benefit learners in either context, but that out-of-school L2 experiences are more likely to benefit community-L2 learners, while immersion learners still rely on in-school exposure.

In contrast to the models for receptive vocabulary, very little variance between individual participants was explained by the models for morphosyntactic ability reported in Chapter Two. For the model predicting immersion learners' accuracy on the judgment task, the adjusted R^2 was .02 as opposed to .28 for receptive vocabulary. For the community-L2 group's morphosyntactic accuracy, the adjusted R^2 for the model was .04 (with the stimulus difficulty variable removed), in contrast to .72 for receptive vocabulary model. Given that the respective R^2 values across domains are based on different types of models that analyze the results of different test instruments (*i.e.*, either the PPVT or the judgment task), and that somewhat different sets of predictors were included in each study, direct comparison of these numbers should be done with caution. Nevertheless, the differences in estimated explained variance are very large. Thus, it is reasonable to conclude, in general terms, that much more of the within-group variance in receptive vocabulary was explained by learner factors, whereas little variation in morphosyntactic ability was explained.

These findings indicate that for participants of both of these high-proficiency adolescent L2 groups, differences between individual learners, particularly in their L2 environment, have a larger impact on L2 receptive vocabulary than on L2 morphosyntactic ability. At first glance, this could be interpreted as inconsistent with a usage-based approach given that these approaches propose common underlying representations for all

types of linguistic constructions (Dąbrowska, 2018; Kidd, Donnelly, & Christiansen, 2018), and thus, both vocabulary and morphosyntax should be similarly impacted by learner factors. As a result, the findings presented here could be interpreted as support for theoretical approaches to language that posit a split between lexical knowledge and rule-based morphosyntactic knowledge (*e.g.*, Pinker, 1999; Ullman et al., 2005). However, it was argued in Chapter Two that development of L2 morphosyntactic ability for the studied constructions is likely to have plateaued for learners in both L2 groups. When examining factors that predict L2 development, there needs to be L2 development still taking place to predict. Thus, the difference in predictive power between the two types of models may not be the result of differences in linguistic domains, but the respective developmental stages of the constructions studied for each L2 domain.

5.2.2. Continued L2 Development in Immersion

The results of the present dissertation indicate that after many years of L2 exposure both the morphosyntactic and receptive vocabulary abilities of the studied immersion learners can remain different, not only from L1 speakers, but community-L2 peers, as well. Overall, results indicate that the L2-English abilities of the community-L2 learners are more convergent with L1 speakers than are the L2 abilities of the immersion learners. However, one background variable for which the two L2 groups were not matched was in their length of English exposure. The immersion participants had approximately eight and a half years of exposure whereas the community-L2 participants had about ten years. When compared directly, there was a significant difference between the groups ($t(61.85) = 2.03$, $p = .046$ based on the participants of Chapter Four). In addition, results presented in Chapter Four

indicate that the community-L2 participants also use their L2 more often within this length of time. Both factors mean that the community-L2 participants have a sizable advantage in their total quantity of L2 exposure. Therefore, it is reasonable to ask whether the L2 development of the two L2 groups is similar, with the immersion learners on a more delayed timeline reflecting their lower quantity of total L2 exposure, or whether the evidence presented here indicates qualitative differences in L2 development between the two groups. The answer to this largely depends on the domain of language examined.

5.2.2.1. L2 Receptive Vocabulary Ability.

For receptive vocabulary, the immersion participants continued to lag well behind the community-L2 group, as well as the monolingual-normed standard score for the PPVT. The two groups also differed from one another in that length of L2 exposure was a significant predictor of L2 receptive vocabulary for the immersion learners but not for the community-L2 learners. However, L2 receptive vocabulary knowledge was gauged using an age-standardized PPVT score, meaning that participants' vocabulary scores were not a reflection of an absolute number of words known, but rather a measure of how the participants compared to same-aged monolingual English speakers. Language exposure over time is a powerful predictor of vocabulary development throughout an individual's lifetime even for an L1 (Keuleers, Stevens, Mandera, & Brysbaert, 2015), and this is reflected in PPVT standard scores which increase with age. Thus, for monolingual English speakers (or community-L2 participants) to match the standard score, vocabulary development has to continue year after year.

As a result, measurement against the PPVT standard score indicates that the L2 receptive vocabulary knowledge of participants in both L2 groups is still developing. The additional impact of more years of L2 exposure for the immersion participants reflects that with each year of immersion, they are getting closer to the standardized score for monolinguals (*i.e.*, they are catching up). In this light, the lack of impact for the community-L2 participants makes sense given that, on average, their receptive vocabulary scores fall within the normal range for English monolinguals, indicating that they have already caught up. No additional impact of years of L2 exposure implies that the community-L2 participants, as a group, match the year-over-year receptive vocabulary development of their L1-English-speaking peers, neither falling behind nor pulling ahead. Thus, the evidence gathered and presented in this dissertation indicates that it is reasonable to anticipate that immersion learners with longer L2 exposure and a richer L2 environment could potentially close the gap in receptive vocabulary ability, though it is not clear that this would happen in practice (*i.e.*, before immersion learners' programs end at high-school graduation). As a result, it can be concluded that the difference in quantity of L2 exposure is an important explanation for the lag in the immersion learners' L2 receptive vocabulary knowledge, and that the receptive vocabulary development of both L2 groups is ongoing.

5.2.2.2. L2 Morphosyntactic Ability.

Findings for L2 morphosyntactic ability had both similarities and differences with findings for L2 receptive vocabulary. On the judgment task, the immersion participants had lower accuracy than the community-L2 participants for Easy stimuli. However, there was no significant difference between these L2 groups for the Difficult stimuli. While the

community-L2 participants were still significantly below the L1-English control group for Easy stimuli, their accuracy was nonetheless much closer to the control group than they were to the immersion group. This indicates that the community-L2 participants had already developed a considerable amount of English morphosyntactic ability, but that the immersion participants, if considered to be on a journey toward L1-like ability, still had a long way to go.

The Easy stimuli for the judgment task probed ability to detect errors with articles, past tense, and plural marking. In L1 acquisition, these grammatical morphemes begin to be mastered in preschool, with articles reaching appropriate use around four years of age (Emslie & Stevenson, 1981; Maratsos, 1974), and with both past-tense and plural showing consistent productive use around the same time, but with errors of omission lasting to around seven years of age (Berko, 1958; Graves & Koziol, 1971; Marchman, 1997; Matthews & Theakston, 2006). Though these grammatical morphemes are mastered fairly early on in L1 acquisition, the immersion learners still appeared to find the Easy stimuli challenging. At the same time, the community-L2s were largely accurate with these items despite still being significantly different from the L1-controls.

Like for L2 receptive vocabulary, it could be that the difference between the immersion and the community-L2 groups on the judgment task is the result of the community-L2 group's additional quantity of L2 exposure. However, there are three main reasons that indicate a continued developmental trajectory toward L1-like ability is unlikely for the immersion learners. First, longer exposure to English did not predict higher accuracy on the judgment task for either L2 group. For the immersion participants, the one standard deviation range for length of English exposure was a window of over two years,

meaning that, there was no evidence the immersion participants' L2 morphosyntactic ability was becoming more target-like over this span of time.

The second reason suggesting immersion learners are not continuing toward L1-like ability is that longitudinal research for community-L2 learners with a Chinese or other isolating L1 indicates that development of grammatical morphology progresses more rapidly in the first few years and then begins to slow (Blom, Paradis, & Sorenson Duncan, 2012; Jia & Fuse, 2007; Paradis, Tulpar, & Arppe, 2016), with other research indicating that this pattern is likely shared by child L2 learners across L1s (Soto-Corominas, Paradis, Rusk, Marinova-Todd, & Zhang, 2020). This suggests that the substantial difference in morphosyntactic ability between the two L2 groups is likely to have emerged in the first few years of L2 exposure rather than in the one and half year gap between the groups from eight and a half to ten years of L2 exposure. Thus, it should be expected that, for both L2 groups, much of their L2 learning for these grammatical morphemes has already taken place prior to reaching middle school.

Finally, previous research has found that immersion learners' morphosyntactic abilities may plateau in a non-L1-like state (Fortune & Tedick, 2015; Hart, Lapkin, & Swain, 1991), and that this is more likely for constructions that contrast from L1 to L2 (*e.g.*, incorrect use of grammatical gender in L2-French immersion by L1-English speakers; Warden, 1997). In contrast to L2 receptive vocabulary, the evidence gathered here indicates that the morphosyntactic ability of the immersion learners appears to have plateaued at a stage different from either the ability of L1 speakers or that of their community-L2 peers.

5.2.3. Emergence of Non-target L2 Morphosyntax in Immersion

It is argued above that the immersion learners' ability with the L2 morphosyntactic constructions examined in this dissertation is no longer converging on an ability that parallels L1-English speakers. This section and the section following include discussions of potential explanations for why the immersion learners' L2 abilities ultimately differ from both L1-speaking and community-L2 peers. This section discusses how the immersion learners' non-target L2 abilities with English grammatical morphology arise. These factors should apply to any L1-Mandarin English learner, and therefore can also serve as explanations for early differences in L2 ability for the community-L2 learners, as well. However, the L2 abilities of immersion and community-L2 learners eventually differ. The following section includes a discussion of why the non-target L2 abilities that emerge for these groups of L2-English learners are more likely to persist for the immersion learners.

