2-2019

Mandarin Assessment in Chinese-English Bilingual Preschoolers

Jennifer A. Chard

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the First and Second Language Acquisition Commons

Recommended Citation

https://academicworks.cuny.edu/gc_etds/3050

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
MANDARIN ASSESSMENT IN CHINESE-ENGLISH BILINGUAL PRESCHOOLERS

by

JENNIFER CHARD

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2019
Mandarin Assessment in Chinese-English Bilingual Preschoolers

by

Jennifer Chard

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date ________________

Gita Martohardjono
Chair of Examining Committee

Date ________________

Gita Martohardjono
Executive Officer

Supervisory Committee:
Gita Martohardjono
Isabelle Barriere
Gisela Jia

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Mandarin Assessment in Chinese-English Bilingual Preschoolers
by
Jennifer Chard

Advisor: Gita Martohardjono

Immigrant children who grow up in linguistically and culturally diverse households are at risk for misdiagnosis for language impairment and inappropriate placement in or exclusion from special education classes. Research shows that native language testing is essential in determining eligibility for disability services, as reflected both in federal law (Individuals with Disabilities Education Improvement Act of 2004). However, despite growing agreement that native language assessment is a critical component to understanding the abilities and challenges bilingual students face, the standard assessments currently used are largely administered in Standard English and normed on monolingual English speakers. Few options are available to practitioners who work with speakers raised in multilingual households.

This investigation presents a pilot study of syntax comprehension in English and Mandarin in 24 four-year-old children who live in Chinese-speaking households in New York City. The study has two aims. One is to show how children from Mandarin-speaking homes perform on language assessments in English and Mandarin. The structures selected for this investigation are coordination and relative clauses, which are cross-linguistically robust and have been previously studied in child language acquisition research in both English and Mandarin monolinguals; as well as Chinese-specific classifiers, or nominal modifiers. Results show that the children had similar accuracy on the coordination sentences in Mandarin and English; however, for the relative clause structures, children had higher accuracy in Mandarin than English, emphasizing the need for multilingual testing. This study concludes with recommendations for a touch-screen tablet based assessment that would be useful to schools and replicable to other languages, potentially addressing the gap between policy and practice.
ACKNOWLEDGEMENTS

I am deeply grateful for the training and support I’ve received throughout my time at the CUNY Graduate Center. In addition to the crucial role each committee member played in advising the research, I appreciate the encouragement, professional support, and mentoring from each of them. Professor Gita Martohardjono has been an outstanding primary advisor. I’ve learned so much from her, including not just the science of linguistics, but the way linguistics can be used to have an impact on society and education. Gita is a fantastic role model in her dedication to the use of linguistics to advance policy, teacher education, and assessment practices so that students can truly show their skills in an educational system that was ready to discount them. I appreciate the guidance that Professor Isabelle Barriere has given me in terms of the research, professional development and teaching opportunities. Each time I met with Isabelle, her comments ranging from the benefits of tablet testing, to the need for someone to tell preschool teachers about language diversity, and even to tips on sticker management, have steered my research and teaching immeasurably. Finally, my thanks to Professor Gisela Jia. I reached out to Gisela for guidance on stimuli, and ended up with Mandarin lessons, connections so I could pilot the pilot study, and an amazing trip around China. Thank you for letting me into your world.

The dissertation is a long road, but I've been lucky to have incredible colleagues with me along the way. I began grad school with Christen N Madsen II as my student mentor. I now count Christen as my dissertation pal, bowling buddy, statistical consultant, work husband and one of my closest friends. Corbinian Neuhauser was an asset to the development of the Chinese items, and an appreciated source of syntax facts, motivation and support throughout this project. I'm grateful to Euna Cho for being my role model for work-work balance (balancing academics and other jobs), as well as the many toasts that got Euna and I through the early years of the program. Syelle Graves has been my minimalist mentor, organization guru, and a valued reality check when things get overwhelming. My thanks to Rachel Rakov, who at every stage understood exactly where my head was at, always offering solidarity and knowing exactly what I needed to hear to get past the next hurdle.
A huge thank you to the research assistants without whom I never could have managed a Chinese-centric dissertation. Qian Min (Sarah) Feng, Xiaomeng (Amy) Ma, Qi Zhang, Lianye (Juliet) Zhu, thank you so much for your energy, professionalism, and putting up with insanely early morning commutes for participant testing. An additionally huge thanks to my office mate and Linguistics APO Nishi Bissoondial for her advice, administrative support, and friendship.

I'm deeply grateful for the opportunity to work with various institutions throughout my graduate school experience. I appreciate the scientific inquiry and unshakable camaraderie I've found at the Second Language Acquisition Lab. It's been a pleasure working for the Research Institute for the Study of Language in Urban Society, the Center for Advanced Study in Education and Bridges to Academic Success, and teaching at Lehman College, Hunter College, and the Bilingual Professional Development Center at Long Island University, Brooklyn. I appreciate all the students who explained to me the situations in schools and clinics as we worked together to reconcile "best practices" with the available resources.

And finally, my gratitude to those who have made up life outside grad school: Jennifer Wiseman, Kumiko Hirama, Sondra Kelkar, Kiyoe Sato, Emily Roman, Nafisa Delgani, and Shinya Hamano. I appreciate the inspirational work of Shana Pughe-Dean and the dedicated staff of the Mohawk Valley Resource Center for Refugees. A special acknowledgement to Akemi Nasu and Mami Nishikawa, who by instruction and example encouraged me to make the most out of each moment of education I was offered. To my sister Steph who gracefully grinned and bore each wacky grad school antic, from the time I called her a prescriptivist to the time she let me paint her tongue with charcoal to practice for my first exam. Finally, to my parents Jeff and Kathy Chard for believing in me, listening to me, and encouraging me every step of the way. I couldn't have done this without you. To everyone on this page and others not listed, I appreciate your support throughout the PhD, and continued goodwill as I start the next chapter.
# Table of Contents

Table of Contents ........................................................................................................... vii
List of Figures ................................................................................................................ ix
List of Tables .................................................................................................................. x
List of Terms .................................................................................................................... xi

1. **Introduction** .............................................................................................................. 1
2. **Early Childhood Language Assessments** ............................................................. 4
   2.1 The Achievement Gap and Disproportionality ...................................................... 5
   2.2 Indicators of Language Impairment ...................................................................... 10
   2.3 Language Difference: Prestige Varieties and Undervalued Varieties .............. 12
   2.4 Language Impairment in Bilinguals .................................................................. 14
   2.5 Bilingualism, Socioeconomic Status, and Language Difference: Implications for Assessment .......................................................... 17
3. **Mandarin and English Language Development** ..................................................... 20
   3.1 Mandarin as the Language of Education in the U.S. and China ...................... 22
   3.2 Bilingual Language Acquisition ....................................................................... 26
   3.3 Child Language Development of Coordination and Relative Clauses .......... 28
      3.3.1 Coordination ............................................................................................ 28
      3.3.2 Relative Clauses ....................................................................................... 30
      3.3.3 Classifiers ............................................................................................... 33
4. **Experiment** ............................................................................................................ 35
   4.1 Research Questions ............................................................................................. 35
   4.2 Methodology ....................................................................................................... 38
      4.2.1 Materials .................................................................................................. 38
      4.2.2 Participants .............................................................................................. 41
      4.2.3 Procedure ................................................................................................. 43
5. **Results** .................................................................................................................. 45
   5.1 Mandarin Coordination ...................................................................................... 47
      5.1.1 Across group ............................................................................................ 47
      5.1.2 Home language ......................................................................................... 49
      5.1.3 Most frequent language ........................................................................... 52
      5.1.4 Summary of Mandarin Coordination Results ......................................... 54
   5.2 Mandarin Relative Clauses ................................................................................ 55
      5.2.1 Across group ............................................................................................ 55
      5.2.2 Home language ......................................................................................... 57
      5.2.3 Most frequent language ........................................................................... 60
      5.2.4 Summary of Mandarin Relative Clause Results .................................... 63
   5.3 English Coordination ......................................................................................... 64
      5.3.1 Across group ............................................................................................ 64
      5.3.2 Home language ......................................................................................... 65
      5.3.3 Most frequent language ........................................................................... 67
      5.3.4 Summary of English Coordination Results ............................................ 69
   5.4 English Relative Clauses .................................................................................... 70
      5.4.1 Across group ............................................................................................ 70
   5.5 Mandarin Classifiers .......................................................................................... 72
      5.5.1 Across group ............................................................................................ 72
      5.5.2 Home language ......................................................................................... 74
      5.5.3 Most Frequent Language ......................................................................... 75
      5.5.4 Summary of Classifier Results .................................................................. 76
   5.6 Sociolinguistic Variable Analysis ....................................................................... 77
5.6.1 Child Results for Home Language Questionnaire ................................................................. 77
5.6.2 Parent Results for HLQ ........................................................................................................... 80
5.6.3 Discussion of Child and Parent Language Results ................................................................. 81
5.6.4 Correlations of Sociolinguistic Variables ............................................................................ 83
6. Conclusion and Recommendations ............................................................................................... 87
7. Appendices .................................................................................................................................. 96
  7.1 Appendix 1: Detailed Description of Materials Presented to Participants ................................. 96
    7.1.1 English RMST ..................................................................................................................... 96
    7.1.2 Mandarin RMST ............................................................................................................... 97
    7.1.3 Mandarin Classifiers ......................................................................................................... 98
References .......................................................................................................................................... 100
List of Figures

Figure 3.1  Sinitic Varieties ................................................................. 24
Figure 4.1  RMST Image Example ......................................................... 40
Figure 4.2  Classifier Image Example .................................................... 41
Figure 5.1  Mandarin Coordination Mean Percent Accuracy Across Group ........... 48
Figure 5.2  Mandarin Coordination Mean Percent Accuracy by Home Language .......... 50
Figure 5.3  Mandarin Coordination Pairwise Comparisons Across Group .............. 51
Figure 5.4  Mandarin Coordination Significant Differences from Chance by Home Language .... 52
Figure 5.5  Mandarin Coordination Mean Percent Accuracy by Most Frequently Used Language .... 53
Figure 5.6  Mandarin Coordination Mean Percent Accuracy Across Group .................. 56
Figure 5.7  Mandarin RC Mean Percent Accuracy Across Group.......................... 58
Figure 5.8  Mandarin RC Significant Pairwise Comparisons by Home Language .......... 59
Figure 5.9  Mandarin RC Mean Percent Accuracy by Home Language .................... 61
Figure 5.10  Mandarin RC Significant Pairwise Comparisons by Most Frequent Language ...... 62
Figure 5.11  English Coordination Mean Percent Accuracy Across Group ............... 65
Figure 5.12  English Coordination Mean Percent Accuracy by Home Language .......... 66
Figure 5.13  English Coordination Significant Pairwise Comparisons Across Group ........ 67
Figure 5.14  English Coordination Mean Percent Accuracy by Most Frequently Used Language .... 68
Figure 5.15  English Coordination Significant Differences from Chance by Most Frequent Language .... 69
Figure 5.16  English RC Mean Percent Accuracy Across Group ......................... 71
Figure 5.17  Mandarin Classifier Mean Percent Accuracy Across Group .................. 73
Figure 5.18  Mandarin Classifier Mean Percent Accuracy by Home Language ............ 74
Figure 5.19  Mandarin Classifier Pairwise Comparisons Across Group .................... 75
Figure 5.20  Mandarin Classifier Mean Percent Accuracy by Most Frequently Used Language .... 76
Figure 5.21  Child Participant Profile .................................................. 78
Figure 5.22  English and Mandarin Coordination Correlation ............................ 83
Figure 5.23  Socioeconomic Status Correlation with English Accuracy .................... 84
Figure 5.24  Mother’s English Proficiency Correlation with English Accuracy ............. 85
Figure 5.25  Time in China Correlation with Mandarin Accuracy .......................... 86
List of Tables

Table 2.1 Disproportionate Enrollment in Special Education Services in New York City .............................................. 7
Table 3.1 “He gave me three books” in Mandarin and Cantonese ................................................................. 23
Table 3.2 English Coordination Types ........................................................................................................... 29
Table 3.3 Mandarin Coordination Types ........................................................................................................ 29
Table 3.4 Relative Clause Position Versus Gap Position ............................................................................. 31
Table 3.5 English Relative Clause Types ......................................................................................................... 31
Table 3.6 Mandarin Relative Clause Types ..................................................................................................... 31
Table 3.7 Mandarin Classifier Examples .................................................................................................... 33
Table 5.1 Mandarin Coordination Types ......................................................................................................... 47
Table 5.2 Mandarin Coordination Mean Percent Accuracy Across Group ......................................................... 48
Table 5.3 Mandarin Coordination Mean Percent Accuracy by Home Language ............................................. 50
Table 5.4 Mandarin Coordination Mean Percent Accuracy by Most Frequently Used Language ............... 53
Table 5.5 Mandarin Relative Clause Types ...................................................................................................... 55
Table 5.6 Mandarin RC Mean Percent Accuracy Across Group .................................................................. 56
Table 5.7 Mandarin RC Mean Percent Accuracy by Home Language ............................................................... 57
Table 5.8 Mandarin RC Mean Percent Accuracy by Most Frequently Used Language ................................. 60
Table 5.9 English Coordination Types ......................................................................................................... 64
Table 5.10 English Coordination Mean Percent Accuracy Across Group ......................................................... 64
Table 5.11 English Coordination Mean Percent Accuracy by Home Language .................................................. 66
Table 5.12 English Coordination Mean Percent Accuracy by Most Frequently Used Language ................. 67
Table 5.13 English Relative Clause Types .................................................................................................... 70
Table 5.14 English RC Mean Percent Accuracy Across Group ......................................................................... 71
Table 5.15 Mandarin Classifier Examples ....................................................................................................... 72
Table 5.16 Mandarin Classifier Mean Percent Accuracy Across Group ......................................................... 73
Table 5.17 Mandarin Classifier Mean Percent Accuracy by Home Language .................................................. 74
Table 5.18 Mandarin Classifier Mean Percent Accuracy by Most Frequently Used Language .................... 76
Table 5.19 Child Language Information ......................................................................................................... 80
Table 5.20 Socioeconomic Status Results ................................................................................................. 80
Table 5.21 Parent English Proficiency ......................................................................................................... 81
List ofAbbreviations and Terms

IP Coordination - Inflectional Phrase coordination; the conjunction of two clauses.

Mandarin – The English word for the prestige variety in China.

NP coordination – Noun Phrase coordination, including subject and object coordination.

OO RC – object-embedded, object gap relative clause.

OS RC – object-embedded, subject gap relative clause.

Putonghua – The Mandarin word for the prestige variety in China.

RC – Relative clause.

SS RC – subject-embedded, subject gap relative clause.

SO RC – subject-embedded, object gap relative clause.
1. Introduction

Research has long shown that children in the U.S. who hear a language other than English at home perform differently on language assessments than age-matched monolingual peers (Roseberry-McKibbin, 2014). Historically, this has led to the over-diagnosis of language impairments in the bilingual school-aged population, causing bilingual children to be recommended for disability services in situations where those children are typically developing bilingual children and language learning services would have been more appropriate. Recognizing the acute need to curtail the over-diagnosis of bilingual children as having language impairments, legislation was passed that requires evaluators to differentiate a language disability from a language difference due to bilingualism or exposure to an undervalued language variety (IDEA, 2004). To this end, native language testing is federally mandated when children are evaluated disabilities in U.S. schools (20 U.S.C. § 1414(b)(3)(A)). While the presence of this legislation reflects positive evolving attitudes toward bilingualism and bilingual children, especially as compared to previous times where bilingualism itself was considered a disadvantage or even a threat to cognitive function, several obstacles to implementation remain.

First, the deficit-based approach to bilingualism that has long been entrenched in U.S. policy, although discredited, remains the approach upon which the current system for working with bilingual students was based. Second, IDEA (2004) specifically states that native language testing is required unless it is not feasible for the school to provide native language testing. There is a documented shortage of access to home language materials for instruction and assessment, and also a documented shortage of certified bilingual therapists who are qualified to work with children in their home language. Although IDEA (2004) requires that standardized test scores of an English-medium assessment not be used for bilingual students, the reality is that schools and districts around the U.S. still require teachers and evaluators to report the scores of standardized English-medium assessments, and use those scores to determine service delivery to bilingual children.

Therefore, despite the gains in attitudes toward bilinguals and also the presence of policy requiring native language testing, implementation lags behind. One reason for this is the dearth of available home language assessments. The BESA is a model for an appropriate bilingual language assessment tool in English and Spanish. Phonetics, phonology, morpho-syntax, semantics, and
pragmatics are assessed in both Spanish and English, and the assessment was normed on Spanish-English bilingual children in the United States. Similar assessments need to be developed for additional languages.