For immersion learners, Harley (1993) noted that even highly frequent L2 constructions could be vulnerable to non-target use if they lacked perceptual salience, did not carry a heavy communicative load, or differed from the L1. These risk factors for non-target acquisition are associated with either prior linguistic knowledge or perceptual mechanisms and are therefore compatible with usage-based approaches which see linguistic knowledge as an emergent result of language experience and domain-general cognitive mechanisms (Beckner et al., 2009). Though L2 acquisition involves the combination of linguistic experience and cognitive mechanisms, in actuality, these are inseparable as they exert reciprocal influences on each other (Ellis, 2006). In both L1 and L2 learning, language experience shapes cognitive learning mechanisms, and these mechanisms shape what is perceived and attended to from subsequent language

experience (Beckner et al., 2009; Ellis, 2008). For L2 learners, this means they begin their new language with cognitive mechanisms optimized for their L1 and entrenched implicit expectations about how language functions based on their L1 language experience, both of which shape learning of the L2 (Wulff & Ellis, 2018).

For morphosyntactic ability, the present dissertation examined L2 receptive ability with English grammatical morphemes, including articles, past tense, plural-singular, and verb agreement forms for BE, DO, and for third person. None of these constructions has a close correlate in Mandarin (see Chapter 1, section 4). Mandarin does not have bound inflectional morphemes, and nor are articles used to mark definiteness. Further, neither obligatory number, nor verb agreement are marked. As a result, L1-Mandarin English learners are unlikely to attend to these grammatical morphemes, and nor will they anticipate their functions as they are absent from Mandarin. Thus, there is little directly transferable L1-Mandarin knowledge to benefit learning these grammatical morphemes in L2 English. In addition to L1-tuned perception of the L2, the difficulty of acquiring grammatical morphemes is compounded as these constructions often have low phonological salience (Ellis, 2008). Low phonological salience has been found to negatively impact ease of grammatical morpheme acquisition for L2 acquisition in general, regardless of L1 (Goldschneider & DeKeyser, 2005).

How prior learning and construction salience impact the early stages of adult L2 learning was investigated in a series of studies by Ellis and colleagues (Cintrón-Valentín & Ellis, 2015; Ellis et al., 2014; Ellis & Sagarra, 2010, 2011; Sagarra & Ellis, 2013). Prior learning was investigated as the impact of pretraining in the L2 (short-term prior learning) and/or properties of the participants' L1 (long-term prior learning). These studies

examined either L2 Spanish, or a miniature language based on Latin, investigating whether participants showed a preference for determining temporal reference using either adverbs, such as the translation equivalents of *yesterday*, *today* or *tomorrow*, or a preference for verb forms marked for tense. The studies showed that both properties of the prior-learned L1 and pretraining in the L2 play a role in determining the L2 cues to which learners attend. Specifically, they showed that attention to either an adverb or verb form when both were present in a stimulus was determined by three factors. The first was whether a participant's L1 had rich verbal morphology, poor verbal morphology, or no verbal morphology. Essentially, L1 speakers of Spanish and Russian were more likely to attend to verb morphology than English L1s given that both Spanish and Russian have richer morphology than English. Speakers of all three of these languages were more likely to attend to verb morphology than were L1 speakers of Chinese, which lacks bound verb morphology entirely.

The second factor determining which temporal reference cue a participant attended to was whether the participant had received pretraining with either the relevant adverbs or verb forms. This impacted participants in all L1 groups showing that pretraining could, to an extent, mitigate L1 knowledge. This means that an L1 speaker of Chinese who received pretraining on verb morphology would be more likely to attend to these constructions than a Chinese speaker who had not, though both would still be more likely to attend to adverbs than speakers of other L1s.

The final factor determining the cue participants used for temporal reference was that all participants had a tendency to attend to the more phonologically salient adverbs, particularly, as the complexity of the stimuli increased. The tendency to prefer lexical

markers of tense to morphosyntactic markers is typical of adult L2 learners regardless of L1 (Shirai, 2009), and is not limited only to tense as a preference for lexical over grammatical markers appears to be a general principle of L2 learning (Klein & Perdue, 1992).

These studies indicate that in the earlier stages of L2 learning, L1 knowledge has an important impact on processing the new language (Sagarra & Ellis, 2013), meaning that subsequent L2 experience is parsed through perceptual mechanism tuned by L1 experience. As a result, and consistent with the studies discussed above, grammatical morphemes are unlikely to be preferred initial cues for either the L1-Mandarin immersion or the community-L2 learners studied in this dissertation.

In the present dissertation, the most closely examined L2 grammatical morpheme was the immersion learners' comprehension of English plural-singular marking. Because there is no directly transferable construction in Mandarin, the immersion learners need to inductively learn how English plural-singular marking with '*-s*' works from their L2 experiences within the immersion context. For L2 comprehension of plural-singular marking, it should be expected that the immersion participants are biased toward attending to lexical indicators of quantification both because these have closer equivalent constructions across the two languages, and because they are more phonologically salient. Specifically, Mandarin speakers should be more likely to attend to English lexical markers of number such as cardinal numbers, or the determiners *this/that* and *these/those*. This potential initial reliance on number information supplied by preceding quantifiers can also attenuate subsequent learning of additional cues. This phenomenon is referred to as *blocking* where only a prior-learned cue is associated with a particular outcome and thus,

only the initially learned cue is reinforced through subsequent exposure (Kruschke & Blair, 2000). As a result, not only is an L1-Mandarin speaker likely to depend on preceding quantifiers at initial stages of learning L2-English, but this initial attendance to these markers could actually slow or prevent learning the redundant '*-s*,' particularly given the latter is less phonologically salient.

For immersion learners, there should be nothing inherent to the immersion context that prevents immersion learners from attending to noun quantification. L1-English learners in Spanish immersion have been shown to use number-marked determiners to identify referents on an online task before completing their second year of immersion (Lew-Williams, 2017). This indicates that when the construction is phonologically salient and consistent with the L1, immersion learners can demonstrate rapid acquisition. It is expected that English grammatical morphemes, and plural-singular marking specifically, should be frequently modelled in the L2 input immersion teachers provide for their students. However, in line with Harley's (1993) risk factors, when grammatical morphemes are less phonologically salient, have their meaning overshadowed by other semantic cues, or are not present in the L1, they may be subject to non-target acquisition. Overall, the research discussed in this section serves to illustrate how L1-Mandarin immersion learners may not be able to attend to English grammatical morphemes in a way that facilitates acquiring ability to use them as L1 speakers do.

The research discussed in this section helps explain why early L1-Mandarin immersion and community-L2 learners' ability with L2-English grammatical morphemes can be non-target at early stages. These explanations are rooted in perceptual mechanisms that are either influenced by L1 knowledge, or, like the impact of low phonological salience,

are expected to be common to all L2 learners regardless of L1 background or learning context. Given that the participants of both L2 groups are L1-Mandarin speakers who began L2-English at similar ages, these explanations do not account for why the immersion participants' ability appears to have plateaued in a state different from the community-L2 learners. The following section discusses how the L2 experiences of learners in the immersion context can be implicated in the persistence of non-target L2 ability.

5.2.4. The Persistence of Non-target L2 Morphosyntax in Immersion

Findings of this dissertation indicate that the community-L2 participants have more L1-like ability with grammatical morphemes than the immersion participants. This suggests that over their years of L2-English exposure, the community-L2 learners have more successfully overcome the potential impediments to L2 development described in the previous section. Absent other factors, entrenchment of L1 knowledge should not be expected to be a permanent obstacle to L2 acquisition as it can be overcome by both older and younger L2 learners (MacWhinney, 2017). Thus, the default expectation for L2 learners, especially those who begin young, should be that their L2 abilities progressively converge with L1 speakers over time as their L2 exposure increases. However, evidence gathered for this dissertation indicates additional L2 exposure does not appear to be bringing immersion learners to a state of target language ability that parallels L1 speakers while the community-L2 learners demonstrate L2 abilities that are much more similar to their L1-English peers. Participants of the two L2 groups were matched for L1, age of beginning English, age at testing, and parental education. Given their similarities, particularly for their L1 and age of beginning English education, the participants across these two L2 groups

should have had cognitive perceptual mechanisms that operated similarly at the outset of learning L2-English. As a result, explanations for these differences are most likely associated with their differences in L2 input and experience across contexts.

In Chapter Four, the out-of-school L2 environments of the two L2 groups were compared. This comparison showed that, while the immersion learners can receive similar amounts of out-of-school L2 input as community-L2 peers, the immersion learners use the L2 significantly less often to interact with others. As foreign language learners, the immersion participants' more limited proportion of L2 use for interaction should be an expected consequence of living in a community where the L2 is not widely used and is thus a reflection of their more limited access to target-language speakers. However, it is also important to emphasize that, of this more limited access to L2 interaction outside of school, the majority of it is likely still with classmates. In contrast, the community-L2 participants are certain to have more opportunities to use English with L1-English speakers and/or L2-English speakers whose L1 is not Mandarin. Thus, the key difference may not just be the amount of L2 interaction, but who the interlocutors are.