According to the census, Chinese speakers are the second largest linguistic group in the U.S. (Zong & Batalova, 2017). However, children from homes where a Sinitic variety is spoken are a relatively understudied and widely misunderstood heritage speaker group. Han, Brebner, & McAllister (2016) describe some of the issues with research on and service delivery for children from Chinese-speaking families who reside in English-speaking countries. One of the biggest issues is a lack of understanding about who Chinese-speakers are and what “Chinese” is. Education professionals who work with children from these families may know of two varieties, often misleadingly called dialects: Mandarin and Cantonese. Even the practitioners who know about the Mandarin / Cantonese distinction are often surprised to hear about how different those varieties are from each other, and even more surprised to hear that there are far more varieties of Chinese present (even low estimates cite seven mutually unintelligible varieties (Ramsey, 1987). To complicate matters, Putonghua, which English speakers usually call Mandarin, is the national language of China, and is the prestige variety taught in schools, but there are many people who do not speak that variety in the home. Han, Brebner, & McAllister call not just practitioner evaluations but also research findings into question, citing cases where studies on Mandarin-speaking children include children whose parents speak another variety in the home, and studies that define their population only as Chinese-speaking children. There is therefore a need for research on the skills of children who live in Chinese-speaking homes to take the varieties of Chinese into account. Furthermore, there is a need for the development of materials that can serve the Chinese-English bilingual population since no bilingual language assessment tool is available for any variety of Chinese.

This investigation presents a pilot study of syntax comprehension in English and Mandarin for four-year-old children who live in Chinese-speaking households in New York City. The study has two aims. One is to show how children from Mandarin-speaking homes perform on language assessments in English and Mandarin. The structures selected for this investigation are coordination and relative clauses, which are cross-linguistically robust and have been previously studied in child language acquisition research in both English and Mandarin monolinguals. There are four sub-conditions for coordination and
four sub-conditions for relative clauses. Additionally, there is an assessment of classifiers, which are Chinese-specific nominal modifiers. Results show two types of information. First, results will show whether the Chinese-English bilingual four-year-olds have the same pattern of acquisition as monolingual Mandarin- and English-speaking children on the syntax assessment in English and Mandarin (coordination > subordination); (IP coordination > other coordination); (subject gap relative clauses > object gap relative clauses), and as monolingual Mandarin-speaking children on the Mandarin specific classifier task (běn (books) > zhī (animals) > tiāo (long thin objects) > zhāng (flat objects)). Second, because children from Cantonese-speaking homes also participated in this study, results are analyzed according to sociolinguistic variables. Performance on the assessment is measured by home language group (Mandarin or Cantonese) and most frequently spoken language (Mandarin, Cantonese, or English). Results of the group analyses will show whether there are significant differences in accuracy depending on what variety of Chinese is spoken in the home, or what language the child speaks most frequently.

An overview of early childhood language assessments is in Chapter 2. Chapter 3 describes coordination and relative clause structures in Mandarin and English, and summarizes the findings for child performance on these structures in monolingual populations. Chapter 4 introduces the research questions and experimental design of the study. Chapter 5 summarizes the performance on the language assessment tasks, the results of the home language questionnaire administered to the parents, and the relationships between sociolinguistic variables and performance on the syntax assessments. Chapter 6 concludes with a discussion of the findings and recommendations for future work on Chinese-English bilingual assessments.
2. Early Childhood Language Assessments

Standardized language assessment tools are used in early childhood education settings to determine whether students are in need of and are eligible for special education services. The assessments are based on a set of milestones that speakers from similar backgrounds are expected to meet at around the same time as their age-matched peers. If a child’s performance is significantly below average, an appropriate intervention, such as working with a Speech-Language Pathologist, will be offered with the objective of bringing that child’s language skills up to the same level as the child’s age-matched peers. In order for this system to work, there are two requirements. First, that the assessment tool itself must be non-biased; it cannot require cultural or linguistic knowledge that the child has not been exposed to. Second, that the “average score” that a given child’s score is compared to must be based on the scores of age-matched peers from similar backgrounds, but often these scores do not exist. These requirements are mandated by federal law in the Individuals with Disabilities Education Improvement Act of 2004 (IDEA, 2004). IDEA requires that evaluation materials must be valid, reliable, and without linguistic or cultural bias (mandating native language testing “unless it is not feasible to so provide or administer”), and that the evaluation must be able to distinguish between a disability, limited English proficiency, and lack of adequate instruction in reading and math (20 U.S.C. § 1414(b)(3)(A)).

Federal, New York State, and New York City guidelines are in place to promote appropriate assessment practices, including native language assessment (IDEA, 2004; NYSED, 2010; NYCDOE, 2014). However, despite these efforts, over-diagnosis, under-diagnosis, and misdiagnosis of language impairments specifically – and disabilities generally – are pervasive in the U.S. (Roseberry-McKibbin, 2014; Morgan et al. 2015). Unfortunately, it is not currently feasible to provide native language testing to most multilingual students. There is a lack of multilingual assessment tools, and instructional resources that education professionals can use to serve multilingual students are limited. There is also an unfulfilled need for education professionals to participate in professional development to learn how to assess culturally and linguistically diverse students, including multilingual students.

This chapter presents an analysis of the key factors related to childhood language assessments and reinforces the need for appropriate language assessment practices. Section 2.1 describes the disparity that exists between the ideals for disability assessment and special education services
contained in the Federal, State, and City guidelines. Section 2.2 introduces the characterizations of a language impairment and the indicators that are used for assessing Standard American English. Since language impairment is characterized by atypical functioning of the mental grammar, it is important that the starting point for any child language assessment is the actual variety the child uses, based on the language input the child is exposed to. If a child’s undervalued language system is compared to the prestige language that is the default for education, then the assessment tool making the comparison is not valid for that child. This holds true whether the child uses an undervalued variety of English (as described in Section 2.3) or a different language (as described in Section 2.4). Section 2.5 explores the interactions between these factors, motivating the need for appropriate language assessment practices for students from culturally and linguistically diverse background.

2.1 The Achievement Gap and Disproportionality

There is a disparity of educational achievement based on race/ethnicity, socioeconomic status, and language background. This well-known phenomenon is often described in terms of disproportionality and the achievement gap. Researching the correlations between socioeconomic status/racial background and academic achievement – and counteracting these trends, have been national priorities since these critical factors were documented in the publication of the 1966 government report, *Equality of Educational Opportunity*, (commonly referred to as the “Coleman Report”: Coleman et al., 1966).

*Disproportionality* refers to the over- or under-representation of a particular demographic group in special education programs relative to the general population (National Association for Bilingual Education, 2002; National Education Association, 2007). Specifically, disproportionality describes the reality that students from (for example) African-American, Hispanic, bilingual, and lower socioeconomic backgrounds are over-represented in special education programs but under-represented in gifted and talented programs across the country, while Asian students are under-represented in special education programs. Both Caucasian and Asian students are over-represented in gifted and talented programs. According to the National Center for Education Statistics (2017), 13% of public school students receive special education services nationally.
In New York City, disproportionality is a reality that needs to be urgently addressed. The publicly available New York City Department of Education [NYCDOE] Demographic Snapshot reports that students with a disability constituted 19.6% of the public school population in the 2015-2016 school year (NYCDOE, 2017). By cross-referencing the total enrollment number reported on the Demographic Snapshot with the Annual Report on Special Education from the same year (NYCDOE, 2016), it becomes evident that New York City has a higher proportion of students receiving special education services (16.8%) than the national average (13%). According to the Annual Report on Special Education, 190,412 students received special education services in New York City Public Schools in the 2015-2016 school year. Using the percentage of students reported for each racial/ethnic group in the Demographic Snapshot, the expected enrollment of each racial/ethnic group in special education services was calculated and compared to the actual enrollment of each racial/ethnic group reported in the Annual Report on Special Education. The percentage of students receiving special education services by ethnic group does not match the proportion of each ethnic group in the general student population, as shown in Table 2.1. In the Demographic Snapshot, 40.5% of students are Hispanic and 27.1% are African American. If each ethnicity were proportionally represented among the group of students who receive special education services in New York City, 77,117 out of the 190,412 students would be Hispanic and 51,601 students would be African American. However, as per the Annual Report on Special Education, 93,137 students receiving special education services are Hispanic (48.9%) and 56,129 are African American (29.5%). Both groups are over-represented in special education. Comparatively, 29,514 of the 190,412 students would be Asian (15.5% of the student population) if each race was proportionally represented, but the number of Asian students receiving services is just 12,422 (6.5%), a substantial under-representation. Caucasian students comprise 15.8% of the student population, so 28,181 would be expected to receive special education services, and the number of students receiving services is 25,233 (13.3%). Additional, students who receive services for English Language Learners (ELLs) are also...

1 Note that while the number of Caucasian students receiving special education services is lower than would be expected based on demographic representation alone, the number of students receiving special education services in New York City is higher than the national average. However, the number of students listed on the Demographic Snapshot as living in poverty is 76.1%, and students from lower socio-economic status backgrounds are also likely to be over-represented in the population receiving special education services. Weighing these factors against each other is outside of the scope of this paper.
over-represented in special education. 23,992 represent the proportional percentage of ELLs in the general student population (12.6%), while 35,256 ELLs actually receive special education services (18.5%).

Table 2.1 Disproportionate Enrollment in Special Education Services in New York City

<table>
<thead>
<tr>
<th>Ethnicity / ELL Status</th>
<th>% of NYC Public School Population</th>
<th>Predicted Proportional Enrollment in Special Education Services</th>
<th>Actual Enrollment in Special Education Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>40.5%</td>
<td>77,117</td>
<td>93,137 (48.9%)</td>
</tr>
<tr>
<td>African-American</td>
<td>27.1%</td>
<td>51,601</td>
<td>56,129 (29.5%)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>15.8%</td>
<td>28,181</td>
<td>25,233 (13.3%)</td>
</tr>
<tr>
<td>Asian</td>
<td>15.5%</td>
<td>29,514</td>
<td>12,422 (6.5%)</td>
</tr>
<tr>
<td>ELLs</td>
<td>12.6%</td>
<td>23,992</td>
<td>35,256 (18.5%)</td>
</tr>
</tbody>
</table>

The over-representation of Hispanic, African American, and ELL students in special education contributes to the achievement gap documented for these groups. This achievement gap refers to the disparity in measures of educational performance among students based on inclusion in specific ethnic groups, lower socioeconomic status, and/or linguistically and culturally diverse backgrounds, including multilingualism. This gap is substantiated by scores on standardized math and literacy assessments, and high school graduation rates (National Education Association, 2017).

The underlying causes of disproportionality and the achievement gap overlap. Overlapping elements include the deficit-based approach to education for children from minority groups, institutionalized and unconscious bias, the magnifying effect low socioeconomic status has on educational performance in terms of student nutrition, readiness to learn, etc. However, the over-representation of African Americans, Hispanics, and bilinguals in special education programs exacerbates the achievement gap because students with disabilities are at risk for being held to lower academic standards and exposed to a less rigorous curriculum, in addition to contributing negative affective factors such as decreased motivation, negative self-perception, and the negative influence of stigma from peers and teachers (U.S. Department of Health and Human Services & U.S. Department of Education, 2017).

---

2 Percent of New York City Public School Population is from the Demographic Snapshot Report for 2015-2016 school year (NYCDOE, 2017). Actual Enrollment in Special Education Services is from the Annual Report on Special Education for 2015-2016 school year, based on a total special education enrollment of 190,412 students, with calculated percentages (NYCDOE, 2016). Predicted Proportional Enrollment in Special Education Services is calculated as the percentage reported in the second column out of the 190,412 students receiving special education services.
The designation of culturally and linguistically diverse students includes not only multilingual students from immigrant families, but also students who are monolingual English speakers, but whose language and interaction practices do not reflect the middle-class American communicative styles upon which the benchmarks in U.S. educational settings are based (California Department of Education, 2009). One example of an interaction difference among students from a monolingual English-speaking background is the practice of teachers and parents asking “known-answer questions” (questions in which the answer is obvious to the adult, and are used as a tool to facilitate interaction and language development in the communities that use them). This questioning style is common among adult-child interactions in middle class families and schools. For example, if an adult were to ask a child “what color is this toy?”, a child who did not grow up in a “known-answer question” culture would be unsettled by the inquiry, since in the home culture, the child would not be asked a question if the answer was obvious to an adult. In the case where a correct answer is offered to a “known-answer question”, a teacher, in noting the child’s hesitation, delay before answering, and discomfort with the question, may believe it to be because the child does not know the answer rather than that the child is reacting cautiously to what may seem like a trick question. As a result, the child may be referred for an evaluation to see if there is a language delay or other disability. A question that the adult did not necessarily know the answer to (such as “which is your favorite toy?”, possibly followed by “Can you tell me about it?”) would be a more appropriate question for a child from this background, thereby reducing the possibility of over-diagnosis based on misunderstanding of a cultural practice. This example is based on Heath’s 1983 examination of the linguistic, cultural, and interaction practices of a lower socioeconomic status community in South Carolina. Heath found that although the language skills of children in that community developed at the same rate as those of middle class children, the trajectory and resulting communicative style were different.

This example of a mismatch between the norms of a child’s cultural community and the milestones and norms that public school standards are based on calls into question whether the assessments in place in early childhood programs are serving their intended purpose. Quantifiable assessment tools at their best can give a teacher who is unfamiliar with a child’s cultural and linguistic background confidence that the child is performing appropriately in comparison with peers from a similar
background. However, if the expectations and curricula of early education programs align with middle class linguistic and cultural practices, and not with culturally and linguistically diverse community practices, students in the latter group will be at a competitive disadvantage. Linguistically diverse students are held to standards deriving from conversational rules and cultural practices that students from middle class, monolingual English-speaking backgrounds experience on a daily basis. Given their more limited exposure to these standards of practice, it would be surprising if children from culturally and linguistically diverse backgrounds did perform similarly to monolingual English speaking students. As students are held to standards and benchmarks based on cultural and linguistic practices that they have not received sufficient input to be able to master, the education of culturally and linguistically diverse students is based on a deficit perspective. This means that educators view a child’s knowledge in terms of the standard variety skills, knowledge, and practices the child does not possess, rather than the skills and knowledge the child does possess, as compared to peers from a similar background.

The deficit-based approach to education, which implicitly targets cultural and linguistic diversity as an obstacle that must be overcome in order to participate with the majority instead of merely a different variety, actively contributes to the achievement gap in two ways. First, by holding all children only to the linguistic and cultural practices that are emphasized in mainstream education, and disregarding the abilities children possess that are valued in the home culture, a representative description of the child’s abilities is not possible. Second, framing a child’s education in terms of “needing to catch up” to students who are familiar with a larger number of school-based practices because they have experienced those practices in the home throughout their childhood, puts the child from a non-mainstream culture at a disadvantage by default. Too often the children from these environments are seen as needing special help to catch up in order to meet the standards that other students meet naturally, without the understanding that the differences are due to factors of environmental exposure. Intentional exposure to target cultural and interaction practices are appropriate forms of support for these children, but must be distinct from special education services, as required by IDEA, 2004. As early as preschool, children who do not interact in a format following school standards based on mainstream middle-class cultural practices may be held apart from their peers and may even be labeled delayed or disabled. The more appropriate designation would be typically developing, though along a different trajectory. This begins a
downward spiral. A student mislabeled as needing special education services may begin to have a negative self-perception. Similarly, the adults who are aware of the special education designation may hold the child to lower standards, believing that the child is less capable of learning than their peers. Over time, these affective variables can have negative consequences for learning and academic achievement. However, the reality is that typically developing children from culturally and linguistically diverse backgrounds will acquire the target skills and practices once they receive sufficient exposure and have the opportunity to practice the skill at school. They should not be measured against other children who received more exposure to a given skill and had the opportunity to practice the skill in the home environment.

This is why appropriate assessments that are tailored to culturally and linguistically diverse groups matter. Mainstream assessments differentiate students who are typically developing along an expected trajectory from students who are developing along other mainstream trajectories that may be equally valid. The unavailability of assessments tailored to specific culturally and linguistically diverse groups makes it difficult to tease apart whether a student is following a different trajectory or is actually developmentally atypical. This section has described general issues with administering standard assessments to children early childhood education settings and the detrimental effects those results can have on students. The next sections narrow the focus to assessments of language impairment.

2.2 Indicators of Language Impairment

Specific Language Impairment (SLI – also called Primary Language Impairment) is a developmental disorder, defined by Leonard (1998) as a significant limitation in language ability in the absence of hearing damage, neurological damage, intellectual disability. SLI is estimated to affect about 7% of the population (Tomblin et al., 1997). Individuals with SLI are a heterogeneous population; even within groups from similar demographics, there is a range of variation in both linguistic and nonlinguistic characteristics.

SLI has traditionally been thought of as a grammatical language impairment since function words and inflectional morphemes are the most vulnerable language properties affected. The American Speech-Language-Hearing Association [ASHA] (2017) identifies several areas of difficulty that constitute a
language impairment. One common indicator is in the realm of functional, rule-based (grammatical) morpho-syntax, including regular and irregular inflectional verb and noun morphology, along with auxiliary verbs, articles, prepositions and pronouns. Leonard (1998) hypothesizes that the abstract principle of inflection has not been internalized in individuals with SLI and proposes that SLI results not only from difficulty identifying and using the appropriate morphology, but also from an underlying lack of knowledge that morphological marking is obligatory. Language impairment is not manifested the same way in all individuals, but is characterized by disordered receptive, expressive, and meta-linguistic skills in the areas of grammar, vocabulary, and/or pragmatics, as described below.

In terms of receptive skills, SLI is characterized by difficulty understanding complex sentences and difficulty distinguishing between morphological forms. Children with a language impairment who find it difficult to understand sentences derived by movement may exhibit misunderstanding or delayed understanding of relative clauses and wh-questions in comparison to typically developing age-matched peers (Friedmann & Novogrodsky, 2011). These comprehension difficulties can interfere with understanding instructions, especially when several complex sentences are used in a row. Children with SLI may also exhibit reduced phonological awareness, and may be unable to use morphological information to identify differences in meaning. Morphemes that are less phonetically salient such as English plural and possessive –s, and past tense –ed, tend to be particularly affected (Leonard, 1998). As a result of underlying receptive difficulties, children with SLI are especially likely to present as dyslexic as they begin learning how to read and write. The effects can last into adulthood in the cases where the individual does not receive intervention services for a language impairment (Poll, Betz, & Miller, 2010).

Expressive grammatical and lexical skills are also affected by SLI. Leonard (1998) notes that expressive errors of omission are the most common grammatical error type. Errors of omission are those that leave out a functional morpheme (e.g. “She run” instead of “She runs”). Omission and inconsistent morpheme use occur more frequently than errors of commission (where an incorrect morpheme is present in the correct position, e.g. “They is happy” rather than “They are happy”) or insertion (adding a superfluous morpheme, e.g. “I sleeps” instead of “I sleep”). Individuals with a grammatical language impairment make errors in complex sentences such as omission of the obligatory relative marker in sentences that contain relative clauses; and furthermore, they tend to use fewer complex sentences
overall as they avoid constructions that are especially error-prone (Schuele & Tolbert, 2001; Novogrodsy & Friedmann, 2006). Lexical issues such as delayed onset of first words, word retrieval difficulties, and a smaller vocabulary with fewer expressive words, may occur in addition to or instead of grammatical errors. There may also be a delay in the onset of multi-length utterances and a shorter average mean length of utterance than typically developing age-matched peers (ASHA, 2017; Armon-Lotem & de Jong, 2015).

Meta-linguistic skills are also affected by SLI. A language impairment makes it difficult for an individual to judge grammaticality, identify and correct grammatical errors, and express knowledge about language such as describing how to manipulate the structure of language for different purposes. Understanding and explaining ambiguities can also be difficult for children and adults with SLI (ASHA, 2017; Poll, Betz, & Miller, 2010).

In Standard American English, the language properties typically assessed for the purpose of identifying a language impairment are present in standardized assessments such as Rice & Wexler’s (2001) Test of Early Grammatical Impairment (TEGI). The language properties examined in this assessment are: third person singular –s, past tense –ed, irregular past tense, “to be” as a copula and as an auxiliary, “do” as an auxiliary, progressive –ing, prepositions “in” and “on”, plural -s, and articles “a” and “the”. The examiner’s manual of the assessment notes that results are only valid for children who speak Standard American English, the population the assessment was created for and normed on.

Assessment norming aims to ensure that the test is valid both in terms of sensitivity (so that children who have an impairment are correctly identified) and specificity (so that children who do not have an impairment are not identified as having an impairment). However, when the population the assessment is normed on is different from a child who is taking the assessment, as will be explored in the next section, the validity of the results must be called into question for those students who grew up in a different linguistic environment from the students the assessment was normed on.

2.3 Language Difference: Prestige Varieties and Undervalued Varieties

While the Standard American English (SAE) assessment described above is appropriate for children who receive most of their language input in that variety, that assessment will not be valid for
children who speak an undervalued variety of English, such as African American English (AAE) or Appalachian English. Historically, speakers of AAE were diagnosed as language impaired because they were considered non-proficient by SAE standards; however, legislation and clinical guidelines for best practices now state that using constructions from an undervalued variety is not the same as having a language impairment, and that clinicians must rule out misdiagnosis due to an undervalued variety before assigning a language impairment diagnosis (IDEA, 2004; Roseberry-McKibbin, 2014). Speaking an undervalued variety is now termed a language difference instead of a language impairment. At the same time, however, children who speak undervalued varieties are not immune to SLI and clinicians should not assume that all patterns diverging from SAE are due to a language difference (Oetting, McDonald, Seidel, & Hegarty, 2016)

The following are some typical characteristics of AAE. These features are systematic within the variety, but are different from SAE, and so must be taken into account when assessing children from an AAE-speaking background. Phonological phenomena include fortition of interdental fricatives to alveolar stops, final consonant deletion, and consonant cluster reduction. Morphological markers in redundant contexts may be absent, such as the plural –s when the noun is modified by an overt number marker (“She has 3 boy” is grammatical without the plural marker because the number is overtly stated. Compare that to the sentence “She has only boys”, where the number is not overtly stated and the plural –s is obligatory). Copula use is another prominent difference, with zero copula and overt copula forms being grammatical in AAE. (Zero copula constructions like “He tired” and “He working” describe a state or action in the present moment. Overt unconjugated copula constructions, also called “habitual be” constructions, like “He be tired” and “He be working” describe a habitual state or action. SAE speakers unfamiliar with the grammar of AAE often believe both of these structures to be undistinguished erroneous manifestations of the sentences “He is tired” and “He is working”) (Payne, 2005).

The differences between SAE and AAE in phonology, morphology, and syntax often mean that a child who receives primarily AAE input would likely to do poorly on assessments developed for and normed on SAE-speaking children whether or not the AAE-speaking child has a language impairment. This situation led to widespread over-diagnosis of language impairment among African-American students in the U.S. Students inappropriately receiving Speech-Language Pathology services due to a
language difference were then “corrected” to SAE structures by clinicians (Payne, 2005). ASHA now instructs clinicians to acknowledge undervalued varieties as inherently valuable and systematic, and not to coach clients to replace AAE structures with SAE structures. This is a vast improvement in policy, though it does not always pan out in practice, and clinicians working with language-impaired AAE speakers typically target SAE structures (ASHA, 2017).

Current clinical opinion is that most language assessments are biased toward SAE speakers and that the results of SAE assessments are not valid for speakers of other varieties of English (Seymour, Roper, & de Villiers, 2005; Roseberry-McKibbin, 2014, and more). Further, as described above, the IDEA (2004) legislation explicitly states that a language difference (i.e. speaking a variety other than SAE) is distinctly different from a language disability, and that speaking such a variety is not cause for remediation. This leaves clinicians in the position where they know that the results of a SAE assessment are invalid, but this knowledge alone does not give clinicians the tools to appropriately assess speakers of undervalued varieties. Therefore, a language assessment that takes into consideration the grammatical properties of AAE, and also account for possible code-switching between SAE and AAE, is necessary.

The Diagnostic Evaluation of Language Variation (DELV: Seymour et al., 2005) is a norm-referenced language assessment that is tailored to speakers of undervalued varieties of American English, and specifically AAE. There is one set of questions that allows answers from both SAE and AAE to be marked correct, and the assessment was both developed for speakers of undervalued varieties, and also normed on children from backgrounds similar to the children who the assessment is intended to serve. In the scoring system, a child’s response may fall into one of three categories: the response could be typical of SAE, typical of AAE, or not typical of either variety. The DELV provides a model for how to assess a prestige language variety without putting speakers of an undervalued variety at a disadvantage, and this model will be explored in the context of bilingualism in the next section.

2.4 Language Impairment in Bilinguals

Bilinguals are individuals who regularly use two or more languages. Bilinguals are often categorized based on relative language proficiency level (balanced bilingual vs. dominant bilingual) and age of acquisition (simultaneous bilingual vs. sequential bilingual). The population investigated in this
dissertation are Dual Language Learners (DLLs) in the U.S., defined by Park, O’Toole, and Katsiaficas (2017) as “children ages eight and under with at least one parent who speaks a language other than English at home”, who are attending early childhood programs or schools where they are learning English. In educational contexts, the term DLL is differentiated from ELL (English Language Learner) in that ELLs are students who begin learning English at age 9 or later, and who therefore have a much stronger foundation of language skills in the home language than in English at the outset of learning English. DLLs can be simultaneous bilinguals or sequential bilinguals, beginning a second language in school at three, four, or five years of age. DLLs can be dominant in some conversational domains such as household routines in one language, and dominant in other conversational domains like school language in their other language.

As students age out of the DLL designation, many of them will eventually fall under the Heritage Speaker label, indicating that they spoke a non-English language regularly at home as a small child, but have used English so often as a school language that they may feel more comfortable communicating about a wide variety of topics in English. Heritage speakers may maintain highly proficient home language skills or may retain only limited understanding skills in the home language. Heritage speakers have typically received negative attention in school, as well as in the bulk of academic research, where the main arguments debate whether their “problem” is incomplete first language acquisition or first language attrition (Montrul, 2008). However, following Cook’s theory of multicompetence (Cook, 1991), rather than investigating heritage speakers from the deficit perspective of comparing a bilingual individual to two monolingual individuals, it makes more sense to describe heritage speakers in terms of their own heritage speaker grammar. The heritage speaker grammar is different from a monolingual grammar in systematic and predictable ways, which does not indicate an impairment or acquisition deficiency (NYSED, 2010; Roseberry-McKibbin, 2014).

The multilingual grammar is shaped by language input and domains of language use, as well as language-external factors such as immigration status, cultural proximity to the majority, and more (Peña, Bedore, & Sheng, 2013; Fishman & Peyton, 2001). Language input is a key factor in determining acquisition and performance in each language (Hart & Risley, 1995; Fernald, Marchman, & Weisleder, 2013). If a child is consistently exposed to rich, interactive input in two or more languages, the underlying
structures of each language will be acquired, though it is not necessarily the case a bilingual child will acquire two languages in the amount of time it takes a monolingual child to acquire one language. Even a child who receives a substantial amount of overall input does not receive the same amount of input in each language as a monolingual peer from a similar background (Hoff & Core, 2013). Therefore, in addition to the overall amount of input a child receives, language performance and dominance are influenced by how much input a child receives in each language. Early childhood education experts also suggest that the quality of input (defined as interactive, meaningful, diverse language communication with a child) is equally important (California Department of Education, 2009), though the designation of high quality input must not be confused with prestige variety input – in English or in the home language.

The challenge facing practitioners who work with bilingual children is differentiating a typically developing bilingual child from a child with a language impairment who is exposed to more than one language. The ideal comparison group consists of peers from a similar socioeconomic status background, who speak the same language varieties and receive a similar amount and quality of overall input. However, even bilingual children from similar backgrounds may have different language skills from each other because the total input in each variety may be different. A bilingual child who is assessed in English often exhibits errors similar to those of a monolingual child with a language impairment, including difficulty producing and comprehending complex sentences, and inconsistent usage or omission of possession markers, plurals, past tense, copulas, auxiliaries, pronouns, articles, and more. In addition, a bilingual child may use a non-canonical word order and/or switch between languages. None of these factors, taken alone, points to a language impairment in a bilingual (Roseberry-McKibbin, 2014). In order to appropriately assess a bilingual child, the relative amount of input and use for each language must be considered (usually measured with a home language questionnaire administered to parents), as well as the differences between English and the home language. For example, a Spanish-English bilingual child may regularly produce plural –s in English, but may struggle with possessive –s, favoring either the Spanish word order or simply omitting the possessive morpheme. If the child is dominant in Spanish, this pattern would be expected, because while plural markers in Spanish and English overlap, possession is expressed with different grammatical constructions. Furthermore, if the child is appropriately able to use possessive constructions in Spanish, then that child does not have a problem with the grammatical
expression of possession. Rather, the child needs more exposure and practice with the English expression of possession. Similarly, children from Russian-speaking households tend to struggle with English articles, children from Chinese-speaking households tend to struggle with English pronouns and plurals, and children from Arabic-speaking households tend to struggle with English word order. Such issues are common for bilinguals from these backgrounds because of the influence of the home language on English, and do not indicate a language impairment. Indicators of an impairment must be present in both languages in order to be deemed representative of a language disability, and not just a language difference. For example, if a child does not mark tense in either language, is not able to understand complex sentences in either language, and/or exhibits shorter utterances and a more limited vocabulary than peers from a similar background in both languages, that could indicate a language impairment (Peña, Bedore, & Sheng, 2013; NYSED, 2010). The heterogeneity of bilinguals as a group has implications for language assessments developed for bilinguals, as discussed in the next section.

2.5 Bilingualism, Socioeconomic Status, and Language Difference: Implications for Assessment

Unfortunately, there is a dearth of assessments developed for and normed on bilinguals, which makes implementation of the ideals expressed in IDEA (2004) difficult to achieve. U.S. clinicians cite this as a major problem in working with clients (Kimble, 2013), and this problem is known to policy makers (U.S. Department of Health and Human Services & U.S. Department of Education, 2017). In order for education professionals to be able to execute linguistically and culturally responsive assessment practices, they need to have the training and the tools that will allow them to do so\(^3\). There is a disparity between the widely-agreed-upon ideals of bilingual assessment practices and the reality. This section summarizes considerations for the development of linguistically and culturally appropriate bilingual

---

\(^3\) One way of serving culturally and linguistically diverse students from monolingual and bilingual backgrounds is to implement the dynamic assessment approach described by Crowley’s (2015) LEADERS Project. Dynamic Assessment measures a child’s ability to learn, rather than static knowledge. An example of dynamic assessment includes a task that measures whether a child can identify and then name two objects given nonsense labels by an examiner. If the child does not correctly point to the object or name the object at the first prompt, the examiner repeats the name of the object and later asks the child again. The benefit of this type of assessment is that it can be done even if a practitioner is not able to work in a child’s home language. IDEA (2004) requires native language testing whenever possible, and the role of this dissertation is to explore best practices for appropriate native language testing. While an excellent accompaniment to static assessments, exploration of dynamic assessment practices is outside the scope of this dissertation.
assessments presented in this chapter, and examines one of the few existing assessments developed for and normed on a bilingual population.

Assessments for bilingual children must measure relevant language-specific skills in two different language varieties, and must be free of bias based on culture, socioeconomic status, or language variety in either language. First, the criteria for language-specific skills for each variety must be used, since bilingual children tend to exhibit observable traits similar to monolingual children with SLI, even when a language impairment is not present. Second, monolingual standardized assessments tend to be developed in the prestige variety of a language, and a bilingual child may speak undervalued varieties of both the home language and also English. A language assessment that does not take undervalued varieties of both languages into account will not be valid for children who speak an undervalued variety of even one language, since appropriate assessment in both languages is required to indicate that a language disability, and not just a language difference, is present. Third, the assessment must be normed on bilingual children from a cultural and socioeconomic background similar to the bilinguals who the assessment is meant to serve. The resulting tool must be used by a practitioner who is trained in working with culturally and linguistically diverse children and able to distinguish between a language disability and differential performance caused by a language difference or limited exposure to the item(s) being assessed. Practitioners must remember that if the population that an assessment was developed for or normed on is different from the bilingual child who is being evaluated, the score derived from that assessment will not be valid for that child.

There is an example of a bilingual language assessment that was developed for and normed on bilingual English-Spanish speakers in the U.S. (The Bilingual English-Spanish Assessment [BESA]: Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). The repertoire of skills measured in this assessment include phonology, morpho-syntay, and semantics in English and in Spanish. The tool includes questionnaires that measure home language use and document parent and teacher concerns about the student’s language ability. The English version of the morpho-syntay part of the test examines grammatical structures that are known indicators of language impairment in English, such as plural –s, past tense –ed, copula constructions, auxiliaries, and more. The Spanish morpho-syntay section examines grammatical structures that are known indicators of language impairment in Spanish, such as
articles (gender), clitics, subjunctives, and complex verb conjugations. Efforts were made to control for language variation in both English and Spanish, though speakers of undervalued varieties such as AAE are identified incorrectly at higher rates than speakers of SAE. A bilingual clinician evaluates a child entirely in one language and then the other, and the child is scored separately for each language. If a child’s scores fall within the average distribution in one or both languages, the child is considered typically developing and not language impaired. If the score is outside the typically developing range in both languages, then the child may have a language impairment (or may speak a variety that the test was not developed for). The BESA meets most of the criteria laid out above, except for the requirement that the assessment be valid for children who speak one or more undervalued varieties. The next chapter describes a proposed approach to development of appropriate bilingual assessments, with a focus on Mandarin.
3. Mandarin and English Language Development

Chapter 2 motivated the pressing need for appropriate assessments free of cultural or linguistic bias for bilinguals, who may speak undervalued varieties of both of their languages and who may also be speakers from lower socioeconomic status backgrounds. In order to comply with federal, state, and city standards, as well as clinical best practices, the development of native language assessment tools must be a priority. This chapter describes the developmentally appropriate structures selected for a native language assessment tool in Mandarin. The reasons Mandarin was selected as the focus of this dissertation are explained below.