In a usage-based approach, learners inductively learn abstract linguistic knowledge from concrete experiences with language (Bybee, 2006; Lieven, 2016), meaning that the knowledge that is acquired is ultimately a reflection of the social contexts in which the learner has been exposed to the language (Beckner et al., 2009; Wulff & Ellis, 2018). As a result, learning an L2 from actual usage within an immersion context may be incompatible with attaining L2 morphosyntactic ability that parallels L1 speakers. Immersion L2 learners begin their programs in classrooms with one proficient speaker of the target L2 (*i.e.*, their teacher), and many additional potential interlocutors with roughly equivalent

linguistic abilities both in terms of the L1 and L2 (*i.e.*, their classmates). Even for the presumably native or native-like input that immersion teachers provide, research has found that this input is shaped by the classroom context and provides limited opportunities for teachers to model the full range of L2 constructions. This can impact the distribution and frequency of constructions like verb forms with some forms being extremely infrequent in classrooms, if they are ever used at all (Peter, Hirata-Edds, & Montgomery-Anderson, 2008; Swain, 1988). In addition, the classroom context does not provide many opportunities with particular form-function mappings reflecting sociolinguistic distinctions (Harley, Allen, Cummins, & Swain, 1987; Mougeon, Nadasdi, & Rehner, 2010; Swain & Lapkin, 1990). These factors will mean that the acquisition of some L2 constructions will be impacted by their low frequency of occurrence within the immersion context.

Nevertheless, the English grammatical morphemes that were examined in the present study are all frequently used, and thus, should not be absent from the L2 input teachers provide for students. However, in contrast to teachers' presumably correct use of L2 grammatical morphemes, the same cannot be assumed for the L2 production of classmates. In the initial stages of an L2, learners are likely to model incorrect use of grammatical morphemes, as errors with these constructions are common regardless of a learner's L1 (Dulay & Burt, 1974; Goldschneider & DeKeyser, 2005; Larsen-Freeman, 1975; Shin & Milroy, 1999). The potential impacts of L2 learners modelling non-target L2 use within classrooms has been previously noted, especially for contexts, like immersion, where L2 learners all share an L1 (Lightbown, 1991; Lightbown, Courchêne, St John, Thérien, & Glidden, 1992). Thus, the immersion learners' low accuracy on the judgment task would suggest that a potential explanation for continued non-target L2 ability is

simply the persistence of overheard incorrect L2 forms used by classmates. However, this is likely only part of the explanation as learners should not be expected to easily acquire non-target L2 constructions that also contrast with the conventional usage of corresponding L1 constructions. For example, an L1-English-speaking learner in Spanish immersion would be unlikely to be misled by a classmate's production of a count noun unmarked for number, interpreting it as evidence that Spanish nouns are optionally quantified. Given that both the L1 and L2 use obligatory number-marking on nouns, the unmarked interpretation, though consistent with the hypothetical classmate's non-target production, should initially be blocked by the form-function mapping of the L1 construction. However, this would not be the case for L1-Mandarin speakers in English immersion, where the default (but incorrect) expectation should be that number is optionally marked on L2 noun phrases as it is in Mandarin. Essentially, non-target use is unlikely to take hold when the produced L2 form does not easily map onto entrenched L1 knowledge.

Where the L1 and the L2 contrast, MacWhinney (2017) lists social participation as a protective factor that can mitigate the impacts of entrenched L1 knowledge on L2 learning, noting that young L2 learners often have greater opportunities and pressure to integrate themselves with peer groups who speak the target language. In contrast, older learners tend to be less open to these types of experiences and thus end up isolated from opportunities for additional L2 input and experience. The importance of social participation is supported by research in community-L2 contexts that shows even among young L2 learners, the richness of their L2 environment predicts L2 morphosyntactic ability (Jia & Fuse, 2007; Paradis, 2011; Paradis & Jia, 2017; Paradis, Soto-Corominas, Chen,

& Gottardo, 2020; Paradis, Tulpar, & Arppe, 2016). Elements of a richer L2 environment can include factors that should directly index social participation such as use of the L2 for participation in extracurricular activities or when interacting with peers.

Though immersion learners are young, their unique L2 learning situation provides more limited access to some of the elements that constitute a richer L2 environment. In Chapter Four, it was found that the immersion learners' L2 use for extracurricular activities was extremely limited, and while individual learners demonstrated a range of L2 use within their peer groups, this type of participation with classmates is still isolated from communities where the L2 is used natively. Because classmates may depend on their shared L1 for the production and comprehension of the L2, it should not be expected that this type of social participation will contribute to mitigation of L1 entrenchment. This dynamic may also underlie findings for community-L2 learners that additional L2 input and experience at home does not always have a detectable benefit on L2 abilities (Paradis, 2011; Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017), and that potential benefits can depend on the L2 fluency of family members (Chondrogianni & Marinis, 2011; Hammer et al., 2012; Paradis, Soto-Corominas, Chen, & Gottardo, 2020; Sorenson Duncan & Paradis, 2020). A rich language environment is characterized as one where the amount of input and output children experience is diverse and complex (Paradis, Soto-Corominas, Chen, & Gottardo, 2020), and this type of experience is less likely to come from individuals with lower levels of L2 proficiency.

An example of how L2 use among L2 speakers who share an L1 may not benefit target-like learning of L2 constructions comes from consideration of the Chapter Three findings for plural-singular marking. Results of that study show that the immersion

learners did not comprehend plural '*-s*' as reliably as the L1-English control group. It was proposed that the immersion learners may use English number marking in a manner consistent with Mandarin, meaning that they rely on use of determiners or lexical quantifiers that precede the noun, and which are only used when pragmatically relevant. In effect, this would mean that the immersion learners' L2-English knowledge includes non-target optional number marking, and that the absence of the bound '*-s*' morpheme would be a less consistently used construction for number marking. If Mandarin-influenced number marking predominates within the immersion learners' social contexts, a usage-based approach should expect that this is what will be acquired. As a result, social participation within an immersion context does not provide access to speakers whose target language use can mitigate the impacts of L1 entrenchment. Instead, for some L2 constructions, it should be expected that participation could have the opposite impact as interaction with classmates contributes to entrenching non-target L2 ability. Essentially, this difference in the dynamics of social participation reflects a key difference between the community-L2 and immersion contexts.

For L2 educational programs like language immersion, a prescriptive standard of the target L2 is typically implicitly assumed to be the end goal. However, from a usage-based perspective, linguistic knowledge is an emergent result of development through language experience and domain-general cognitive mechanisms (Beckner et al., 2009; Ellis, 2008), and thus, L2 acquisition in an immersion context needs to be understood as a product of these elements. A usage-based approach assumes no a priori L2 representation to be completely or incompletely acquired, and thus, there is no final L2 competence that exists beyond what results from inductive learning in context. This means that ultimately what

immersion learners acquire of the L2 is a reflection of their actual L2 input and experience, and therefore any prescriptive L2 standard that is removed from their actual experiences is irrelevant. For community-L2 learners, living in a community where the target language is used natively ensures that L2 input and experience will be closely aligned with standard use of the target language.

5.2.5. Limitations

The aim of this dissertation was to investigate the L2 ability of the immersion participants in relation to other English-speaking groups and to investigate factors that impact their L2 ability. The presented studies demonstrate that the English abilities of the L2 learners in the immersion context differ from those of the comparison groups and that factors associated with learning context can predict these differences. Nevertheless, the scope of these conclusions should be interpreted in light of limitations in the study methodology. Here, limitations are acknowledged pertaining to both findings for L2 morphosyntactic ability and L2 vocabulary.

First, it is important to note that the morphosyntactic constructions studied were not selected to give a general survey of immersion learners' L2 English morphosyntactic ability. Instead, the studies focused on English grammatical morphemes because prior research evidence indicates that L1-Mandarin/Chinese speakers find them challenging (*e.g.,* Jia & Fuse, 2007; Jiang, 2007; Luk & Shirai, 2009; Paradis, Tulpar, & Arppe, 2016; Trenkic, Mirković, & Altmann, 2014). Thus, the lag in ability found between the immersion group, and both the community-L2 and L1-English control group, as well as the lag between the community-L2 and control group should not be generalized to other specific

L2-English morphosyntactic constructions. For example, while L1-Chinese community-L2 learners may have a more delayed rate of acquisition for English verbal morphology as compared to speakers of more typological similar L1s (Blom, Paradis, & Sorenson Duncan, 2012), this is not the case for acquisition of English complex syntax where an impact of L1 is not found (Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017). Given such findings, it is likely that both adolescent L1-Mandarin immersion and community-L2 learners would demonstrate different developmental profiles for English complex syntax or other grammatical constructions, as compared to their profile for grammatical morphemes.

Second, the differing levels of assessment for L2 ability across the domains of morphosyntax and vocabulary should be noted. For L2 ability with English grammatical morphemes, it was concluded here that development had plateaued, while for L2 receptive vocabulary, it was concluded that the immersion learners' vocabulary knowledge would continue to increase as they accumulated a quantity of L2 exposure equivalent to the community-L2 participants. Thus, it may appear that the L2 development of vocabulary ability is less impacted by context than is the acquisition of morphosyntax. However, this does not mean that there are not important differences in the L2 vocabulary knowledge of the immersion learners, as compared to either the community-L2 or L1-control groups. For example, Harley (1993) discusses the highly frequent French verbs AVOIR (*have*) and ÊTRE (*be*) and how learners in French immersion may continue to use these verbs consistently with their L1 English (*e.g.*, saying the French equivalent of '*I am hungry*' when '*I have hunger*' is the conventional way to express this in French). The PPVT does not probe this type of information, and thus, the possibility remains that the immersion participants'

vocabulary knowledge would show differences with community-L2 and L1 speaking peers if examined in this way.