“Chinese-speakers” in the U.S are a large but ill-defined group. The U.S. census cites “Chinese” as the second most widely spoken language other than English in the U.S. (Zong & Batalova, 2017). 13.39% of ELLs in New York City speak Chinese, which is the second largest ELL population after Spanish speakers. However, “Chinese” can refer to all of the many varieties in the Sinitic language family. There are at least seven major mutually unintelligible varieties of spoken “Chinese” (Ramsey, 1987), with higher estimates reaching around 275 varieties (Simons & Fennig, 2018). (See Section 3.1 for more detail.) Despite the wide variation of languages that fall under the label of “Chinese”, the population of “Chinese-speaking ELLs” is not clearly defined by the NYCDOE. The home language questionnaire used by the NYCDOE and New York State Education Department (NYSED) collects home language information as an open-ended question (NYSED, 2014), which leads to answers like “Chinese” instead of a specific variety. Within the current circumstances, the languages of children whose parents speak “Chinese” cannot even be divided into the two mutually unintelligible varieties that are the most common in New York City: Mandarin and Cantonese. In some New York City classrooms, teachers don’t know that there are different varieties of Chinese, or which variety/varieties their students speak. This confusion is not limited to New York City schools; Han, Brebner, and McAllister (2016) make the case that the lack of understanding about Chinese languages has led to misdiagnoses of language impairments and has had negative repercussions for research. Because the terminology is problematic but also widespread, the term “Chinese” or “Chinese-speaker” will be used in this dissertation to refer to the Sinitic language family or speakers of those varieties, respectively. This labeling is necessary where the variety or varieties are not recorded in the original source material, such as numbers reported by the census and the NYCDOE.
The lack of understanding about the profile of ELLs who speak Chinese leads to the second point: there is a dearth of assessment tools to support students from homes where a Sinitic variety is spoken. Assessments of language skills are not available in preschools for most languages, including Mandarin. Kimble (2013) reports that speech-language pathologists working in English-dominant contexts widely report not having access to language assessments tailored to languages other than English, with the exception of some Spanish language resources. McLeod et al.’s (2014) “Review of 30 Speech Assessments in 19 Languages other than English” (which includes only assessments of phonetics and phonology) does not include Mandarin materials. The review does include two Cantonese assessments of speech sounds developed for monolingual children in Hong Kong, but noted that any assessment developed for monolinguals must be validated on bilinguals before bilingual children can be held to the standards of these assessments. There are no assessments that have been developed for and normed on Mandarin-English bilinguals or Cantonese-English bilinguals.

As described above, the problem facing professionals who work with children from Chinese-speaking homes is two-fold. Language assessments designed for these children are not available to the professionals who need them. Furthermore, care must be taken when designating a language assessment as a tool for Chinese-speakers since there is such a large number of distinct, mutually unintelligible varieties that fall into this category and a widespread misunderstanding about Sinitic varieties among practitioners. As a result of the lack of understanding about the home languages of Chinese-English bilinguals, and the absence of valid language assessment tools, there is a dire need to improve how Chinese-English bilinguals are diagnosed for disabilities and how those services are administered. Chinese-speaking students who attend preschools where there are few other Chinese speakers may be under-diagnosed for language impairments and other disabilities that are present because of the pervasive stereotype of Asians as the “model minority”, as well as the stigma attached to disabilities in the Chinese community (Roseberry-McKibbin, 2014). This generalization points to an under-representation of Asian students receiving special education services that pans out in New York City schools, as shown in Table 2.1 Disproportionate Enrollment in Special Education Services in New York City (Section 2.1; page 7). Asian students represented 15.5% of the student population during the 2015-16 school year, (NYCDOE, 2017), but only 6.5% of students receiving special education services during
that same period (NYCDOE, 2016). Data for how many Chinese-speaking ELLs receive disability services in New York City is not publicly available, but it is the case that ELLs generally are over-represented in special education. ELLs represented 12.6% of the student population during the 2015-16 school year, (NYCDOE, 2017), but 18.5% of students receiving special education services during that same period (NYCDOE, 2016). Although the statement that Asian students are under-represented in special education services generally but over represented as ELLs seems counterintuitive, Chinese-speaking children may receive speech therapy services because their English is not as developed as monolingual English-speaking peers. That means that for the Asian students who are ELLs, there is a possibility of misdiagnosis. The need for appropriate bilingual assessments developed for the Chinese-English population is urgent.

In summary, the motivation behind this dissertation is the pressing need for assessment materials to serve the large Chinese-speaking student population in New York City, despite the group of Chinese-speakers being relatively ill-defined. A brief overview of the Chinese languages and a description of the status of Mandarin in the People’s Republic of China (P.R.C) and New York City are given in Section 3.1. Section 3.3 outlines considerations that must be made for bilingual language development, commonly reported as language milestones, and describes language structures that are appropriate for use on language assessments of English and Mandarin. These structure will be the focus of the experimental component of this investigation.

3.1 Mandarin as the Language of Education in the U.S. and China

*Mandarin* is the term used in English to the national language of the P.R.C. Speakers of the language call it *Putonghua*, or the “common language”. For readability purposes, this dissertation will use the term *Putonghua* to refer to the standardized national language of the P.R.C. and *Mandarin* to refer to the spoken variety used widely in the northern P.R.C., which shares many similarities with the prescriptive

---

4 In addition to the presence of the theoretical possibility, parents of children receiving speech pathology services and preschool teachers who work with Chinese students receiving disability services in New York City have stated that their impression that this is the case.

5 When referring to policies and figures for a specific region, that region will be named (i.e. the People’s Republic of China, (P.R.C.) and the Republic of China (R.O.C.)). In general, the term “China” will be used to refer in general to areas where Sinitic varieties are the majority languages.
Putonghua. However, the colloquial varieties of Mandarin spoken in the northern P.R.C. have some features that are different from the prescriptive prestige variety. For example, the variety of Mandarin spoken in Beijing has political prestige, though it is not a variety that sounds like an unaccented or default version of Putonghua. The variety has a strong rhoticization feature not used elsewhere. Putonghua is a prescriptive construct, developed by politicians so that the P.R.C. would have a national language. Most people in the southern P.R.C. use a regional variety that is very different from Putonghua at home and in the community, and are taught Putonghua in schools (Ramsey, 1987).

The number of varieties spoken in China differs depending on how variety is defined: from a political standpoint, Putonghua is the official language of the P.R.C. and all other varieties are considered even by native speakers to be dialects of Chinese, despite mutual unintelligibility. Ramsey (1987) lists seven major mutually unintelligible varieties of Chinese, while Ethnologue (Simons & Fennig, 2018) lists 275 varieties as indigenous to the P.R.C. Of those varieties indigenous to the P.R.C., around 120 of them are Sinitic, with the rest representing a wide range of language families, most prominently Tibetan, Altaic, Tungusic, and Semitic. Table 3.1 presents the same sentence in Mandarin and Cantonese, demonstrating differences in the phonetic realization of some words (e.g. Mandarin ㄍㄟ vs. Cantonese ㄆㄟ for "give") and a structural difference whereby in Mandarin the indirect object precedes the direct object and vice versa for Cantonese. A brief overview of the relationship between Sinitic languages, compiled based on the varieties noted by Ramsey (1987) and Simons & Fennig (2018) is in Figure 3.1 below. The varieties most commonly known in the U.S. are at the top of the figure.

Table 3.1 “He gave me three books” in Mandarin and Cantonese

<table>
<thead>
<tr>
<th>Mandarin</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tā ㄍㄟ wǒ sān-bên shū</td>
<td>K'ōi⁴ pei³ sa:m¹-pun³ sǔ ngo⁴</td>
</tr>
<tr>
<td>He give me 3-CL book</td>
<td>He give 3-CL book me</td>
</tr>
<tr>
<td>“He gave me three books”</td>
<td>“He gave me three books”</td>
</tr>
</tbody>
</table>
Putonghua is meant to be the common language of speakers throughout the P.R.C., and is the language of learning and teaching in schools. However, most regions and ethnic groups have their own version of the standard which differs in vocabulary and pronunciation from the prescriptive varieties in Beijing and the Republic of China (ROC: Taiwan) (Ramsey, 1987). The local variation in Putonghua
creates a dichotomy such that, for example, people in Guangdong speak the local language (Cantonese) for everyday communication, but use Putonghua in educational contexts. Despite the intention of the government in assigning a standard language, the Putonghua spoken in Guangdong has lexical, phonetic, and structural features of Cantonese which makes that version of Putonghua very different from the Putonghua spoken in other places.

In New York City, various neighborhoods like Chinatown in Manhattan, Sunset Park in Brooklyn, and Flushing in Queens have concentrations of immigrants from China, with varieties including Mandarin, Cantonese, Shanghainese, and varieties of Min present. Just as Putonghua is the language of education in the P.R.C., it is also the language typically used for education in New York City Schools in places where any variety of Chinese is offered. Accordingly, most resources for Chinese language programs and materials in New York City are produced in Mandarin.

Mandarin was chosen as the focal point for this dissertation because of its widespread use. As the official language of the P.R.C., Mandarin is the spoken variety taught in schools across the P.R.C. Further, Ethnologue notes that 70% of Chinese language users speak a Mandarin dialect as a native language (Simons & Fennig, 2018; all figures in this section are from 2015). Ethnologue lists 896,000,000 native speakers of Mandarin in China and 178,000,000 second language speakers in China. Worldwide, there are 1,107,162,230 speakers of Mandarin, counting both native language users and second language users. By comparison, Shanghainese has 80,700,000 native speakers in China and 80,797,910 speakers worldwide and Cantonese has 62,000,000 native speakers in China and 73,757,610 speakers worldwide (Simons & Fennig, 2018; also see Ramsey, 1987). Further, Mandarin was chosen because as the language of learning and teaching in the P.R.C., R.O.C., and dual language Chinese programs in New York City, it is often the starting point for most educational tools. Additionally, because the population of Chinese-speakers in New York City is relatively ill-defined, an important next step will be to gather data about what varieties of Chinese are used most commonly in New York City, and whether Mandarin is actually a useful language for a diagnostic.
3.2 Bilingual Language Acquisition

Cross-linguistically, children follow a similar arc of language development, which are often represented in terms of language development milestones. Babbling begins at around six months and the first word is spoken at around the one-year mark. Typical two-year-olds are able to produce two-word telegraphic utterances, with content words but little to no grammatical machinery. By four years of age, children can perform basic speech acts, using correct word order and morphological markers most of the time, though they still struggle with complex sentences and pragmatic use of language. Throughout the stages of language development, children understand more words and structures than they can produce. Production of complex sentence structures such as relative clauses continue to develop after age five, and even up through age eight for typically developing children. These generalizations are true for English-speaking and Mandarin-speaking children (Erbaugh, 1992; Lightbown & Spada, 2013; Zukowski, 2016).

Monolingual language development milestones represent an expectation for the order in which children are able to use features of a language, along with the average age at which those features emerge. Even within typically developing monolingual children there is a great deal of variation regarding when the first word is spoken and other milestones. Despite the presence of individual difference, the milestones remain a useful guideline for the typical trajectory of language development, at least in the prestige variety that the milestones are based on.

It is important to note that language development milestones are based largely on features of language production, and very little research has been conducted on language development milestones for comprehension. This gives an incomplete description of language development, and likely an inaccurate description of language ability. Since that children are able to understand structures that they do not yet produce, this is a gap in the child language development literature that needs to be addressed. Language production is not necessarily representative of a speaker’s underlying grammatical knowledge. An array of performance effects, such as discourse and message planning, lexical look-up, phonetic and phonological output mechanisms, and the integration of all of those systems can be the cause of variability and/or omission in production. Therefore, production alone is not a valid measure of what learners know (Martohardjono, Valian, & Klein, 2018).
While monolingual language development milestones can be a useful starting point to make predictions for bilingual children, monolingual language development milestones cannot be assumed to strictly apply to children developing two languages. Research measuring the language skills of children acquiring two languages has shown that while the order of acquisition described by monolingual milestones in each language seems to be similar for bilingual children, the typical age at which bilingual children reach those milestones is different from monolingual children. This is to be expected, since a child exposed to two languages will likely receive less input in each language than an age-matched monolingual child (Roseberry-McKibbin, 2014). Furthermore, bilingual grammars interact and have bi-directional influence on each other, so the grammar of a bilingual child (or adult) could not reasonably be expected to be the same as the grammar of a monolingual (Cook, 1973; 1991).

In terms of assessment, monolingual milestones can provide a starting point for target structures to be used in bilingual assessments, but several factors need to be kept in mind. First, assessments must include aspects of syntax comprehension. A measure of how many words a child can understand and produce is not enough. Grammatical relationships must be included in an appropriate assessment of language development. Many tools for parents and educators include a list of words commonly spoken or understood by children at a given age, and give guidance indicating that children should be able to appropriately respond to simple requests at age one and to follow more complex instructions by age four. However, assessments directly measuring language in a child (not through parent report) are also necessary, and these must include systematic study of structures that have been established in child language development literature. Furthermore, the trajectory of monolingual language development cannot be assumed to apply for bilinguals. Rather, expectations for bilingual development must be established through evidence drawn from research studies examining the language development of bilingual children in each of their languages.

This investigation examines structures that are attested in the language development milestone literature in English and Mandarin, and presents an assessment of structures chosen from evidence-based acquisition hierarchies. These structures are described in Section 3.3. A comprehension task incorporating these structures administered to Chinese-English bilingual four-year olds provides context for bilingual language development patterns in Mandarin and English.
3.3 Child Language Development of Coordination and Relative Clauses

This section discusses the structures examined in this investigation. This dissertation examines coordination and relative clauses, two structures that are common to English and Mandarin, and that are presumed to be cross-linguistic universals; as well as classifiers, which are present in Mandarin but not English.

Coordination and subordination, specifically relative clauses, are cross-linguistically robust structures that have been studied in child language acquisition in English (Lust, Chien & Flynn, 1987; Slobin, 1971; de Villiers, Flusberg, Hakuta, & Cohen, 1979) and to a lesser extent Mandarin (Erbaugh, 1992; Lust & Chien, 1984; Hu, Gavarró, Vernice, & Guasti, 2016). Both structures continue to develop through the preschool years (McKee, McDaniel, & Snedeker, 1998; Lightbown & Spada, 2013; Zukowski, 2016). Coordination is relatively simple and early acquired; and relative clauses (hereafter RCs), are relatively more complex and later acquired. Coordination in Mandarin and English is described in Section 3.3.1 and RC structures in Mandarin and English are described in Section 3.3.2. In addition to the structures shared by Mandarin and English, a Mandarin-specific structure that is not present in English is also included. Classifiers are particles that modify nouns and are obligatory following numerals, quantifiers and demonstratives. Acquisition of specific classifiers develops through the preschool years. Classifiers are described in Section 3.3.3.

3.3.1 Coordination

Children as young as 2;6 begin to conjoin ideas using and in English (Limber 1973; Bowerman, 1979). Previous literature has examined four sub-types of coordination: Inflectional Phrase (IP) coordination, which is the conjunction of two clauses; subject coordination, object coordination, and verb coordination. Examples of each type of coordination in English are given in Table 3.2, below.

In English, all four types of coordination use only the lexical item and. In Mandarin, er 而 is used for IP coordination, and he 和 is used for subject and object coordination. Verb coordination takes the form of two neighboring conjuncts, either with or without an overt coordinator. For the overt coordinator, each conjunct is preceded by the string yi bian 一边. See examples of Mandarin coordination structures in Table 3.3, below.
Table 3.2 English Coordination Types

<table>
<thead>
<tr>
<th>Coordination Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Coordination</td>
<td>The bear swims and the dog walks</td>
</tr>
<tr>
<td>Subject Coordination</td>
<td>The bear and the dog chase the cat</td>
</tr>
<tr>
<td>Object Coordination</td>
<td>The cat chases the bear and the dog</td>
</tr>
<tr>
<td>Verb Coordination</td>
<td>The rabbit hits and kicks the cat</td>
</tr>
</tbody>
</table>

Table 3.3 Mandarin Coordination Types

<table>
<thead>
<tr>
<th>Coordination Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Coordination</td>
<td>小熊在走路而小狗在游泳 [little bear PROG walk] CNJ [little dog PROG swim] 'The bear is walking and the dog is swimming'</td>
</tr>
<tr>
<td>Subject Coordination</td>
<td>小熊和小猫追着小狗 [little bear CNJ little cat] chase DUR little dog 'The bear and the cat are chasing the dog'</td>
</tr>
<tr>
<td>Object Coordination</td>
<td>小狗拉着小熊和小猴 [little bear CNJ little monkey] pull DUR 'The dog is pulling the bear and the monkey'</td>
</tr>
<tr>
<td>Verb Coordination</td>
<td>小狗一边跳着一边抱着小熊 [one side jump DUR one side hug DUR little bear] 'The dog is jumping and hugging the bear.'</td>
</tr>
</tbody>
</table>

Results from elicited imitation tasks (Lust, Chien & Flynn, 1987), show that young English-speaking children (ages two and three years old, but replicated with older children) can successfully repeat IP coordination earlier than subject, object, and verb coordination.⁶ Lust et al. also report better performance on object coordination than subject and verb coordination for English-speaking children, with a coordination acquisition hierarchy of IP coordination > object coordination > subject coordination = verb coordination.