5.2.6. Future directions

Currently, little research exists that specifically addresses the L2-English ability that L1-Mandarin immersion learners develop, and as such, potential future research directions are numerous. Nevertheless, there are two key extensions to the present dissertation that would benefit understanding of L2 development in immersion contexts. The following future directions examine either the emergence or persistence of non-target L2 ability more closely.

First, in order to better understand how non-target L2 ability emerges, research should examine factors that impact how immersion learners perceive their initial L2 input and how this influences subsequent development of the L2. As noted, research with adults indicates initial perception of L2 input is impacted not only by L1-tuned perceptual mechanisms, but also by factors common to L2 learners regardless of L1 (Sagarra & Ellis, 2013). Similarly, research with community-L2 learners shows that the rate of L2 morphosyntactic development can be impacted by factors including L1 (Blom & Baayen, 2013; Blom, Paradis, & Sorenson Duncan, 2012; Paradis, 2011), as well as variations between learners in cognitive abilities (Paradis, 2011; Paradis, Rusk, Sorenson Duncan, & Govindarajan, 2017; Verhagen, Leseman, & Messer, 2015). Thus, in order to understand factors that influence L2 acquisition in immersion, longitudinal research monitoring learner development from the very beginning of programs should examine learners' emerging production and comprehension of grammatical morphemes. Emerging ability

should be compared across programs that involve L2-English learners of different L1s (*i.e.*, comparing L2-English immersion learners in programs in Taiwan with ones in Spain). In addition, how cognitive abilities that impact initial perception of the L2 such as verbal memory, and those that may reflect ability to abstract patterns from L2 input such as analytic reasoning or statistical learning ability should also be examined for influences on individual learners' L2 development.

Next, in order to understand why non-target L2 forms potentially persist, future research should more closely examine how L1-Mandarin immersion learners use English to interact with their immersion classmates. For English plural-singular marking, it was proposed based on comprehension data that the immersion participants continue to use number marking in a manner consistent with Mandarin pragmatics. Direct observation of immersion learners' variable use of plural marking with peers and examination of discourse factors that may predict the suppliance or omission of different forms of number marking would strengthen conclusion made here. These issues can both be addressed in a future study that specifically collects information about whether peers interact with English speakers who are not classmates and observes a dyadic interaction task designed to elicit number-marking constructions.

5.3. Conclusions

The three studies presented in this dissertation compared the L2-English abilities of the studied L1-Mandarin immersion learners with other groups of English speakers, and examined how L2 input and experience impacted those abilities. In each study, differences in receptive L2 abilities were found between the immersion learners and comparison

groups. For a range of highly frequent English grammatical morphemes, the L1-Mandarin immersion learners were different from both community-L2 learning peers, and from L1-English speakers. For English receptive vocabulary, immersion learners were different from community-L2 peers and fell below the age-standardized test norm for L1 speakers. When the out-of-school L2 environments were compared, the immersion and community-L2 groups were largely similar in their access to additional English input. However, they differed in the amount that they used the L2 to interact with others. In discussion, the differences between experiences in the two L2 learning contexts were linked to differences in morphosyntactic ability concluding that even a semantically transparent English grammatical morpheme can be susceptible to sustained non-target use in an immersion context due to lack of opportunities for interaction with others who do not share the immersion learners' L1. This dissertation provides evidence that the receptive L2 ability immersion learners acquire is distinct from that acquired by other high-proficiency early L2 learners and underscores the importance of L2 experience in impacting what is acquired of an L2.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306.

Adlof, S., & Perfetti, C. (2013). Individual differences in word learning and reading ability. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), *Handbook of language and literacy: Development and disorders, second edition* (pp. 246–264). Guilford New York, NY.

Alanís, I. (2000). A Texas two-way bilingual program: Its effects on linguistic and academic achievement. *Bilingual Research Journal*, *24*(3), 225–248. https://doi.org/10.1080/15235882.2000.10162763

Allen, P., Swain, M., Harley, B., & Cummins, J. (1990). Aspects of classroom treatment: Toward a more comprehensive view of second language education. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 57–81). Cambridge University Press. https://doi.org/10.1017/CBO9781139524568.007

Arias-Trejo, N., Cantrell, L. M., Smith, L. B., & Canto, Elda, A. A. (2014). Early comprehension of the Spanish plural*. *Journal of Child Language*, *41*(6), 1356–1372. https://doi.org/10.1017/S0305000913000615

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, *20*(1), 50–57. https://doi.org/10.1111/j.1540-5826.2005.00120.x

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press. https://doi.org/10.1558/sols.v2i3.471

Ballinger, S., Lyster, R., Sterzuk, A., & Genesee, F. (2017). Context-appropriate crosslinguistic pedagogy: Considering the role of language status in immersion education. *Journal of Immersion and Content-Based Language Education*, *5*(1), 30–57. https://doi.org/10.1075/jicb.5.1.02bal

Barner, D., Thalwitz, D., Wood, J., Yang, S. J., & Carey, S. (2007). On the relation between the acquisition of singular-plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science*, *10*(3), 365–373. https://doi.org/10.1111/j.1467-7687.2007.00591.x

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bartoń, K. (2020). *MuMIn: Multi-model inference*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., … Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(December), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bild, E. R., & Swain, M. (1989). Minority language students in a French immersion programme: Their French proficiency. *Journal of Multilingual and Multicultural Development*, *10*(3), 255–274. https://doi.org/10.1080/01434632.1989.9994377

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249.

Blom, E., & Baayen, R. H. (2013). The impact of verb form, sentence position, home language, and second language proficiency on subject–verb agreement in child second language Dutch. *Applied Psycholinguistics*, *34*(04), 1–35. https://doi.org/10.1017/S0142716412000021

Blom, E., Boerma, T., Bosma, E., Cornips, L., van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual

children. *First Language*, *40*(2), 151–171. https://doi.org/10.1177/0142723719892794

Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *56*(February), 281–294. https://doi.org/10.1044/1092-4388(2012/11-0112)b

Blom, E., Paradis, J., & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular -s in child L2 English. *Language Learning*, *62*(3), 965–994. https://doi.org/10.1111/j.1467-9922.2012.00715.x

Bock, K., & Middleton, E. L. (2011). Reaching agreement. *Natural Language and Linguistic Theory*, *29*(4), 1033–1069. https://doi.org/10.1007/s11049-011-9148-y

Bock, K., & Miller, C. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93.

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org/

Brandt-Kobele, O. C., & Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. *Lingua*, *120*(8), 1910–1925. https://doi.org/10.1016/j.lingua.2010.02.008

Brown, R. (1973). *A first language: The early stages*. Cambridge MA: Harvard University Press.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.72

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353

Butler, Y. G., & Le, V.-N. (2018). A longitudinal investigation of parental social-economic status (SES) and young students' learning of English as a foreign language. *System*, *73*, 4–15. https://doi.org/10.1016/j.system.2017.07.005

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Carhill, A., Suárez-Orozco, C., & Páez, M. (2008). Explaining English language proficiency among adolescent immigrant students. *American Educational Research Journal*, *45*(4), 1155–1179. https://doi.org/10.3102/0002831208321443

Cazabon, M. T., Nicoladis, E., & Lambert, W. E. (1998). *Becoming bilingual in the Amigos two-way immersion program*. Santa Cruz, CA: Center for Research on Education, Diversity; Excellence, UC Berkeley.

Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, *35*(3), 243–262. https://doi.org/10.1093/applin/amt011

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chatterjee, S., & Hadi, A. S. (2006). Simple linear regression. In S. Chatterjee & A. S. Hadi (Eds.), *Regression analysis by example* (pp. 21–51). John Wiley & Sons, Inc.

Cheng, L. (2012). English immersion schools in China: Evidence from students and teachers. *Journal of Multilingual and Multicultural Development*, *33*(4), 379–391. https://doi.org/10.1080/01434632.2012.661436

Cheng, L. L.-S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, *30*(4), 509–542. https://doi.org/10.1162/002438999554192

Cheng, L. L.-S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese*. In G. Cinque & R. S. Kayne (Eds.), *Handbook of comparative syntax* (pp. 259–292). Oxford: Oxford University Press.

Cheng, L., Li, M., Kirby, J. R., Qiang, H., & Wade-Woolley, L. (2010). English language immersion and students' academic achievement in English, Chinese and mathematics. *Evaluation & Research in Education*, *24*(2), 163–163. https://doi.org/10.1080/09500790.2010.489150

Chomsky, N. (1976). *Reflections on language*. Fontana.

Chondrogianni, V. (2018). Child L2 acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 103–126). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.54.06cho

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, *15*(01), 5–21. https://doi.org/10.1017/S1366728911000368

Cintrón-Valentín, M., & Ellis, N. C. (2015). Exploring the interface: Explicit focus-on-form instruction and learned attentional biases in L2 Latin. *Studies in Second Language Acquisition*, *37*(02), 197–235.

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(01), 107–126. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395.

Cobo-Lewis, A. B., Pearson, B., Eilers, R. E., & Umbel, V. (2002). Effects of bilingualism and bilingual education on oral and written English skills: A multifactor study of standardized test outcomes. In D. K. Oller & R. E. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 64–97). Clevedon: Multilingual Matters.

Collins, M. F. (2010). ELL preschoolers' English vocabulary acquisition from storybook reading. *Early Childhood Research Quarterly*, *25*(1), 84–97. https://doi.org/10.1016/j.ecresq.2009.07.009

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge Textbooks in Linguistics.

Cummins, J. (1991). The politics of paranoia: Reflections on the bilingual education debate. In O. García (Ed.), *Bilingual education: Essays in honor of Joshua A. Fishman*. Amsterdam: John Benjamins Publishing.

Cummins, J. (1998). Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. In M. R. Childs & R. M. Bostwick (Eds.), *Learning through two languages: Research and practice. Second Katoh Gakuen international symposium on immersion and bilingual education* (pp. 34–37). Katoh Gakuen, Japan.

Cummins, J. (2009). Bilingual and immersion programs. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 161–181). Oxford: Blackwell Publishing Ltd.

Davies, B., Xu Rattanasone, N., & Demuth, K. (2017). Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development*, *13*(1), 38–53. https://doi.org/10.1080/15475441.2016.1219257

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Day, E. M., & Shapson, S. M. (1987). Assessment of oral communicative skills in early French immersion programmes. *Journal of Multilingual and Multicultural Development*, *8*(3), 237–260.

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*(July 2015), 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267–278. https://doi.org/10.1007/BF01067106

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728918001062

de Winter, J. F. C., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). Practical Assessment, Research, and Evaluation, 15. https://doi.org/10.7275/BJ1P-TS64

DeKeyser, R. (2012). Age effects in second language learning. In S. M. Gass & A. Mackey (Eds.), *Routledge handbook of second language acquisition* (pp. 442–460). Abingdon: Routledge.

Derrick, B., & White, P. (2017). Comparing two samples from an individual Likert question. *International Journal of Mathematics and Statistics*, *18*(3).

Dressman, M., & Sadler, R. W. (2020). *The handbook of informal language learning*. John Wiley & Sons.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*(1), 37–53.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed). Bloomington, MN: Pearson.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. https://doi.org/10.1093/applin/aml015

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). London: Routledge.

Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., & Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics*, *35*(3), 547–579. https://doi.org/10.1017/S0142716412000501

Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, *32*(4), 553–580. https://doi.org/10.1017/S0272263110000264

Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, *33*(4), 589–624. https://doi.org/10.1017/S0272263111000325

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*.

Emslie, H. C., & Stevenson, R. J. (1981). Pre-school children's use of the articles in definite and indefinite referring expressions. *Journal of Child Language*, *8*(02), 313–328. https://doi.org/10.1017/S0305000900003214

Erdos, C., Genesee, F., Savage, R., & Haigh, C. A. (2011). Individual differences in second language reading outcomes. *International Journal of Bilingualism*, *15*(1), 3–25. https://doi.org/10.1177/1367006910371022

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171. https://doi.org/10.1191/1362168805lr161oa

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, *32*(4), 711–738. https://doi.org/10.1017/S0142716411000038

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104. https://doi.org/10.1006/jmla.1999.2638

Forgays, D. K., Hyman, I., & Schreiber, J. (2014). Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. *Computers in Human Behavior*, *31*, 314–321. https://doi.org/10.1016/j.chb.2013.10.053

Fortune, T. W., Christian, D., & Tedick, D. J. (2011). *Immersion education: Practices, policies, possibilities*. Bristol: Multilingual Matters.

Fortune, T. W., & Tedick, D. J. (2015). Oral proficiency assessment of English-proficient K-8 Spanish immersion students. *Modern Language Journal*, *99*(4), 637–655. https://doi.org/10.1111/modl.12275

Fortune, T. W., Tedick, D. J., & Walker, C. L. (2008). Integrated language and content teaching: Insights from the immersion classroom. In T. W. Fortune & D. J. Tedick (Eds.), *Pathways to Multilingualism* (pp. 71–96). Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847690371-007

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gao, X., & Wang, W. (2017). Bilingual education in the People's Republic of China. In O. García, A. M. Y. Lin, & S. May (Eds.), *Bilingual and multilingual education* (pp. 219–231). Springer International Publishing.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, V. C. M., Laporte, N., & Thomas, E. M. (2005). Differentiation, carry-over, and the distributed characteristic in bilinguals: Structural "mixing" of the two languages? *Proceedings of the 4th International Symposium on Bilingualism*, 838–851.

Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning*, *26*(2), 267–280. https://doi.org/10.1111/j.1467-1770.1976.tb00277.x

Genesee, F. (1978). Individual differences in second-language learning. *Canadian Modern Language Review*, *34*(3), 490–504.

Genesee, F. (1981). A comparison of early and late second language learning. *Canadian Journal of Behavioural Science*, *13*(2), 115–128. https://doi.org/10.1037/h0081168

Genesee, F. (1984). *Historical and theoretical foundations of immersion education* (pp. 32–57). California State Department of Education.

Genesee, F. (1987). *Learning through two languages: Studies in immersion and bilingual education*. Cambridge, MA: Newbury House Publishers.

Genesee, F. (2006). What do we know about bilingual education for majority-language students? In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 547–576). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756997.ch21

Genesee, F., & Hamayan, E. (1980). Individual differences in second language learning. *Applied Psycholinguistics*, *1*(1980), 95–110.

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. https://doi.org/10.1075/jicb.1.1.02gen

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge. *Studies in Second Language Acquisition*, *37*(02), 269–297.

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, *29*, 41–65.

Goldschneider, J. M., & DeKeyser, R. M. (2005). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants.

Language Learning, *55*(S1), 27–77. https://doi.org/10.1111/j.0023-8333.2005.00295.x

Goriot, C., Van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., Unsworth, S., … James, M. (2018). Using the Peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, 1–23. https://doi.org/10.1080/13670050.2018.1494131

Graves, M., & Koziol, S. (1971). Noun plural development in primary grade children. *Child Development*, *42*(4), 1165–1173.

Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.

Grosjean, F. (2016). The complementarity principle and its impact on processing, acquisition, and dominance. In C. Silva-Corvalan & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 66–84). Cambridge University Press. https://doi.org/10.1017/CBO9781107375345.004

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, *26*, 321–351.

Hamayan, E., Genesee, F., & Tucker, G. R. (1977). Affective factors and language exposure in second language learning. *Language Learning*, *27*(2), 225–241. https://doi.org/10.1111/j.1467-1770.1977.tb00120.x

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., Goldstein, B., … Goldstein, B. (2012). Predicting Spanish-English bilingual children's language

abilities. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1251–1264. https://doi.org/10.1044/1092-4388(2012/11-0016)

Harley, B. (1986). *Age in second language acquisition*. Clevedon: Multilingual Matters.

Harley, B. (1992). Patterns of second language development in French immersion*. *French Language Studies*, *2*(September 1991), 159–183.

Harley, B. (1993). Instructional strategies and SLA in early French immersion. *Studies in Second Language Acquisition*, *15*, 245–259.

Harley, B., Allen, P., Cummins, J., & Swain, M. (1987). *The nature of language proficiency* (Vol. 1, p. 3).

Harley, B., Cummins, J., Swain, M., & Allen, P. (1990). The nature of language proficiency. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 7–25). Cambridge University Press. https://doi.org/10.1017/CBO9781139524568.003

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*(3), 379–400.

Harley, B., & Jean, G. (1999). Vocabulary skills of French immersion students in their second language. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, *4*(2).

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition*, *11*(4), 415–439. https://doi.org/10.1017/S0272263100008421

Harley, B., & Swain, M. (1978). An analysis of the verb system used by young learners of French. *Interlanguage Studies Bulletin*, *3*(1), 35–79.

Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Criper, & A. P. R. E. Howatt (Eds.), *Interlanguage* (Vol. 291, p. 311). Edinburgh University Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing.

Hart, D., Lapkin, S., & Swain, M. (1991). Secondary level immersion French skills: A possible plateau effect. In L. M. Malave & G. Duquette (Eds.), *Language, culture and cognition: A collection of studies in first and second language acquisition*. Multilingual Matters Ltd.

Hermanto, N., Moreno, S., & Bialystok, E. (2012). Linguistic and metalinguistic outcomes of intense immersion education: How bilingual? *International Journal of Bilingual Education and Bilingualism*, *15*(2), 131–145. https://doi.org/10.1080/13670050.2011.652591

Herschensohn, J., Stevenson, J., & Waltmunson, J. (2005). Children's acquisition of L2 Spanish morphosyntax in an immersion setting. *International Review of Applied Linguistics in Language Teachings*, *43*(3), 193–217. https://doi.org/10.1515/iral.2005.43.3.193

Hoff, E. (2020). Lessons from the study of input effects on bilingual development. *International Journal of Bilingualism*, *24*(1), 82–88. https://doi.org/10.1177/1367006918768370

Hoff, E., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics*, *190*, 241–245.e1. https://doi.org/10.1016/j.jpeds.2017.06.071

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, *38*(4), 271–278. https://doi.org/10.1016/j.jcomdis.2005.02.003

Hootsuite. (2019). *Digital 2019 Q4 global digital statshot (October 2019) v01* [Internet]. Retrieved from https://www.slideshare.net/DataReportal/digital-2019-q4-global-digital-statshot-october-2019-v01

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Hwu, F., & Sun, S. (2012). The aptitude-treatment interaction effects on the learning of grammar rules. *System*, *40*(4), 505–521. https://doi.org/10.1016/j.system.2012.10.009

Izquierdo, J. (2007). *Multimedia environments in the foreign language classroom: Effects on the acquisition of the French perfective and imperfective distinction* (PhD thesis). McGill University.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis. *Studies in Second Language Acquisition*, *21*(03). https://doi.org/10.1017/S0272263199003034

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. https://doi.org/10.1037/a0021284

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1297–1311. https://doi.org/10.1044/1092-4388(2003/101)

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131–161.

Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1280–1299. https://doi.org/10.1044/1092-4388(2007/090)

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*(3), 940–967. https://doi.org/10.1111/j.1467-9922.2010.00627.x

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Kalia, V., Lane, P. D., & Wilbourn, M. P. (2018). Cognitive control and phonological awareness in the acquisition of second language vocabulary within the Spanish-English dual immersion context. *Cognitive Development*, *48*, 176–189. https://doi.org/10.1016/j.cogdev.2018.08.010

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed). San Antonio, TX: Pearson.

Keating, G. D. (2009). Sensitivity to violation of gender agreement in native and nonnative Spanish. *Language Learning*, *59*(September), 503–535.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Kidd, E., & Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, *6*(1), 319–340. https://doi.org/10.1146/annurev-linguistics-011619-030326

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Klein, W., & Perdue, C. (1992). *Utterance structure: Developing grammars again*. Amsterdam: John Benjamins Publishing Co.

Knell, E., & Chi, Y. (2012). The roles of motivation, affective attitudes, and willingness to communicate among Chinese students in early English immersion programs.

International Education, *41*(2), 66–87. https://doi.org/10.1207/s1532799xssr0901_5

Knell, E., Siegel, L. S., & Lin, Z. (2007). Early English immersion and literacy in Xi'an , China. *The Modern Language Journal*, *91*(3), 395–417.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25.

Kowal, M., & Swain, M. (1997). From semantic to syntactic processing. In R. K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 284–309). Cambridge: Cambridge University Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636–645. https://doi.org/10.3758/BF03213001

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85. https://doi.org/10.1080/17439880903561876

Kuznetsova, A., Brockhoff, P. B., & Christiansen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal Of Statistical Software*, *82*(13), 1–26. https://doi.org/doi: 10.18637/jss.v082.i13

Kwon, S., & Zribi-Hertz, A. (2004). Number from a syntactic perspective: Why plural marking looks 'truer' in French than in Korean*. In O. Bonami & P. Cabredo Hofherr (Eds.), *Empirical issues in formal syntax and semantics 5* (pp. 133–158).

Lapkin, S., Swain, M., & Shapson, S. (1990). French immersion research agenda for the 90s. *Canadian Modern Language Review*, *46*(4), 638–674.

Larsen-Freeman, D. E. (1975). The acquisition of grammatical morphemes by adult ESL students*. *Tesol Quarterly*, *9*(4), 409419.

Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Levshina, N. (2015). *How to do linguistics with R* (p. 443). Amsterdam: John Benjamins Publishing Company.

Lew-Williams, C. (2017). Specific referential contexts shape efficiency in second language processing: Three eye-tracking experiments with 6- and 10-year-old children in Spanish immersion schools. *Annual Review of Applied Linguistics*, *37*(2017), 128–147. https://doi.org/10.1017/S0267190517000101

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar* (p. 691). University of California Press.

Li, M., Cheng, L., & Kirby, J. R. (2012). *Phonological awareness and listening comprehension among Chinese English-immersion students*. 21.

Li, M., Kirby, J. R., Cheng, L., Wade-Woolley, L., & Qiang, H. (2012). Cognitive predictors of English reading achievement in Chinese English-immersion students. *Reading Psychology*, *33*(5), 423–447. https://doi.org/10.1080/02702711.2010.528663

Li, P., Ogura, T., Barner, D., Yang, S. J., & Carey, S. (2009). Does the conceptual distinction between singular and plural sets depend on language? *Developmental Psychology*, *45*(6), 1644–1653. https://doi.org/10.1037/a0015553

Li, S. (2015). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*(3), 693–722. https://doi.org/10.1017/S0142716418000796

Li, Y. H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics*, *8*(1), 75–99. https://doi.org/10.1023/A:1008306431442

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, *8*(03), 346–368. https://doi.org/10.1017/langcog.2016.16

Lightbown, P. M. (1991). What have we here? Some observations on the role of instruction in second language acquisition. In R. Phillipson, E. Kellerman, L. Selinker, M. S. Smith, & M. Swain (Eds.), *Foreign language pedagogy research: A commemorative volume for Claus Faerch*. Clevedon: Multilingual Matters.

Lightbown, P. M., Courchêne, R., St John, J., Thérien, C., & Glidden, J. I. (1992). Can they do it themselves? A comprehension-based ESL course for young children. In *Comprehension-based second language teaching*. Ottawa: University of Ottawa Press.

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878–919. https://doi.org/10.1080/09588221.2018.1541359

Lin, L. C., & Johnson, C. J. (2016). Mandarin-English bilingual vocabulary development in an English-immersion preschool: How does it compare with monolingual development? *International Journal of Bilingualism*, *20*(2), 173–189. https://doi.org/10.1177/1367006914547662

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Lindholm-Leary, K. (2011). Student outcomes in Chinese two-way immersion programs: Language proficiency, academic achievement, and student attitudes. In D. J. Tedick, D. Christian, & T. Williams Fortune (Eds.), *Immersion education: Practices, policies, possibilities* (pp. 81–103). Bristol: Multilingual Matters.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning*, *24*(6), 515–525. https://doi.org/https://doi.org/10.1111/j.1365-2729.2008.00289.x

Luk, Z. P.-S., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive 's. *Language Learning*, *59*(4), 721–754.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins Publishing.

Lyster, R. (2008). Evolving perspectives on learning French as a second language through immersion. In D. Ayoun (Ed.), *Studies in French applied linguistics* (Vol. 21, pp. 3–36). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.21.05lys

Lyster, R., & Genesee, F. (2019). Immersion education. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781405198431.wbeal0525.pub2

Lyster, R., & Tedick, D. J. (2014). Research perspectives on immersion pedagogy: Looking back and looking forward. *Journal of Immersion and Content-Based Language Education*, *2*(2), 210–224. https://doi.org/10.1075/jicb.2.2.04lys

Macmillian, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah: Psychology Press.

MacWhinney, B. (2017). Entrenchment in second-language learning. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*.

Maratsos, M. P. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, *45*(2), 446–455.

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science*, *21*(3), 283–304. https://doi.org/10.1207/s15516709cog2103_2

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, *34*(1), 9–34.

Matthews, D. E., & Theakston, A. L. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: The effects of frequency, phonology, and competition. *Cognitive Science*, *30*(6), 1027–1052. https://doi.org/10.1207/s15516709cog0000_66

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*(3), 395–423. https://doi.org/10.1017/S0142716400003064

McSweeney, M. A. (2017). I text English to everyone: Links between second-language texting and academic proficiency. *Languages*, *2*(3), 7. https://doi.org/10.3390/languages2030007

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/a0021890

Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English bilingual children. *International Journal of Bilingualism*, *11*(3), 301–328.

Mougeon, R., Nadasdi, T., & Rehner, K. (2010). *The sociolinguistic competence of immersion students*. Bristol: Multilingual Matters.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*(4), 578–596. https://doi.org/10.1093/applin/amm056

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*(4), 463–482. https://doi.org/10.1093/applin/amu024

Muñoz, C., Cadierno, T., & Casas, I. (2018). Different starting points for English language learning: A comparative study of Danish and Spanish young learners. *Language Learning*, *68*(4), 1076–1109. https://doi.org/10.1111/lang.12309

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, *44*(01), 1–35. https://doi.org/10.1017/S0261444810000327

Nakanishi, K., & Tomioka, S. (2004). Japanese plurals are exceptional. *Journal of East Asian Linguistics*, *13*(2), 113–140. https://doi.org/10.1023/B:JEAL.0000019058.46668.c1

Nicolay, A.-C., & Poncelet, M. (2013). Cognitive abilities underlying second-language vocabulary acquisition in an early second-language immersion education context: A longitudinal study. *Journal of Experimental Child Psychology*, *115*(4), 655–671. https://doi.org/10.1016/j.jecp.2013.04.002

Ní Dhiorbháin, A., & Ó Duibhir, P. (2017). An explicit-inductive approach to grammar in Irish-medium immersion schools. *Language Awareness*, *26*(1), 3–24. https://doi.org/10.1080/09658416.2016.1261870

Olsson, E., & Sylvén, L. K. (2015). Extramural English and academic vocabulary. A longitudinal study of CLIL and non-CLIL students in Sweden. *Apples: Journal of Applied Language Studies*, *9*. https://doi.org/10.17011/apples/urn.201512234129

Oniski, K. K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Ó Duibhir, P. (2011). 'I thought we had good Irish': Irish immersion students' insights into their target language use. In D. J. Tedick, D. Christian, & T. W. Fortune (Eds.), *Immersion education : Practices, policies, possibilities*. Bristol: Multilingual Matters.