---

⁶ However, see Tager-Flusberg, DeVilliers and Hakuta (1982) for a longitudinal study of three English-speaking children showing that phrasal coordination precedes sentential coordination, or Bloom, Lahey, Hood, Lifter & Fiess (1980) for a narrative task showing there is no fixed order of acquisition in natural speech patterns of English-speaking children. Though these studies have different findings, please also note the issues with production-based tasks discussed in Section 3.2.
IP coordination precedes other types of coordination in Mandarin, just as in English. Lust & Chien (1984) tested young Mandarin-speaking children in Taiwan on an elicited imitation task. The children were more accurate repeating sentences with IP coordination than any other type, just as the English-speaking children were. Lust & Chien also found that children were more accurate on null-conjunction verb coordination than subject or object coordination in Mandarin, yielding a coordination acquisition hierarchy of IP coordination > verb coordination > subject coordination = object coordination (with earliest acquired structure listed first).

3.3.2 Relative Clauses

Sentences with relative clauses develop later than coordination, with the earliest signs of production and comprehension at around two to three years of age (Sheldon, 1974; McKe, McDaniel, & Snedeker, 1998; Martohardjono et al. 2004, 2005). A relative clause is a subordinate clause which modifies a nominal head. There is parametric variation in whether RC structures precede or follow the head noun they modify. Whereas English RCs obligatorily follow the head noun (Fodor, Bever & Garrett, 1974), Mandarin relative clauses must precede the head noun (Lust & Chien, 1984). The differences are further explained and examples are given in the tables below.

In English, RCs can be introduced with one of several overt complementizers (e.g. that, who, etc.) or in some cases, a null complementizer. The surface order of the constituents is influenced by embeddedness (the position of the relative clause as the subject or the object of a sentence), and by the gap position (whether the nominal head is associated with the subject position or object position of the subordinate phrase). In subject-gap RCs, the gap is in the subject position of the subordinate clause, and in object-gap RCs, the gap is in the object position of the subordinate clause. The interaction between embeddedness and gap position is shown in Table 3.4, with accompanying examples in Table 3.5, below.

Like English, Mandarin follows a SVO word order, but in contrast with English, Mandarin RCs precede the head noun (Lust & Chien, 1984). Mandarin RCs are introduced with the particle de 的. Mandarin RCs also differ according to embeddedness of the relative clause and gap position. Embeddedness involves the position of the RC as the subject or object of a sentence, and the gap...
position refers to the position of the relativized element in the subordinate clause. Examples of each type of Mandarin RC are in Table 3.6, below.

Table 3.4 Relative Clause Position Versus Gap Position

<table>
<thead>
<tr>
<th>Type</th>
<th>Subject-Gap</th>
<th>Object-Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object-Embedded</td>
<td>OS</td>
<td>OO</td>
</tr>
<tr>
<td>Subject-Embedded</td>
<td>SS</td>
<td>SO</td>
</tr>
</tbody>
</table>

Table 3.5 English Relative Clause Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Embeddedness</th>
<th>Gap Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>The bear hugs the monkey [who ₯ kisses the dog]</td>
<td>object-embedded</td>
<td>subject gap</td>
</tr>
<tr>
<td>SS</td>
<td>The monkey [who ₯ kisses the dog] dances</td>
<td>subject-embedded</td>
<td>subject gap</td>
</tr>
<tr>
<td>OO</td>
<td>The monkey kisses the dog [who the bear hugs ₯ ]</td>
<td>object-embedded</td>
<td>object gap</td>
</tr>
<tr>
<td>SO</td>
<td>The dog [who the bear hugs ₯ ] sings</td>
<td>subject-embedded</td>
<td>object gap</td>
</tr>
</tbody>
</table>

Table 3.6 Mandarin Relative Clause Types

<table>
<thead>
<tr>
<th>RC Type</th>
<th>Example</th>
<th>Embeddedness &amp; gap position</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>小熊 抱 着 跳 的 小狗 &lt;br&gt;xiao xiong bao zhe tiao de xiao gou &lt;br&gt;wwl little bear hug DUR [jump REL little dog]</td>
<td>object-embedded, subject gap</td>
</tr>
<tr>
<td>SS</td>
<td>跳舞 的 小狗 在 打 小熊 &lt;br&gt;tiao wu de xiao gou zai da xiao xiong &lt;br&gt;[dancing REL little dog] PROG hit little bear</td>
<td>subject-embedded, subject gap</td>
</tr>
<tr>
<td>OO</td>
<td>小猴 抱 着 小猫 推 着 的 小熊 &lt;br&gt;xiao hou bao zhe xiao mao tui zhe de xiao xiong &lt;br&gt;little monkey hug DUR [little cat push DUR REL little bear]</td>
<td>object-embedded, object gap</td>
</tr>
<tr>
<td>SO</td>
<td>小熊 摸 着 的 小猴 抱 着 小狗 &lt;br&gt;xiao xiong mo zhe de xiao hou bao zhe xiao gou &lt;br&gt;[little bear touch DUR REL monkey] hold DUR little dog</td>
<td>subject-embedded, object gap</td>
</tr>
</tbody>
</table>
Early studies of RC elicited imitation (Smith, 1974), binary picture-point comprehension (Brown, 1971), and act-out comprehension (de Villiers, Flusberg, Hakuta, & Cohen, 1979) report higher accuracy rates for subject-gap RCs than object-gap RCs in English-speaking children. This asymmetry is robust and has dominated the modern RC literature (Traxler et al., 2002). Currently, researchers seem to agree that in child language acquisition of English there is a preference for subject-gap relative clauses over object-gap relative clauses, and a preference for object-embedded sentences over subject-embedded sentences (Slobin, 1971; Cook, 1973; Lahey, 1974). The hierarchies regarding which is more influential has not been resolved, and a number of studies have studied only embeddedness, without regard for gap position, or only gap position without regard for embeddedness, which has led to results that can seem contradictory (de Villiers & de Villiers, 1985; Eisenberg, 2002; Hu, Gavarró, Vernice, & Guasti, 2016).

Studies on child language development in Mandarin have also found a subject-gap preference for Mandarin RCs (Vasishth, Chen, Li & Guo, 2013; Hu, Gavarró, Vernice, & Guasti, 2016; Yan & Matthews, 2017), though Lee (1992) and Cao, Goodluck & Shan (2005) did not find that asymmetry for children under five years old. Some studies have found a preference for object-gap RCs and have hypothesized that that is because the relativized element precedes rather than follows the head noun in Mandarin (Hisao & Gibson, 2003; Gibson & Wu, 2013). However, Lin (2006) and Lin & Bever (2006) found fault with the methodology from Hisao & Gibson (2003).

The study that stands out as a model for child acquisition of Mandarin is Hu, Gavarró, Vernice, & Guasti (2016). Hu et al. conducted a picture-point comprehension task with 20 children ages three to eight and report the results for children by age. Results showed that while accuracy increased with age, children ages three to seven showed the expected asymmetry and performed with better accuracy on subject-gap relative clauses than object-gap relative clauses (eight year old children performed near ceiling on both structures). However, there was no contrast between subject-embeddedness and object-embeddedness within the subject-gap and object-gap RC items. The question of whether the gap asymmetry or embeddedness asymmetry is more influential has not been resolved in Mandarin.
3.3.3 Classifiers

In addition to universal structures like coordination and relative clauses, this dissertation examines Mandarin classifiers, which are not present in English. Classifiers occur before head nouns and are obligatory following numerals, quantifiers and demonstratives. Bare head nouns are grammatical without classifiers, though classifier use in that context is not ungrammatical. Mandarin has a general classifier, \textit{ge 个}, which can be used with all objects and people. In addition, there are sortal classifiers, which are specific and based on feature bundles that label a group of objects. These categories range from very narrow, such as the classifier indicating books (\textit{bĕn 本}) and the classifier indicating flowers (\textit{duo 朵}), to very broad categories based on shape such as \textit{tiáo 条}, which indicates long thin objects including fish, rivers, roads, pants, and more (Erbaugh, 2006).

Classifiers develop according to an acquisition hierarchy that follows universal patterns of classifier development (Erbaugh, 1986; 1990). The patterns are: specific categories before general categories, concrete categories before abstract categories, valued items (people, animals, other conversationally prominent referents) before ordinary items, quantifiers (count before mass), and finally, within shape types, and for reasons unclear, long objects before flat objects.

The classifiers used for this investigation are listed in Table 3.7 below. Classifiers are included in this assessment as a language-specific structure, and no English equivalent is presented. This assessment will examine comprehension of classifier meaning, and not constraints regarding their presence or absence.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Meaning</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{bĕn 本}</td>
<td>books/bound objects</td>
<td>book, dictionary</td>
</tr>
<tr>
<td>\textit{zhī 只}</td>
<td>small to medium animals</td>
<td>dog, cat</td>
</tr>
<tr>
<td>\textit{tiáo 条}</td>
<td>long, thin objects</td>
<td>river, road, fish</td>
</tr>
<tr>
<td>\textit{zhāng 张}</td>
<td>flat objects</td>
<td>table, painting</td>
</tr>
</tbody>
</table>

This chapter has provided information about the language varieties and grammatical constructions under investigation for this dissertation. The next chapter presents the methodology for an
empirical study of classifiers, RCs, and coordination structures in Chinese-English bilingual preschool children.
4. Experiment

This investigation is a cross-sectional study of four year-old children (N=24) from households where “Chinese” is spoken. Classifiers, RCs and coordination structures are presented as picture-point tasks on a touch screen tablet to child participants. This study serves as a pilot to identify children’s comprehension skills on structures identified as language development milestones in English and Mandarin. Results from this investigation will be used to shape a subsequent version of the assessment, with the goal of creating an appropriate assessment of language development for Chinese-English bilinguals. Because bilinguals are a heterogeneous population, a home language questionnaire is administered to the parents of the preschool participants in order to collect linguistic information about participants’ home varieties, language input, and frequency of use; as well as information about parents’ language profiles and demographic data.

4.1 Research Questions

The research questions for this study are as follows.

RQ (1) Do children from Chinese-speaking homes exhibit the same asymmetries in acquisition hierarchies as established in the literature?

The first asymmetry under investigation is coordination. Higher accuracy is expected on IP coordination structures than subject, verb, and object structures on English-medium tasks (Lust, Chien, & Flynn, 1987) and Mandarin-medium tasks (as per Lust & Chien, 1984). Second is the RC asymmetry. Higher accuracy is expected on subject-gap than object-gap RCs on English-medium tasks (as per de Villiers et al. 1979) and Mandarin-medium tasks (as per Hu et al. 2016). Finally, for classifiers, the order of accuracy (běn (books) > zhī (animals) > tiáo (long thin objects) > zhāng (flat objects)) is expected on a Mandarin-medium task (as per Erbaugh, 1992). It is expected that children will follow each of these hierarchies in line with the literature reviewed in Section 3.3. The acquisition hierarchies will be calculated based on performance accuracy for each structure by sub-condition. The acquisition hierarchies represent the order of acquisition of these structures, with items having higher accuracy being those that are earlier acquired.
For each of the five conditions listed above (i.e. Mandarin-medium coordination, English-medium coordination, Mandarin-medium RCs, English-medium RCs, and Mandarin-medium classifiers), four outcomes are possible. First, the children may show the expected asymmetries based on the review of the monolingual acquisition literature. Second, it may be the case that there is no asymmetry due to participants being at ceiling on the tasks, which would mean that the children tested develop those structures faster than previous studies of monolingual and bilingual children have shown. There may also be no asymmetry if children are at chance on the structures, which would mean that the children had not yet acquired those structures. Performance at chance would not mean that the structures will not be acquired, but following Jia and Paradis (2018), who studied RC production and comprehension in children from Mandarin-speaking households in Canada, there is evidence that bilingual children may exhibit “protracted acquisition”, developing the structures later than monolingual children. Finally, it is possible that children will exhibit a different development pattern that does not include the expected patterns of asymmetry. This could possibly reflect language transfer, as shown in Kidd, Chan & Chiu’s (2015) study on Cantonese-speaking children living in Australia. They found that eight-year olds wrongly interpreted the RC subject as the head referent in object-gap RCs in Cantonese, likely due to influence from English. It is also possible that a different development pattern could be exhibited but not reflect language transfer. The results section for each sub-condition in each language will discuss whether participant responses patterned with the expected hierarchy.

RQ (2) What different patterns emerge when performance data are analyzed by group according to a child’s home language or most frequently used language. Specifically:

a. Do children from Mandarin-speaking households outperform children from Cantonese-speaking households on the Mandarin-medium tasks or English-medium tasks?

b. Do children who most frequently speak English outperform children who most frequently speak Mandarin or Cantonese on the Mandarin-medium tasks or English-medium tasks?

Parents report the language(s) spoken in the household and the language the child speaks most frequently on a Home Language Questionnaire. For all participants, at least one of the household
languages is Mandarin or Cantonese. It is outside the scope of this dissertation to argue for or against the appropriateness of using parent report of most frequently used language as a measure of dominance, since the most frequently used language at the time of testing may not have been the most frequently used language before the child began preschool.

To address RQ2(a), it is expected that performance on Mandarin-medium tasks will be influenced by whether the home language of a child is Mandarin or Cantonese due to the wide typological distance between these varieties (see Chapter 3 for a discussion of variation in the “Chinese” languages). Children from Mandarin-speaking households are expected to outperform children from Cantonese-speaking households on the Mandarin-medium tasks, but not on the English-medium tasks, since the Sinitic variety spoken in the home is not expected to differentially affect a child’s English performance. To address RQ2(b), it is expected that performance on English-medium tasks will be influenced by the language a child speaks most frequently as discussed in Chapter 1. Children who most frequently speak English are expected to outperform children who most frequently speak Mandarin or Cantonese on the English-medium tasks because frequency of use has is one of the factors that influences language proficiency (Roseberry-McKibbin, 2014). Children who most frequently speak Mandarin are not expected to perform differently from children who speak Cantonese on the English-medium tasks, again because the most frequently used Sinitic variety is not expected to differentially affect a child’s English performance. Children who most frequently speak Mandarin are, however, expected to outperform children who speak both Cantonese and English on the Mandarin-medium tasks.

In addition to information about home language and most frequently used language, a number of other demographic and sociolinguistic questions are asked on the Home Language Questionnaire. Variables such as socioeconomic status, a child’s mother’s level of English, and the amount of time a child has spent in China are expected to have an impact on child performance, and so these relationships will be explored. As described in Chapter 2, socioeconomic status is expected to predict a child’s performance on language assessments such that children from higher socioeconomic status households are expected to outperform children from lower socioeconomic status households. Following Hollingshead (1975) and Bornstein, Hahn, & Suwalsky (2014), socioeconomic status is determined using the mother’s highest completed level of education. The other two variables selected for examination are a
child’s mother’s self-reported proficiency level of English and the amount of time a child has spent in China. Children whose mother self-reports higher English proficiency are expected to outperform children whose mother self-reports lower English proficiency on the English-medium tasks, but not the Mandarin-medium tasks. Similarly, Children who have spent over a year in China are expected to outperform children who have not on the Mandarin-medium tasks, but not the English-medium tasks. Further information about how information was collected on the Home Language Questionnaire and how the data were quantified are in Section 5.6.

4.2 Methodology

This section provides an explanation of the research methodology used for this investigation, including descriptions of the research tools used for data collection (Section 4.2.1), the participant recruitment process (Section 4.2.2), and protocols for administration (Section 4.2.3).

4.2.1 Materials

The materials used for this investigation include a battery of comprehension assessments administered as picture-point tasks to preschool children, and a home language questionnaire administered to the children’s parents. The comprehension tasks include an assessment of coordination and RCs in English (three sets of 20 stimuli each, including fillers), an assessment of coordination and RCs in Mandarin (three sets of 16 stimuli each, including fillers), and a classifier task in Mandarin (one set of 20 stimuli). The home language questionnaire is a survey with 27 questions, which parents completed in Mandarin or English, according to their preference. Stimuli are described in further detail below.

Stimuli for the picture-point comprehension tasks were presented on Microsoft Surface 3 Tablets (10.8-Inch, Windows 10), which recorded and stored responses with E-Prime 2.0 software (Psychology Software Tools, 2016). The tablet played the recorded stimuli and showed the corresponding pictures. Presentation on a tablet reflects the effort to meet the need laid out in Chapter 3 for an assessment that will be maximally useful to education professionals and evaluators working with preschool aged children, given the circumstance that Mandarin-speaking evaluators and speech-language pathologists are not available to all of the students who would benefit from their services. Since it is currently the case that
Mandarin speakers may be assessed using ad-hoc translations of English assessments, if they are assessed in Mandarin at all, an assessment that could be deployed to non-Mandarin speaking professional evaluators, with the assistance of a Mandarin-speaking assistant was developed. The children would be able to take the test fully in a Mandarin speech setting, and the tablet would record the answers and then create an English-language report for the English-speaking professional to understand the skills and abilities in Mandarin. This type of tool has been developed and is in use in English (Quick Interactive Language Screener™ (QUILS™): Golinkoff, de Villiers, Hirsh-Pasek, Iglesias & Wilson, 2017).