Ó Duibhir, P., Ní Dhiorbháin, A., & Cosgrove, J. (2016). An inductive approach to grammar teaching in grade 5 & 6 Irish immersion classes. *Journal of Immersion and Content-Based Language Education*, *4*(1), 33–58. https://doi.org/10.1075/jicb.4.1.02dui

Paradis, J. (2010). Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning*, *60*(3), 651–680. https://doi.org/10.1111/j.1467-9922.2010.00567.x

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. M. Brown & B. Dailey (Eds.), *43rd Boston university conference on language development*. Somerville: Cascadilla Press.

Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: Language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, *20*(1).

Paradis, J., Rusk, B. V., Sorenson Duncan, T., & Govindarajan, K. (2017). Children's second language acquisition of English complex syntax: The role of age, input and cognitive factors. *Annual Review of Applied Linguistics*, (37), 148–167.

Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, 1–27. https://doi.org/10.1017/S014271642000017X

Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: Individual variation in long-term outcomes. *Journal of Child Language*, *43*(03), 553–580. https://doi.org/10.1017/S0305000915000562

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning by listening to songs. *Studies in Second Language Acquisition*, *41*(4), 745–768. https://doi.org/10.1017/S0272263119000020

Peter, L., Hirata-Edds, T., & Montgomery-Anderson, B. (2008). Verb development by children in the Cherokee language immersion program, with implications for teaching. *International Journal of Applied Linguistics*, *18*(2), 166–187. https://doi.org/10.1111/j.1473-4192.2008.00199.x

Peter, L., Sly, G., & Hirata-Edds, T. (2011). Using language assessment to inform instruction in indigenous language immersion. In Diane. J. Tedick, D. Christian, & T. Williams

Fortune (Eds.), *Immersion education: Practices, policies, possibilities* (pp. 187–208). Bristol: Multilingual Matters.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, *169*(1), 142–168. https://doi.org/10.1075/itl.00010.pet

Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, *69*(3), 747–782. https://doi.org/10.1111/lang.12351

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds*. *Bilingualism: Language and Cognition*, *19*(5), 1023–1041. https://doi.org/10.1017/S1366728915000322

Plester, B., Wood, C., & Joshi, P. (2009). Exploring the relationship between children's knowledge of text message abbreviations and school literacy outcomes. *British Journal of Developmental Psychology*, *27*(1), 145–161. https://doi.org/https://doi.org/10.1348/026151008X320507

Porretta, V., Kyröläinen, A.-J., van Rij, J., & Järvikivi, J. (2018). *VWPre: Tools for preparing visual world data*.

Prevoo, M. J. L., Malda, M., Mesman, J., Emmen, R. a. G., Yeniad, N., Ijzendoorn, M. H. V., & Linting, M. (2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language*. *Journal of Child Language*, *41*(5), 963–984. https://doi.org/10.1017/S0305000913000299

Psychology Software Tools. (2012). *E-Prime 2.0*. Pittsburgh: Psychology Software Tools. Retrieved from http://www.pstnet.com

Qiang, H., & Kang, Y. (2011). English immersion in China as a case of educational transfer. *Frontiers of Education in China*, *6*(1), 8–36. https://doi.org/10.1007/s11516-011-0120-8

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rezzonico, S., Goldberg, A., Milburn, T., Belletti, A., & Girolametto, L. (2017). English verb accuracy of bilingual Cantonese–English preschoolers. *Language, Speech, and Hearing Services in Schools*, *48*(3), 153–167. https://doi.org/10.1044/2017_LSHSS-16-0054

Rojas, R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., & Reese, L. (2016). Interlocutor differential effects on the expressive language skills of Spanish-speaking English learners. *International Journal of Speech-Language Pathology*, *18*(2), 166–177.

Rullmann, H., & You, A. (2006). General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese. In K. von Heusinger & K. Turner (Eds.), *Where semantics meets pragmatics* (pp. 175–196). Leiden: Brill.

Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, *35*(02), 261290. https://doi.org/10.1017/S0272263112000885

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses*. Peter Lang Publishing Incorporated.

Scheidnes, M. (2020). Sentence repetition and non-word repetition in early total French immersion. *Applied Psycholinguistics*, *41*(1), 107–131. https://doi.org/10.1017/S0142716419000420

Schütze, C. T., & Sprouse, J. (2014). Judgment data. *Research Methods in Linguistics*, *27*, 50.

Sheng, L., Lam, B. P. W., Cruz, D., & Fulton, A. (2016). A robust demonstration of the cognate facilitation effect in first-language and second-language naming. *Journal of Experimental Child Psychology*, *141*, 229–238. https://doi.org/10.1016/j.jecp.2015.09.007

Shin, S. J., & Milroy, L. (1999). Bilingual language acquisition by Korean schoolchildren in New York City. *Bilingualism: Language and Cognition*, *2*(2), 147–167. https://doi.org/10.1017/S1366728999000243

Shirai, Y. (2009). *Temporality in first and second language acquisition* (W. Klein & P. Li, Eds.). De Gruyter Mouton. https://doi.org/10.1515/9783110199031.167

Siegel, L., Knell, E., Pei, M., Qiang, H., Zhao, W., & Zhao, L. (2010). Teaching English to Chinese-speaking children. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner* (pp. 233–255). Dordrecht: Springer. https://doi.org/10.1007/978-90-481-3840-1

Song, Y. (2015). L2 processing of plural inflection in English. *Language Learning*, *65*(2), 233–267. https://doi.org/10.1111/lang.12100

Sorenson Duncan, T., & Paradis, J. (2020). Home language environment and children's second language acquisition: The special status of input from older siblings. *Journal of Child Language*, 1–24. https://doi.org/10.1017/S0305000919000977

Soto-Corominas, A., Paradis, J., Rusk, B. V., Marinova-Todd, S., & Zhang, X. (2020). Oral language profiles of English second language learners in adolescence: Cognitive and input factors influence how they compare to their monolingual peers. *Studies in Second Language Acquisition*, 1–24.

Spinner, P., & Gass, S. M. (2019). *Using judgments in second language acquisition research*. Abingdon: Routledge.

Sun, H., Steinkrauss, R., Tendeiro, J., & de Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition*, *19*(3), 550–566. https://doi.org/10.1017/S1366728915000243

Swain, M. (1985). *Communicative competence: Some roles of comprehensible input and output in its development*.

Swain, M. (1988). Manipulating and complementing content teaching to maximize second language learning. *TESL Canada Journal*, *6*(I), 68–83.

Swain, M. (2000). French immersion research in Canada: Recent contributions to SLA and Applied Linguistics. *Annual Review of Applied Linguistics*, *20*, 199–212.

Swain, M. (2005). The output hypothesis: Theory and research. *Handbook of Research in Second Language Teaching and Learning*, 471–483. https://doi.org/10.4324/9781410612700-38

Swain, M., & Carroll, S. (1987). The immersion observation study. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of bilingual proficiency: Final report* (Vol. 2, pp. 190–341). Modern Language Centre.

Swain, M., & Johnson, R. K. (1997). Immersion education: A category within bilingual education. In R. K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 1–16). Cambridge: Cambridge University Press.

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

Swain, M., & Lapkin, S. (1990). Aspects of the sociolinguistic performance of early and late French immersion students. In R. C. Scarcella, E. S. Andersen, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 41–54). New York: Newbury.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371–391.

Tarone, E., & Swain, M. (1995). A sociolinguistic perspective on second language use in immersion classrooms. *The Modern Language Journal*, *79*(2), 166–178. https://doi.org/10.2307/329617

Tedick, D. J., & Wesely, P. M. (2015). A review of research on content-based foreign/second language education in US K-12 contexts. *Language, Culture and Curriculum*, *28*(1), 25–40. https://doi.org/10.1080/07908318.2014.1000923

Tedick, D. J., & Young, A. I. (2016). Fifth grade two-way immersion students' responses to form-focused instruction. *Applied Linguistics*, *37*(6), 784–807. https://doi.org/10.1093/applin/amu066

Thomas, E. M., Williams, N., Jones, L. A., Davies, S., & Binks, H. (2014). Acquiring complex structures under minority language conditions: Bilingual acquisition of plural morphology in Welsh. *Bilingualism: Language and Cognition*, *17*(3), 478–494. https://doi.org/10.1017/S1366728913000497

Trenkic, D., Mirković, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*(2), 237–257. https://doi.org/10.1017/S1366728913000321

Tucker, G. R. (2008). St. Lambert immersion study. In J. González (Ed.), *Encyclopedia of Bilingual Education*. Thousand Oaks CA: SAGE Publications, Inc. https://doi.org/10.4135/9781412963985.n307

Turnbull, M., Lapkin, S., & Hart, D. (2001). Grade 3 immersion students' performance in literacy and mathematics: Province-wide results from Ontario (1998-99). *Canadian Modern Language Review*, *58*(1), 9–26. https://doi.org/10.3138/cmlr.58.1.9

Turnbull, M., Lapkin, S., Hart, D., & Swain, M. (1998). Time on task and immersion graduates' French proficiency. In S. Lapkin (Ed.), *French second language education in Canada: Empirical studies* (pp. 31–55). Toronto: University of Toronto Press.

Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergarteners: What are the best predictors? *Bilingualism: Language and Cognition*, *9*(1), 33–49. https://doi.org/10.1017/S1366728905002361

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, *93*(2), 185–238. https://doi.org/10.1016/j.bandl.2004.10.001

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*(1), 88–110.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). On the role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, *35*(04), 765–805.