The English syntax assessment is the RISLUS Multilingual Syntax Test (RMST), developed by Klein & Martohardjono (2009) as a test of comprehension for coordination, RCs, temporal adverbials, and control structures. The test consists of four coordination conditions, with five items per condition, four relative clause conditions, with five items per condition, two temporal adverbial conditions, with five items per condition, and two control conditions, with five items per condition. A list of examples of each type of structure on the assessment is in Appendix 1. Additionally, the examples of coordination and RCs presented in Table 3.2 and Table 3.5 are test items. In total, there are 60 test items and five practice items. In order to make the task length suitable for preschool students, the item types were divided across three sets of 20 items each. Responses were collected for all structures, but only the coordination and RC items are analyzed for this investigation.

The Mandarin syntax assessment was trans-adapted from the English version of the RMST. A team of linguists, native speakers, and educators developed the Mandarin items using the English version as a base. The Mandarin version of the test consists of four coordination conditions, with five items per condition, four RC conditions, with five items per condition, and one temporal adverbial condition, with five items per condition. Again, a list of examples of each type of structure on the assessment is in Appendix 1. Additionally, the examples of coordination and RCs presented in Table 3.3 and Table 3.6 are test items. In total, there are 45 test items and five practice items. In order to make the task length suitable for preschool students, the item types were divided across three sets of 15 items each. Responses were collected for all structures, but only the coordination and RC items are analyzed for this investigation.

The items were audio-recorded by a native speaker of Mandarin for the Mandarin tasks, and a native speaker of English for the English tasks. Each item is accompanied by three images, one of which
matches the sentence. An example is in Figure 4.1, below. The same images are used for the English and Mandarin versions of the test.

**Figure 4.1 RMST Image Example**

```
xiao xiong  zai zou lu  er  xiao gou  zai  you yong
little bear  PROG walk  CNJ  little dog  PROG swim

The bear is walking and the dog is swimming
```

The Mandarin classifier assessment was developed by the same team that trans-adapted the Mandarin version of the RMST. The test consists of four classifier conditions, with five items per condition. The total number of test items is 20, with two practice items. The classifier labeling task consists of the base sentence “Please point to that x”, where x is one of four classifiers: zhī 只 ‘small to medium animals’; tiáo 条 ‘long, thin objects’; běn 本 ‘books/bound objects’; or zhāng 张 ‘flat objects’. An example is in Figure 4.2 below. Participants are trained on the task using two practice items with general classifier ge 个, which can be used with all objects and people. Children hear each sentence once, and the constellation of pictures is different for each item. Five sentences are included for each condition.
The Home Language Questionnaire consists of 27 questions. 17 questions target information about the child, such as most frequently used language (generally), most frequently used language with different family members, and whether the child has spent a protracted amount of time outside of the U.S. 10 questions target information about the parents’ language and educational background. The Home Language Questionnaire is written in Mandarin and in English. Questions are multiple choice except for child’s birth date, parents’ occupations, and parents’ hometowns.

4.2.2 Participants

24 children ($M_{\text{age}} = 4;4, \ SD = 0;3, \ Range = 4;0–4;9$) participated in this study. The child participants for this study are preschool students who live with at least one Chinese-speaking parent or caregiver. This section describes the recruitment process and provides a brief overview of general characteristics of the participants for this study. This project was approved by the University Integrated Institutional Review Board of the City University of New York (file #2016-0788) and the Institutional Review Board of the New York City Department of Education (file #1323).

19 preschools across four boroughs of New York City were contacted about this study, and five schools chose to participate. Parents of children who attended the schools chose to opt in to the study. All recruitment and consent materials were provided in Mandarin and English. English-speaking, Mandarin-speaking, and Cantonese-speaking research assistants were available to respond to parent questions. Parents who opted in to the study signed a permission form allowing researchers to work with their
children, as well as a consent form which was required so that parents could complete a Home Language Questionnaire. Children were recruited regardless of proficiency level in Mandarin or English, socio-economic status, or variety of Chinese spoken in the home.

A Home Language Questionnaire was developed to gather sociolinguistic information about child participants and their parents. The questionnaire consists of 27 questions. 30 responses to the Home Language Questionnaire were received, and responses from 24 questionnaires are presented for this investigation.

Most inquiries on the Home Language Questionnaire are in multiple choice format. Questions about the child’s most commonly used language, the language varieties spoken by the adults in the household, parents’ level of education, and parents’ English proficiency are presented in multiple choice format to ensure consistency of answers for the purpose of quantifiability and comparison. Questions about parents’ occupations and parents’ hometowns are presented as open-ended short-answer questions. Questionnaires were distributed as a single packet with all questions first in Mandarin and then all questions in English. A cover sheet explained the instructions in both Mandarin and English, and invited parents to respond to questions in the language they preferred.

To report results of the language comprehension assessments, the children are grouped first by home language: Mandarin (n=11), and Cantonese (n=13). This distinction is important, because while “Chinese” speakers are often grouped together in demographic reports, whether a child is exposed to Mandarin or another Sinitic variety at home is expected to affect performance on a Mandarin assessment tool. Results for all assessments are also reported by group according to whether children most frequently speak Mandarin (n=8), Cantonese (n=7), and English (n=7). For the two children who are reported to most frequently speak both Cantonese and English, results for the language assessment are compared to the groups of frequent Cantonese speakers and frequent English speakers, and the children are combined with the group that is the closer match. Frequency of use is an important measure because it is expected that performance on the English assessment and Mandarin assessment will be affected by which language the child uses most often. More information about how these groups were determined based on answers to the Home Language Questionnaire is in Section 5.6.
Other factors that are included in the analysis of sociolinguistic variables include amount of time a child has spent in China, Mother’s English proficiency, and socioeconomic status as measured by mother’s level of education. Time spent in China is a measure of input in Mandarin, Cantonese, or another Sinitic variety. Mother’s level of education has typically been used as the key socioeconomic status indicator (Hollingshead, 1975; Bornstein, Hahn, & Suwalsky, 2014), and is used the same way in this study. In this sample, children have been assigned to one of three socioeconomic status groups based on mother’s level of education: high socioeconomic status (n=4), middle socioeconomic status (n=13), and low socioeconomic status (n=7). Mother’s level of English is used as a measure of English input based on the status of mothers as the most frequent primary caregiver, as is standard practice in language assessment studies (Hammer, Davison, Lawrence, & Miccio, 2009). Children are therefore grouped by mother’s self-reported level of English including those with excellent (n=6), good (n=11), or limited (n=4) English skills. Children whose mothers did not report English level are excluded from that analysis (n=3).

4.2.3 Procedure

This section explains the details of administration and the setting for data collection. Data collection was conducted at the five participating schools. Children were tested individually during the school day in a small area of the school such as a cubicle used by Speech-Language Pathologists, or multipurpose room. After permission, consent, and Home Language Questionnaire forms were completed by the parents, children were asked if they wanted to play a language game on a tablet. A Mandarin-speaking research assistant explained the Mandarin tasks, guiding each student through the practice session of each task and offering neutral encouragement as the child progressed through the task. At the end of each task, the child selected a sticker and chatted or colored with the research assistant before deciding whether to participate in another task. The same process was repeated with an English-speaking research assistant for the English tasks. The child opted in to each of eight sessions\(^7\), distributed over two to three days for each child. If the child wanted to stop participating in the middle of a

---

\(^7\) In addition to the 3 sets of English RMST items, 3 sets of Mandarin RMST items, and 1 set of classifier items, an additional Mandarin grammaticality judgment task was administered. Example items are in Appendix 1.
task, the research assistant brought the child back to join their class. As a result, not all data sets are complete for all children. Parents and schools did not receive compensation for participation.

For each of the language tasks, participants are first introduced to all of the characters or objects that would appear on the tasks in the appropriate language. The child is led through the practice items with a research assistant in order to understand how to complete the task. For each RMST item, the child hears a sentence and then three pictures appeared such as those in Figure 4.1 (page 40). The sentence automatically plays again when the pictures appear, so each child hears each sentence twice. Children are asked to touch the picture that best matched the sentence they heard. Touching one picture causes all of the pictures to disappear and a solid color screen to appear as the next sentence is played. Touching the screen again caused three new pictures to appear, and so on. Pictures remain on the screen until one picture is selected. After every three sentences, children are shown a colorful filler picture. For the Classifier task, because the sentences are much shorter, they were played only once. Example images are above, in Figure 4.2 (page 41).
5. Results

To address research question (1) *Do children from Chinese-speaking homes exhibit the same asymmetries in acquisition hierarchies as established in the literature?*, results are reported for child performance on each of the five conditions described above: Mandarin coordination, English coordination, Mandarin RCs, English RCs, and Mandarin classifiers. Means and standard deviations are reported across group for each condition and sub-condition (i.e. coordination types IP, subject, object, and verb, as described in Section 3.3.1; RC types OS, SS, SO, and OO, as described in Section 3.3.2; and classifiers běn (books), zhī (animals), tiáo (long thin objects), and zhāng (flat objects), as described in Section 3.3.3). The means will suggest the order of accuracy for the structure types in each condition. The order of accuracy suggests the order of acquisition, as earlier acquired structures will be expected to have higher accuracy rates. Furthermore, for each sub-condition, t-tests from chance are run in order to determine whether children score above chance on a structure type. If t-tests show no significant difference from chance, that means that there is no evidence that the structure has (or has not) been acquired. For t-tests that do show a significant difference from chance, the structure is in development, but a claim that the structure has been fully developed cannot be made unless the means are also at or near ceiling. For all items in this investigation, chance is 33%, since children select one out of three pictures.

In addition to the across-group analysis which will show the mean results for each structure type for all children from homes where a “Chinese” is spoken, the data from the across group study will be divided based on whether a child’s home language is Mandarin or Cantonese. This will address research question (2a) *Do children from Mandarin-speaking households outperform children from Cantonese-speaking households on the Mandarin-medium tasks or English-medium tasks?*. Means and standard deviations for each home language group will be reported, and a 2 X 4 mixed ANOVA will be run in order to determine whether the main effects of group and condition are significant, and whether there is an interaction between group and condition. Where the main effect is significant, pairwise t-tests with a Bonferroni-Holm correction will be run in order to explore the significant main effect, and the relationships will be reported. Additionally, where the main effect of group is significant, t-tests from chance will be run for each structure type in order to illustrate whether there are differences between which structures show
trends toward development (as suggested by t-test scores above chance) and which structures are at chance and therefore development cannot be determined. To address research question (2b) *Do children who most frequently speak English outperform children who most frequently speak Mandarin or Cantonese on the Mandarin-medium tasks or English-medium tasks?*, the same process is repeated as in the home language analysis described above, using a 3 X 4 mixed ANOVA to examine the groups of children who most frequently speak Mandarin, Cantonese, and English.

Results for developmental hierarchies to address RQ 1 and language groups to address RQ 2 will be presented for each condition in the following order: Mandarin coordination, Mandarin RCs, English coordination, English RCs, and finally Mandarin classifiers. Next, in order further examine the “Chinese-speakers” in this study, who are a heterogeneous group, results of the Home Language Questionnaire are described in detail, including information about the children’s demographics, language input, and language use, as well as information about parents’ language background, English proficiency, and socioeconomic status. Correlations between performance on Mandarin and English syntax structures, and correlations between performance on the Mandarin and English syntax tasks and the sociolinguistic variables, socioeconomic status, mother’s English proficiency, and amount of time a child spent in China will be reported after the within-condition results are reported.
5.1 Mandarin Coordination

The Mandarin task consists of 45 items, including 20 coordination items (five per condition).

Examples of coordination items are in Table 5.1 (reproduced from Table 3.3).

<table>
<thead>
<tr>
<th>Coordination Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence Coordination</strong></td>
<td>小熊 在 走路 而 小狗 在 游泳</td>
</tr>
<tr>
<td></td>
<td>xiao xiong zai zou lu er xiao gou zai you yong</td>
</tr>
<tr>
<td></td>
<td>[little bear PROG walk] [little dog PROG swim]</td>
</tr>
<tr>
<td></td>
<td>‘The bear is walking and the dog is swimming’</td>
</tr>
<tr>
<td><strong>Subject Coordination</strong></td>
<td>小熊 和 小猫 追 着 小狗</td>
</tr>
<tr>
<td></td>
<td>xiao xiong he xiao mao zhi zhe xiao gou</td>
</tr>
<tr>
<td></td>
<td>[little bear CNJ little cat] chase DUR little dog</td>
</tr>
<tr>
<td></td>
<td>‘The bear and the cat are chasing the dog’</td>
</tr>
<tr>
<td><strong>Object Coordination</strong></td>
<td>小狗 拉 着 小熊 和 小猴</td>
</tr>
<tr>
<td></td>
<td>xiao gou la zhe xiao xiong he xiao hou</td>
</tr>
<tr>
<td></td>
<td>Little dog pull DUR [little bear CNJ little monkey]</td>
</tr>
<tr>
<td></td>
<td>‘The dog is pulling the bear and the monkey’</td>
</tr>
<tr>
<td><strong>Verb Coordination</strong></td>
<td>小狗 一边 跳 着 一边 抱 着 小熊</td>
</tr>
<tr>
<td></td>
<td>xiao gou yi bian tiao zhe yi bian bao zhe xiao xiong dog</td>
</tr>
<tr>
<td></td>
<td>[one side jump DUR one side hug DUR little bear]</td>
</tr>
<tr>
<td></td>
<td>‘The dog is jumping and hugging the bear.’</td>
</tr>
</tbody>
</table>

5.1.1 Across group

The first research question asks whether the bilingual children in this sample display developmental patterns similar to those previously attested in the literature. Acquisition of IP coordination is expected to precede subject, object, and verb coordination, and if this is the case, the accuracy for IP coordination structures would be higher than the other coordination structures. Table 5.2 presents the mean scores and standard deviations for accuracy on Mandarin coordination by sub-condition across group (n=23). Figure 5.1 displays the means for each sub-condition with 95% confidence interval error bars. Because children selected one of three pictures, the line in the figure indicating chance is set at 33%.
Table 5.2 Mandarin Coordination Mean Percent Accuracy Across Group

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>71.43*** (45.39)</td>
</tr>
<tr>
<td>Object</td>
<td>57.14*** (49.72)</td>
</tr>
<tr>
<td>Subject</td>
<td>56.86**  (49.77)</td>
</tr>
<tr>
<td>Verb</td>
<td>47.62**  (50.18)</td>
</tr>
<tr>
<td>Total</td>
<td>58.27    (49.37)</td>
</tr>
</tbody>
</table>

Figure 5.1 Mandarin Coordination Mean Percent Accuracy Across Group

Mandarin coordination has the following order of accuracy, with highest performance listed first: IP coordination > NP coordination > verb coordination. This is in line with findings from Lust & Chien’s (1984) elicited imitation task, in which IP coordination was successfully repeated earlier than other types of coordination in Mandarin. T-tests from chance run on participant-averaged data show that individuals are significantly more accurate than chance on all Mandarin coordination items, IP: $t(22)=6.16$, $p<.001$, $d=1.28$; object: $t(22)=4.44$, $p<.001$, $d=0.93$; subject: $t(22)=3.54$, $p<.01$, $d=0.74$; verb: $t(22)=2.98$, $p<.01$, $d=0.62$. This suggests that participants have acquired the coordination structures, as performance is...
above chance. This finding is in line with experimental findings that coordination is a relatively early-acquired structure that is robust in four-year old children (Sheldon, 1974).

The next two sections address the second research question by presenting statistical tests to show whether the children perform significantly differently when divided by home language (Mandarin or Cantonese) or most frequently spoken language (Mandarin, Cantonese, or English). A description of how these groups were formed based on answers from the Home Language Questionnaire is in Chapter 4. Home language is expected to be relevant for the Mandarin language tasks, but not the English language tasks, as children who are tested in the home language (Mandarin) are expected to outperform children who are not (in this case, the Cantonese group). Most frequently used language is not expected to be more predictive than home language for the Mandarin language tasks, but both constellations of groups will be analyzed for the English stimuli and Mandarin stimuli for all tasks in order to allow for comparison of group performance on Mandarin and English tasks.

5.1.2 Home language

In order to determine whether there is an effect of home language on Mandarin coordination items, the data are separated into two groups based on home language designation of Mandarin (n=10) or Cantonese (n=13). Scores for mean accuracy by sub-condition are reported for each group, Mandarin (CMN) and Cantonese (YUE), in Table 5.3 below. Figure 5.2 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure suggest, the home language background groups have a similar hierarchy of sub-conditions, IP coordination > NP coordination > verb coordination, with the Mandarin home language group having higher overall scores. This hierarchy was the same as that observed for the across-group results.
Table 5.3 Mandarin Coordination Mean Percent Accuracy by Home Language

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Mandarin</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>84.78***</td>
<td>61.02**</td>
</tr>
<tr>
<td></td>
<td>(36.32)</td>
<td>(49.19)</td>
</tr>
<tr>
<td>Object</td>
<td>67.39**</td>
<td>49.15*</td>
</tr>
<tr>
<td></td>
<td>(47.40)</td>
<td>(50.42)</td>
</tr>
<tr>
<td>Subject</td>
<td>68.89*</td>
<td>47.37*</td>
</tr>
<tr>
<td></td>
<td>(46.82)</td>
<td>(50.37)</td>
</tr>
<tr>
<td>Verb</td>
<td>56.52*</td>
<td>40.68</td>
</tr>
<tr>
<td></td>
<td>(50.12)</td>
<td>(49.54)</td>
</tr>
<tr>
<td>Total</td>
<td>69.40 %</td>
<td>49.57 %</td>
</tr>
<tr>
<td></td>
<td>(46.21)</td>
<td>(50.11)</td>
</tr>
</tbody>
</table>

Figure 5.2 Mandarin Coordination Mean Percent Accuracy by Home Language

A 2 x 4 mixed ANOVA was run in R. The main effect of group is significant, $F(1,21)=6.24$, $p<.05$, $\eta^2=.10$, with the Mandarin home language group having a significantly higher mean accuracy than the Cantonese home language group. The main effect of condition is also significant, $F(3,63)=3.85$, $p<.05$, $\eta^2=.10$. Pairwise t-tests with a Bonferroni-Holm correction explore the significant main effect of condition across group. The relationships are shown in Figure 5.3 below. The solid line in the figure indicates that participants are significantly ($p<.05$) more accurate on IP coordination items than verb coordination items. There are no other significant relationships between sub-conditions. All sub-conditions are significantly
different from chance as reported in Section 5.1.1. The interaction between group and condition is not significant, $F(3,63)=0.27$, $p=.84$, $\eta^2=.01$.

**Figure 5.3 Mandarin Coordination Pairwise Comparisons Across Group**

Since the main effect of home language group is significant, t-tests from chance were run on participant-averaged data divided by home language group. The Mandarin home language group responses are significantly more accurate than chance on all Mandarin coordination items, IP $t(9)=7.65$, $p<.001$, $d=2.42$; object: $t(9)=4.16$, $p<.01$, $d=1.32$; subject: $t(9)=2.71$, $p<.05$, $d=0.86$; verb: $t(9)=2.69$, $p<.05$, $d=0.85$, and the Cantonese home language group responses are significantly more accurate than chance on all Mandarin coordination items except verb coordination items, IP $t(12)=3.16$, $p<.01$, $d=0.88$; object: $t(12)=2.45$, $p<.05$, $d=0.68$; subject: $t(12)=2.36$, $p<.05$, $d=0.66$; verb: $t(12)=1.60$, $p=.14$, $d=0.44$. In Figure 5.4 below, the darkened boxes indicate the sub-conditions which are significantly different from chance (see Table 5.3 for means). The outlined box with no background shading for Cantonese verb coordination indicates that it is not different from chance.
5.1.3 Most frequent language

The data for the same participants are next divided according to each child’s most frequently used language: Mandarin (n=7), Cantonese (n=7), or English (n=9). Scores for mean accuracy by sub-condition for each group are in Table 5.4 below. Figure 5.5 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure suggest, the group of participants who most frequently speak Mandarin has a higher mean accuracy across condition than the groups of participants who most frequently speak Cantonese or English. The participants who most frequently speak Mandarin and Cantonese have a similar hierarchical pattern as the home language groups for the coordination condition (IP coordination > NP coordination > verb coordination), while the group of participants who most frequently speak English has a different ranked order of accuracy: subject coordination > IP coordination > verb coordination > object coordination. English is the only group with
better accuracy on verb coordination than one type of NP coordination, and the English group also has the least differentiation between structures as the averages for each condition fall closer together than for the Mandarin or Cantonese groups.

Table 5.4 Mandarin Coordination Mean Percent Accuracy by Most Frequently Used Language

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Most Frequent Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mandarin</td>
</tr>
<tr>
<td>IP</td>
<td>81.82</td>
</tr>
<tr>
<td></td>
<td>(39.17)</td>
</tr>
<tr>
<td>Object</td>
<td>69.70</td>
</tr>
<tr>
<td></td>
<td>(46.67)</td>
</tr>
<tr>
<td>Subject</td>
<td>60.61</td>
</tr>
<tr>
<td></td>
<td>(49.62)</td>
</tr>
<tr>
<td>Verb</td>
<td>58.82</td>
</tr>
<tr>
<td></td>
<td>(49.96)</td>
</tr>
<tr>
<td>Total</td>
<td>67.67</td>
</tr>
<tr>
<td></td>
<td>(46.95)</td>
</tr>
</tbody>
</table>

Figure 5.5 Mandarin Coordination Mean Percent Accuracy by Most Frequently Used Language

A 3 x 4 mixed ANOVA shows that the main effect of group is not significant, $F(2,20)=1.18, p=.33, \eta^2=.05$. Participants who most frequently speak Mandarin, English, and Cantonese do not perform significantly differently from each other on Mandarin coordination items. The main effect of condition is again significant with IP coordination having significantly higher accuracy than verb coordination as
described for the home language group. The interaction between group and condition is not significant, $F(6,60)=1.03$, $p=.42$, $\eta^2=.06$. Because there was no significant effect of group, t-tests from chance by most frequently used language group are not reported.

5.1.4 Summary of Mandarin Coordination Results

In summary, there is a significant effect of condition for Mandarin coordination, with participants having significantly higher accuracy on IP coordination than verb coordination. T-tests from chance across group show that all sub-conditions are significantly more accurate than chance. Children from Mandarin- and Cantonese-speaking homes have a similar hierarchical order of accuracy on the coordination structures, though the Mandarin group has a higher across-condition average. The hierarchical pattern for the children who most frequently speak English is slightly different, with higher accuracy on verb coordination than NP coordination; and less differentiation between the strongest and weakest coordination structures. However, since most frequently spoken language group is not significant, there is not statistical power behind this difference.

When responses are analyzed according to participants' home language, results show a significant effect of group, with Mandarin home language children outperforming Cantonese home language children. There is no significant interaction between home language group and condition found in the ANOVA. When responses are separated according to participant frequency of language use instead of home language, results suggest that despite the trend that children who most frequently speak Mandarin have a higher across-condition average than children who most frequently speak Cantonese and English, there is no significant difference between the groups, and also no significant interaction between most frequently used language and condition.

Taken together, these results suggest that children whose home language is Mandarin significantly outperform children whose home language is Cantonese on Mandarin coordination items, while the frequency of language use analysis does not show a significant difference between the Mandarin, Cantonese, and English groups.
5.2 Mandarin Relative Clauses

Mandarin RC items are presented in the same task as Mandarin coordination. There are 20 RC items (five per condition). Examples of RC items are Table 5.5, below (reproduced from Table 3.6).

Table 5.5 Mandarin Relative Clause Types

<table>
<thead>
<tr>
<th>RC Type</th>
<th>Example</th>
<th>Embeddedness &amp; gap position</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>小熊 抱着 跳的 小狗 (xiao xiong bao zhe tiao de xiao gou)</td>
<td>object-embedded, subject gap</td>
</tr>
<tr>
<td></td>
<td>‘The bear is hugging the dog who is jumping’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[The bear is hugging the jumping dog]</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>跳舞的 小狗 在打 小熊 (tiao wu de xiao gou zai da xiao xiong)</td>
<td>subject-embedded, subject gap</td>
</tr>
<tr>
<td></td>
<td>[dancing REL little dog] PROG hit little bear</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘The dog who is dancing is hitting the bear’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[The dancing dog is hitting the bear]</td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>小猴 抱着 小猫 推着的 小熊 (xiao hou bao zhe xiao mao tui zhe de xiao xiong)</td>
<td>object-embedded, object gap</td>
</tr>
<tr>
<td></td>
<td>little monkey hug DUR [little cat push DUR REL little bear]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘The monkey is hugging the bear who the cat is pushing’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[The monkey is hugging the cat-pushed bear]</td>
<td></td>
</tr>
<tr>
<td>SO</td>
<td>小熊 摸着的 小猴 抱着 小狗 (xiao xiong mo zhe de xiao hou bao zhe xiao gou)</td>
<td>subject-embedded, object gap</td>
</tr>
<tr>
<td></td>
<td>[little bear touch DUR REL monkey] hold DUR little dog</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘The monkey who the bear is touching is holding the dog’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[The bear-touched monkey is holding the dog]</td>
<td></td>
</tr>
</tbody>
</table>

5.2.1 Across group

As with the analysis for coordination, the first part of the RC analysis addresses the patterns of RC development. Results on similar tasks have suggested an asymmetry between subject-gap and object-gap RCs in Mandarin, so the discussion of Mandarin RCs will examine whether an asymmetry exists in addition to the sequence of the four RC sub-conditions. An asymmetry is present if there is different performance on the subject-gap structures (OS and SS) than on the object-gap structures (OO and SO). Table 5.6 presents the mean scores and standard deviations of performance on Mandarin RCs.
by sub-condition across group (n=23). Figure 5.6 displays the means for each sub-condition with 95% confidence interval error bars. As before, the line in the figure indicates chance at 33%.

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object-Object</td>
<td>45.10</td>
</tr>
<tr>
<td>(50.00)</td>
<td></td>
</tr>
<tr>
<td>Object-Subject</td>
<td>62.26</td>
</tr>
<tr>
<td>(48.70)</td>
<td></td>
</tr>
<tr>
<td>Subject-Object</td>
<td>39.22</td>
</tr>
<tr>
<td>(49.06)</td>
<td></td>
</tr>
<tr>
<td>Subject-Subject</td>
<td>53.33</td>
</tr>
<tr>
<td>(50.13)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50.12</td>
</tr>
<tr>
<td>(50.06)</td>
<td></td>
</tr>
</tbody>
</table>

The order of accuracy for Mandarin RCs is OS > SS > OO > SO. This pattern suggests higher performance on the subject-gap structures than the object-gap structures, though the order of means is not evidence of an asymmetry. Still, the pattern of subject-gap structures having higher accuracy rates than object-gap structures replicates Hu et al.’s (2016) findings on a Mandarin picture-point comprehension task. T-tests from chance run on participant-averaged data show that individuals are significantly more accurate than chance on all Mandarin RC items expect for SO RC items, OS:
5.2.2 Home language

To determine whether there is an effect of home language, the analysis performed for the Mandarin coordination (Section 6.1.1) was repeated. The data are separated into two groups based on home language designation of Mandarin (CMN; n=10) or Cantonese (YUE; n=13). Scores for mean accuracy by sub-condition are reported for each home language group in Table 5.7, below. Figure 5.7 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure suggest, both home language groups have the highest accuracy on the OS sub-condition and the lowest accuracy on the SO sub-condition. The Mandarin home language group has higher mean scores on the subject-gap structures than the Cantonese home language group, and lower mean scores on the object-gap structures than the Cantonese home language group, with a higher average across condition. The wider range of averages for the Mandarin home language group indicates that the group has more differentiated results, indicative of an asymmetry, which is visually present in Figure 5.7.

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Mandarin Mean Percent Accuracy</th>
<th>Cantonese Mean Percent Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object-Object</strong></td>
<td>42.22 (49.95)</td>
<td>47.37 (50.37)</td>
</tr>
<tr>
<td><strong>Object-Subject</strong></td>
<td>78.72 (41.37)</td>
<td>49.15 (50.42)</td>
</tr>
<tr>
<td><strong>Subject-Object</strong></td>
<td>37.78 (49.03)</td>
<td>40.35 (49.50)</td>
</tr>
<tr>
<td><strong>Subject-Subject</strong></td>
<td>69.57 (46.52)</td>
<td>40.68 (49.54)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>57.38 (49.59)</td>
<td>44.40 (49.79)</td>
</tr>
</tbody>
</table>
The main effect of group in a 2 x 4 mixed ANOVA approaches significance, $F(1,21)=4.12$, $p=.06$, $\eta^2=.06$, with the Mandarin home language group significantly ($p<.05$) more accurate on RC items than the Cantonese home language group. The main effect of condition is significant ($F(3,63)=3.39$, $p<.05$, $\eta^2=.10$). However, none of the post-hoc comparisons are significant after the Bonferroni-Holm correction.

The interaction between group and condition is significant, $F(3,63)=2.87$, $p<.05$, $\eta^2=.08$. Pairwise t-tests run by group explore the significant interaction of condition and group, and are presented in Figure 5.8 below. As before, the solid line in the figure shows that Mandarin home language participants have significantly ($p<.05$) higher accuracy in OS RCs than SO RCs. The dotted lines in the figure show that Mandarin home language participants perform better on OS RCs than OO RCs, and on SO RCs than SS RCs to a degree approaching significance ($p=.07$). These relationships suggest that the asymmetry observed in Figure 5.7 approaches significance. There are no other significant relationships between sub-conditions for the Mandarin group, and there are no significant relationships at all between sub-conditions for the Cantonese home language group.
All Mandarin RC items are significantly different from chance except for SO RC items in the across-group condition. T-tests from chance by home language group show that responses for the Mandarin home language group are significantly more accurate than chance for only Mandarin subject-gap RC items, OS: $t(9)=4.35, p<.01, d=1.38$; SS: $t(9)=4.89, p<.001, d=1.55$, but not object-gap RC items OO: $t(9)=1.32, p=.22, d=0.42$; SO: $t(9)=0.71, p=.50, d=0.22$. In Figure 6.8 above, the solid boxes indicate the sub-conditions which are significantly different from chance. The Cantonese home language group responses are significantly more accurate than chance only for Mandarin OS RC items, OS: $t(12)=2.33, p<.05, d=0.65$; SO: $t(12)=1.46$; OO: $t(12)=1.68, p=.12, d=0.47$; SS: $t(12)=0.91, p=.38, d=0.25$. The outlined boxes with no background shading in Figure 5.8 indicate the sub-conditions that are not different from chance. The t-tests from chance further suggest a subject-gap / object-gap asymmetry in the Mandarin home language children. Both subject-gap structures (OS and SS) are significantly different from chance, and both object-gap structures (OO and SO) are not. For the Cantonese group, there is not a similar subject-gap / object-gap asymmetry, but the OS structures, which are expected to be the easiest in both English and Mandarin, are the only structures with accuracy above chance for the Cantonese home language group. It is also notable that the Mandarin home language group has a much
higher mean for OS RCs ($M=78.72$, $SD=41.37$) than the Cantonese home language group ($M=49.15$, $SD=50.42$). This could indicate that the Cantonese home language children, who may or may not have been exposed to Mandarin show the same trend toward the asymmetry; however the difference between their home language and Mandarin results accuracy at chance on most structures.

5.2.3 Most frequent language

As with coordination, a second analysis is run on the Mandarin RC data to determine whether there is an effect of most frequently used language: Mandarin (n=7), Cantonese (n=7), or English (n=9). Scores for mean accuracy by sub-condition for each group are in Table 5.8, below. Figure 5.9 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure show, the groups of participants who most frequently speak Mandarin, Cantonese, and English all have the highest accuracy on the OS sub-condition and the lowest accuracy on the SO sub-condition; the same pattern that occurs with the home language group. The Mandarin group has the highest average across the RC condition and also the greatest range between highest and lowest mean average. The English group has the lowest average across the RC condition and also the least amount of differentiation between highest and lowest mean average. The Cantonese group falls in the middle in terms of both average across condition and range between sub-conditions with the highest and lowest accuracy.

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Mandarin</th>
<th>English</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object-Object</td>
<td>48.48</td>
<td>38.89</td>
<td>48.48</td>
</tr>
<tr>
<td></td>
<td>(50.75)</td>
<td>(49.44)</td>
<td>(50.75)</td>
</tr>
<tr>
<td>Object-Subject</td>
<td>85.29</td>
<td>43.59</td>
<td>60.61</td>
</tr>
<tr>
<td></td>
<td>(35.95)</td>
<td>(50.24)</td>
<td>(49.62)</td>
</tr>
<tr>
<td>Subject-Object</td>
<td>33.33</td>
<td>38.89</td>
<td>45.45</td>
</tr>
<tr>
<td></td>
<td>(47.87)</td>
<td>(49.44)</td>
<td>(50.56)</td>
</tr>
<tr>
<td>Subject-Subject</td>
<td>72.73</td>
<td>42.11</td>
<td>47.06</td>
</tr>
<tr>
<td></td>
<td>(45.23)</td>
<td>(50.04)</td>
<td>(50.56)</td>
</tr>
<tr>
<td>Total</td>
<td>60.15</td>
<td>40.94</td>
<td>50.38</td>
</tr>
<tr>
<td></td>
<td>(49.14)</td>
<td>(49.34)</td>
<td>(50.19)</td>
</tr>
</tbody>
</table>

Table 5.8 Mandarin RC Mean Percent Accuracy by Most Frequently Used Language
A 3 x 4 mixed ANOVA shows a significant main effect of group, $F(2,20)=4.65$, $p=.05$, $\eta^2=.12$. Participants who most frequently speak Mandarin are significantly ($p<.01$) more accurate on Mandarin RC items than participants who most frequently speak English. Participants who most frequently speak Cantonese do not perform significantly differently than individuals who most frequently speak Mandarin or English. The main effect of condition is significant, ($F(3,60)=3.23$, $p<.05$, $\eta^2=.10$). However, none of the post-hoc comparisons are significant after the Bonferroni-Holm correction.