Unsworth, S., Brouwer, S., Bree, E. de, & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, 1–31. https://doi.org/10.1017/S0142716419000225

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. *Empirical Perspectives on Anaphora Resolution*, (277), 267–293.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*(2), 390–416. https://doi.org/10.1111/lang.12105

Vandergrift, L., & Baker, S. C. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *Canadian Modern Language Review*, *74*(1), 79–100. https://doi.org/10.3138/cmlr.3906

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed). New York: Springer.

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, *141*, 65–82. https://doi.org/10.1016/j.jecp.2015.06.015

Verhagen, J., Leseman, P., & Messer, M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, *65*(2), 417–448. https://doi.org/10.1111/lang.12101

Verheijen, L. (2013). The effects of text messaging and instant messaging on literacy. *English Studies*, *94*(5), 582–602. https://doi.org/10.1080/0013838X.2013.795737

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing: CTOPP2*. San Antonio, TX: Pearson.

Walker, D., Sepulveda, S. J., Hoff, E., Rowe, M. L., Schwartz, I. S., Dale, P. S., … Bigelow, K. M. (2020). Language intervention research in early childhood care and education: A systematic survey of the literature. *Early Childhood Research Quarterly*, *50*, 68–85. https://doi.org/10.1016/j.ecresq.2019.02.010

Warden, M. (1997). *The effect of form-focused instruction on control over grammatical gender by French immersion students in grade 11* (PhD thesis). University of Toronto, Toronto.

Webb, S. (2019). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (p. 15). London: Routledge.

Webb, S., & Chang, A. C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in and extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651–675. https://doi.org/10.1017/S0272263114000606

Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). Proficiency effects and distance effects in nonnative processing of English number agreement. *Proceedings of the Annual Boston University Conference on Language Development*, *34*, 445–456.

Wesche, M. (1993). French immersion graduates at university and beyond: What difference has it made? *Georgetown University Round Table on Languages and Linguistics*, 208–240.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *The Language Learning Journal*, *40*(1), 125–137. https://doi.org/10.1080/09571736.2012.658232

Wojcik, S., & Hughes, A. (2019). *Sizing up Twitter users*. Pew Research Centre. Retrieved from Pew Research Centre website: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

Wood, J. N., Kouider, S., & Carey, S. (2009). Acquisition of singular-plural morphology. *Developmental Psychology*, *45*(1), 202–206. https://doi.org/10.1037/a0014432.Acquisition

Wood, S. N. (2006). *Generalized additive models: An introduction with R* (p. 410). London: Chapman; Hall. https://doi.org/10.1111/j.1541-0420.2007.00905_3.x

Woumans, E., Ameloot, S., Keuleers, E., & Van Assche, E. (2019). The relationship between second language acquisition and nonverbal cognitive abilities. *Journal of Experimental Psychology: General*, *148*(7), 1169–1177. https://doi.org/10.1037/xge0000536

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Studies in bilingualism* (Vol. 54, pp. 37–56). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.54.03wul

Xiong, T., & Feng, A. (2018). Localizing immersion education: A case study of an international bilingual education program in south china. *International Journal of Bilingual Education and Bilingualism*, *0*(0), 1–14. https://doi.org/10.1080/13670050.2018.1435626

Xu Rattanasone, N., Davies, B., Schembri, T., Andronos, F., & Demuth, K. (2016). The iPad as a research tool for the understanding of English plurals by English, Chinese, and other L1 speaking 3-and 4-year-olds. *Frontiers in Psychology*, *7*, 1–11. https://doi.org/10.3389/fpsyg.2016.01773

Yilmaz, Y., & Granena, G. (2019). Cognitive individual differences as predictors of improvement and awareness under implicit and explicit feedback conditions. *The Modern Language Journal*, *103*(3), 686–702. https://doi.org/10.1111/modl.12587

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*(4), 541–572.

Zapf, J. A., & Smith, L. B. (2007). When do children generalize the plural to novel nouns? *First Language*, *27*(1), 53–73. https://doi.org/10.1177/0142723707070286

Zdorenko, T., & Paradis, J. (2008). The acquisition of articles in child second language English: Fluctuation, transfer, or both? *Second Language Research*, *24*(2), 227–250. https://doi.org/10.1177/0267658307086302

APPENDICES

Appendix A: Grammaticality Judgment Task Stimulus List

1. Easy Stimulus Items

***Articles* – Article in () omitted in ungrammatical stimuli**

1. I am sitting at (the) same desk I sat at yesterday.
2. My sister was wearing (a) hat but she still got a sunburn.
3. I don't have (an) iPhone, but I have had two Android phones.
4. The player scored (the) last goal of the hockey game last night.
5. Some people taking (a) walk at the park saw swans in the pond.
6. My classmate just borrowed (the) last pencil from my pencil box.

Past tense - grammatical/ungrammatical

1. Last night, my father cooks/cooked dinner for all of my friends.
2. Two days ago, my friends played/play a soccer game after school.
3. Someone tipped/tips over the garbage can outside our classroom earlier today.
4. The students finished/finish all of their homework at school yesterday.
5. A car crashed/crashes into a tree near my house last week.
6. Yesterday afternoon, my sister and I walked/walk home from school together.

Plural Marking

1. We saw a few cars/car going east along the highway.
2. The manager told her new employees many tables/table needed clearing.
3. An airplane carrying many passengers/passenger landed safely on the runway.
4. This summer, his mother's car needs all the tires/tire replaced.
5. I will write three separate tests/test this Tuesday coming up.
6. My friend wants to buy a new pair of shoes/shoe for gym.

2. Difficult Stimulus Items

***BE* - grammatical/ungrammatical**

1. The problems with the computer were/was fixed by the technician.
2. The letter from the lawyers was/were concerned with illegal file sharing.
3. The wheels on the car were/was repaired in the mechanic's shop.
4. The puppy beside the kittens was/were very soft and fluffy.
5. The songs on this album were/was recorded in an excellent studio.
6. The student near the children was/were making a lot of noise.

DO

1. The owners of the cat do/does not like dogs very much.
2. The bottle beside the buckets does/do not get filled with water.
3. The doctors for the patient do/does not think the medicine is working.
4. The chair near the tables does/do not need to be painted.
5. The lawyers for the defendant do/does not have a trial today.
6. The wallet beside the keys does/do not belong to my father.

Third-person marking

1. The keys to the door open/opens others in the house too.
2. The owner of the businesses earns/earn a lot of money in profit.
3. The girls near the teacher read/reads the book to everyone.
4. The lizard on the rocks feels/feel very warm in the sun.
5. The sentences on the paper show/shows the answers to the questions.
6. The road with the potholes causes/cause a lot of damage to tires.

Appendix B: Correlations for Modeled Variables

1. Correlations between IL2 Variables

	receptive vocabulary	verbal memory (VM)	nonverbal reasoning (NVR)	relative proficiency (RP)	home language use (HLU)	peer language use (PLU)	Eng. hours per week (EHW)
VM	.144						
NVR	.184	-.201					
RP	.275	.082	-.014				
HLU	.187	.215	-.359	.117			
PLU	.408*	-.028	-.072	.513**	.186		
EHW	.211	.086	-.250	.376*	.154	.422*	
L1 fluency	-.246	.069	.108	-.222	-.122	-.642***	.045

2. Correlations between ML2 Variables

	receptive vocabulary	verbal memory (VM)	nonverbal reasoning (NVR)	relative proficiency (RP)	home language use (HLU)	peer language use (PLU)	Eng. hours per week (EHW)
VM	.331						
NVR	.596**	.150					
RP	.451*	.331	.101				
HLU	.475*	.275	.157	.458*			
PLU	.664***	.298	.051	.717***	.491*		
EHW	-.054	.361	.003	.149	.089	.054	
L1 fluency	-.289	-.014	-.137	-.663***	-.504*	-.378	.182

Note. Correlations between variable for each adolescent participant group. * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix C: Plural-Singular Sentence Stimuli

1. The apple/apples had been picked from our tree.
2. The brush/brushes could be used to paint pictures.
3. The box/boxes might have something inside.
4. The egg/eggs might not be fresh.
5. The elephant/elephants might like to eat peanuts.
6. The bus/buses must wait for the students.
7. The fork/forks can get put in the drawer.
8. The mailbox/mailboxes will get mail every day.
9. The banana/bananas from the market should be ready to eat.
10. The pencil/pencils can be used for writing the test.
11. The orange/oranges will be cut up and eaten.
12. The cup/cups will be filled with coffee.
13. The strawberry/strawberries will taste nice this time of year.
14. The doughnut/ doughnuts would not be a healthy snack.
15. The carrot/carrots will be nice in the soup.
16. The monkey/monkeys living here should not be fed.

Note. The 16 stimulus nouns are bolded. Sentence stimuli with one of the two noun forms were divided into two lists with equal numbers of plural and singular stimuli. Participants would have heard only one from each pair and the order was randomized for each participant. Nouns were roughly balanced for the three potential plural allomorphs /s/ (carrots, cups, doughnuts, elephants, and forks), /z/ (apples, bananas, eggs, monkeys, pencils, and strawberries), and /əz/ (boxes, brushes, buses, mailboxes, and oranges) when in their plural form. According to the Corpus of Contemporary American English (Davies, 2010), the median frequency for the 16 nouns was 12,818 tokens per million words (*IQR* = 27,223 to 7,591) with the least frequent lemma being 'mailbox' at 2,736 tokens per million words.