The interaction between group and condition approaches significance, $F(6,60)=1.96$, $p=.09$, $\eta^2=.12$. Pairwise t-tests by group with a Bonferroni-Holm correction explore the interaction of condition and group, and are presented in Figure 5.10 below. The solid lines in the figure show that the Mandarin group performs significantly ($p<.01$) better on OS than SO RCs, and on SS than SO RCs. The dotted lines in the figure show that the Mandarin group performs better on OS than OO RCs to a degree approaching significance ($p=.07$). This result is similar to that reported for the Mandarin home language group, and further suggests a subject-gap / object-gap asymmetry. There are no other significant relationships between sub-conditions for the Mandarin group, and there are no significant relationships at all between sub-conditions for participants who most frequently speak Cantonese or English.
As described in Section 5.2.1, t-tests from chance show that participants are significantly more accurate than chance on all Mandarin RC items except for SO RC items in the across group condition. Additionally, t-tests from chance were run on participant-averaged data by group according to most frequently used language. T-tests from chance show that Mandarin home language group responses are significantly more accurate than chance only on Mandarin OS and SS RC items, OO: $t(6)=1.58$, $p=.16$, $d=0.60$; OS: $t(6)=7.44$, $p<.001$, $d=2.81$; SO: $t(6)=0.00$, $p=.99$, $d=0.00$; SS: $t(6)=4.64$, $p<.01$, $d=1.76$. The Cantonese home language group responses are significantly more accurate than chance only on Mandarin OS RC items, OS: $t(6)=6.18$, $p<.001$, $d=2.34$; OO: $t(6)=1.63$, $p=.16$, $d=0.62$; SO: $t(6)=1.58$, $p=.16$, $d=0.60$; SS: $t(6)=1.57$, $p=.17$, $d=0.59$. The English home language group responses are at chance on all Mandarin RC items, OO: $t(8)=0.51$, $p=.62$, $d=0.17$; OS: $t(8)=0.66$, $p=.53$, $d=0.22$; SO: $t(8)=1.15$, $p=.28$, $d=0.38$; SS: $t(8)=0.69$, $p=.51$, $d=0.23$. In Figure 5.10 above, the solid boxes indicate the sub-conditions which are significantly different from chance (see Table 5.8 for means). The outlined boxes with no background shading indicate the sub-conditions that are not different from chance.
5.2.4 Summary of Mandarin Relative Clause Results

In summary, home language group approaches significance and most frequently used group is a significant main effect of the Mandarin RC analysis. In both cases, the Mandarin group performs with the most accuracy: in the home language group comparison the Mandarin group significantly outperforms the Cantonese group; and for the frequency of use comparison, the Mandarin group significantly outperforms the English group, though there are no significant relationships with the Cantonese group in that comparison.

Furthermore, there seems to be support for the subject-gap / object-gap asymmetry for the Mandarin home language group and Mandarin most frequently used group. The asymmetry is supported by t-tests from chance run by group, which shows that the subject-gap RCs (OS and SS) are different from chance, while the object-gap RCs (OO and SO) are not. The asymmetry is further supported by pairwise t-tests showing significantly higher accuracy on OS and SS RCs than and SO RCs, and higher accuracy on OS than OO RCs to a degree approaching significance. This asymmetry is not as strongly present for the Cantonese home language or most frequently used language groups. In both analyses, only the OS RCs are different from chance, and there are no significant relationships between the structures. As OS RCs are expected to be the easiest and earliest acquired, the Cantonese groups show an emergent development of the asymmetry. Whether this is because even the children who hear and speak Cantonese as a home language have had some exposure to Mandarin, or whether there is enough mutual intelligibility between the varieties to promote transfer, or a combination of those two is not able to be addressed by this investigation. However, the asymmetry would likely become present if the children from Cantonese-speaking homes received more exposure to Mandarin. There was no evidence of an asymmetry for the children who most frequently use English, as the responses were not different from chance on any of the RC structures, and there was no relationship between the structures. Again, with more exposure to Mandarin, it would be expected that these children would also develop the asymmetry.

Research indicates that by the age of 9 or 10, most monolingual Mandarin-speaking children have mastered RCs and perform at ceiling on RC tasks (Hu et al. 2016; Jia & Paradis, 2018). With consistent exposure to Mandarin it is expected that all children in the sample would also perform at ceiling on the
tasks, though perhaps at a different rate from the monolinguals (c.f. protracted development analysis in Jia & Paradis, 2018).

5.3 English Coordination

The English syntax task consists of 60 items, including 20 coordination items (five per condition). Examples of coordination items are in Table 5.9 (copied from Table 3.2). Examples of the other structures that served as fillers can be found in Appendix 1.

<table>
<thead>
<tr>
<th>Coordination Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Coordination</td>
<td>The bear swims and the dog walks</td>
</tr>
<tr>
<td>Subject Coordination</td>
<td>The bear and the dog chase the cat</td>
</tr>
<tr>
<td>Object Coordination</td>
<td>The cat chases the bear and the dog</td>
</tr>
<tr>
<td>Verb Coordination</td>
<td>The rabbit hits and kicks the cat</td>
</tr>
</tbody>
</table>

5.3.1 Across group

This section repeats the analysis presented in Section 6.1 for coordination structures in English. As with Mandarin coordination development, IP coordination is expected to precede subject, object, and verb coordination in English as well. Table 5.10 presents the mean accuracy on English coordination by sub-condition across group. Figure 5.11 displays the means for each sub-condition with 95% confidence interval error bars.

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>71.17***</td>
</tr>
<tr>
<td></td>
<td>(45.50)</td>
</tr>
<tr>
<td>Object</td>
<td>51.35**</td>
</tr>
<tr>
<td></td>
<td>(50.21)</td>
</tr>
<tr>
<td>Subject</td>
<td>59.63***</td>
</tr>
<tr>
<td></td>
<td>(49.29)</td>
</tr>
<tr>
<td>Verb</td>
<td>51.38**</td>
</tr>
<tr>
<td></td>
<td>(50.21)</td>
</tr>
<tr>
<td>Total</td>
<td>58.41</td>
</tr>
<tr>
<td></td>
<td>(49.34)</td>
</tr>
</tbody>
</table>
The order of accuracy for English coordination is IP coordination > subject coordination > object coordination = verb coordination. As expected, IP coordination again has higher accuracy than the other types, in line with Lust, Chien & Flynn’s (1987) elicited imitation task. T-tests from chance run on participant-averaged data show that individuals are significantly more accurate than chance on all English coordination items, IP: $t(23)=6.60, p<.001, d=1.35$; object: $t(23)=3.59, p<.01, d=0.73$; subject: $t(23)=4.44, p<.001, d=0.91$; verb: $t(23)=3.02, p<.01, d=0.62$.

5.3.2 Home language

To determine whether there is an effect of home language, the same analysis as for Mandarin coordination was repeated on English coordination items. Scores for mean accuracy by sub-condition are reported for each home language group, Mandarin (CMN; n=11) and Cantonese (YUE n=13), in Table 5.11 below. Figure 5.12 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure show, the Mandarin and Cantonese home language participants have the same mean average across condition, and also a hierarchy of sub-conditions similar to each other and to the across group pattern (IP > S > O ≈ V).
Table 5.11 English Coordination Mean Percent Accuracy by Home Language

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Mandarin</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>70.00</td>
<td>72.13</td>
</tr>
<tr>
<td></td>
<td>(46.29)</td>
<td>(45.21)</td>
</tr>
<tr>
<td>Object</td>
<td>54.00</td>
<td>49.18</td>
</tr>
<tr>
<td></td>
<td>(50.35)</td>
<td>(50.41)</td>
</tr>
<tr>
<td>Subject</td>
<td>60.00</td>
<td>59.32</td>
</tr>
<tr>
<td></td>
<td>(49.49)</td>
<td>(49.54)</td>
</tr>
<tr>
<td>Verb</td>
<td>52.00</td>
<td>50.85</td>
</tr>
<tr>
<td></td>
<td>(50.47)</td>
<td>(50.42)</td>
</tr>
<tr>
<td>Total</td>
<td>59.00</td>
<td>57.92</td>
</tr>
<tr>
<td></td>
<td>(49.31)</td>
<td>(49.47)</td>
</tr>
</tbody>
</table>

Figure 5.12 English Coordination Mean Percent Accuracy by Home Language

A 2 x 4 mixed ANOVA shows no effect of group, $F(1,22)=0.42$, $p=.52$, $\eta^2=.01$. There is a significant main effect of condition, $F(3,66)=3.68$, $p<.05$, $\eta^2=.08$. Pairwise t-tests with a Bonferroni-Holm correction are illustrated in Figure 5.13, below. The dotted lines in the figure indicate that participants are more accurate on English IP than verb and object coordination items to a degree approaching significance ($p=.09$). There are no other significant relationships between sub-conditions. All sub-
conditions are significantly different from chance as reported in Section 5.3.1. The interaction between group and condition is not significant, \( F(3,66)=0.24, p=.87, \eta^2=.01 \).

**Figure 5.13 English Coordination Significant Pairwise Comparisons Across Group**

5.3.3 Most frequent language

Data are next analyzed by most frequently used language: Mandarin (n=8), Cantonese (n=9), and English (n=7). Scores for mean accuracy by sub-condition for each group are in Table 5.12, below. Figure 5.14 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure show, the English group’s average across condition is higher than the average across condition for the Mandarin and Cantonese groups, which are similar to each other. The hierarchy of sub-conditions is similar for all three groups (IP > S > O ≈ V).

**Table 5.12 English Coordination Mean Percent Accuracy by Most Frequently Used Language**

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Mandarin</th>
<th>English</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IP</strong></td>
<td>62.50</td>
<td>93.33</td>
<td>63.41</td>
</tr>
<tr>
<td></td>
<td>(49.03)</td>
<td>(25.37)</td>
<td>(48.77)</td>
</tr>
<tr>
<td><strong>Object</strong></td>
<td>45.00</td>
<td>70.00</td>
<td>43.90</td>
</tr>
<tr>
<td></td>
<td>(50.38)</td>
<td>(46.61)</td>
<td>(50.24)</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>55.00</td>
<td>80.00</td>
<td>48.72</td>
</tr>
<tr>
<td></td>
<td>(50.38)</td>
<td>(40.68)</td>
<td>(50.64)</td>
</tr>
<tr>
<td><strong>Verb</strong></td>
<td>45.00</td>
<td>66.67</td>
<td>46.15</td>
</tr>
<tr>
<td></td>
<td>(50.38)</td>
<td>(47.95)</td>
<td>(50.50)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>51.88</td>
<td>77.50</td>
<td>50.63</td>
</tr>
<tr>
<td></td>
<td>(50.12)</td>
<td>(41.93)</td>
<td>(50.15)</td>
</tr>
</tbody>
</table>
A 3 x 4 mixed ANOVA showed a significant effect of group, \(F(2,21)=8.92, p<.01, \eta^2=.22\). Participants who most frequently use English are significantly \((p<.001)\) more accurate on English coordination items than participants who most frequently speak Mandarin or Cantonese. Participants who most frequently speak Cantonese or Mandarin do not perform differently from each other. The main effect of condition is also significant, \(F(3,63)=3.68, p<.05, \eta^2=.10\), as described in Section 5.3.2.

Since the main effect of home language group is significant, t-tests from chance were run on participant-averaged data by group. Responses for participants who most frequently use English are significantly more accurate than chance on all English coordination items, IP: \(t(6)=10.67, p<.001, d=4.03\); object: \(t(6)=4.32, p<.01, d=1.63\); subject: \(t(6)=5.39, p<.01, d=2.04\); verb: \(t(6)=3.39, p<.05, d=1.28\). Responses for participants who most frequently use Mandarin are significantly more accurate than chance only on English IP coordination items, and approached difference form chance on English subject coordination items, IP: \(t(7)=2.66, p<.05, d=0.94\), subject: \(t(7)=2.21, p=.06, d=0.78\); object: \(t(7)=1.86, p=.11, d=0.66\); verb: \(t(7)=1.59, p=.16, d=0.56\). Responses for participants who most frequently use Cantonese are significantly more accurate than chance only on English IP coordination items, IP: \(t(8)=3.23, p<.05, d=1.08\); object: \(t(8)=1.04, p=.33, d=0.35\); subject: \(t(8)=1.53, p=.17, d=0.51\); verb: \(t(8)=1.03, p=.33, d=0.34\). In Figure 5.15 below, the darkened boxes indicate the sub-conditions which are significantly different from chance (see Table 5.12 for means). The faded boxes indicate the sub-conditions that approach chance. The outlined boxes with no background shading indicate the sub-
conditions that are not different from chance. The interaction between group and condition approaches significance ($F(6, 63) = 0.21, p = .10, \eta^2 = .01$); however, none of the post-hoc comparisons within group are significant after the Bonferroni-Holm correction.

**Figure 5.15 English Coordination Significant Differences from Chance by Most Frequent Language**

**Legend:**
- : significantly different from chance
- : not significantly different from chance

5.3.4 Summary of English Coordination Results

In summary, there is a significant effect of condition for English coordination, with participants having a higher accuracy on IP than verb and object coordination to a degree approaching significance. The other sub-conditions are not significantly different from each other. T-tests from chance run on participant-averaged data show that all sub-conditions are significantly different from chance in the across-group comparison.

When responses are separated according to participants’ home language, results do not show a significant effect of group or interaction between home language and condition, indicating that children with a home language of Mandarin or Cantonese do not perform differently from each other. This is expected as variety of Chinese is not expected to influence results on the English-medium task. When responses are separated according to participant frequency of language use, results show a significant effect of group, with children who most frequently speak English outperforming children who most
frequently speak Mandarin or Cantonese. This result is predicted, because children who have a higher frequency of use for English than a Chinese variety are likely to have more developed English skills than children who more frequently use a Chinese variety. Furthermore, t-tests from chance show that when data are divided by group according to most frequently spoken language, responses for participants who most frequently speak English are significantly different from chance for all sub-conditions, and responses for the groups who most frequently speak Mandarin and Cantonese are significantly more accurate from chance only on IP coordination.

5.4 English Relative Clauses

English RC items are included in the same task as English coordination, described above. There are 20 RC items (five per condition). Examples of RC items are in Table 5.13 (reproduced from Table 3.5).

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Embeddedness</th>
<th>Gap Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>The bear hugs the monkey [who __ kisses the dog]</td>
<td>object-embedded</td>
<td>subject gap</td>
</tr>
<tr>
<td>SS</td>
<td>The monkey [who __ kisses the dog] dances</td>
<td>subject-embedded</td>
<td>subject gap</td>
</tr>
<tr>
<td>OO</td>
<td>The monkey kisses the dog [who the bear hugs __ ]</td>
<td>object-embedded</td>
<td>object gap</td>
</tr>
<tr>
<td>SO</td>
<td>The dog [who the bear hugs __ ] sings</td>
<td>subject-embedded</td>
<td>object gap</td>
</tr>
</tbody>
</table>

5.4.1 Across group

The subject-gap / object-gap asymmetry observed in Mandarin RCs is also expected to be present for English RCs. Table 5.14 presents the mean scores and the standard deviations of accuracy on Mandarin RCs by sub-condition across group (n=24). Figure 5.16 displays the means for each sub-condition with 95% confidence interval error bars.
Table 5.14 English RC Mean Percent Accuracy Across Group

<table>
<thead>
<tr>
<th>Sub-Condition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object-Object</td>
<td>45.87</td>
</tr>
<tr>
<td></td>
<td>(50.06)</td>
</tr>
<tr>
<td>Object-Subject</td>
<td>42.34</td>
</tr>
<tr>
<td></td>
<td>(49.63)</td>
</tr>
<tr>
<td>Subject-Object</td>
<td>35.78</td>
</tr>
<tr>
<td></td>
<td>(48.16)</td>
</tr>
<tr>
<td>Subject-Subject</td>
<td>33.33</td>
</tr>
<tr>
<td></td>
<td>(47.35)</td>
</tr>
<tr>
<td>Total</td>
<td>39.32</td>
</tr>
<tr>
<td></td>
<td>(48.90)</td>
</tr>
</tbody>
</table>

Figure 5.16 English RC Mean Percent Accuracy Across Group

English RC performance has a very small range of mean accuracy between the most successful and least successful conditions. The order of accuracy for is OO > OS > SS > SO. T-tests from chance run on participant-averaged data show that individuals are not significantly more accurate than chance on any English relative clause sub-conditions although performance on English object-subject relative clause items approaches significance, object-subject: $t(23)=2.05, p=.06$; object-object: $t(23)=1.68, p=.11, d=0.34$; $d=0.42$; subject-object: $t(23)=0.66, p=.52, d=0.14$; subject-subject: $t(23)=-0.30, p=.77, d=0.06$. As none of
the sub-conditions is significantly different from chance and only one sub-condition even approaches chance, statistical analyses cannot be run to differentiate group performance on English RC structures.

This result does replicate the subject-gap / object-gap asymmetry found in studies of monolingual (and some bilingual) English-speaking children. However, the structure that approaches significance is the OS structure, which is expected to be the first acquired. Therefore, while the asymmetry is not present, the trend is in the expected direction.

5.5 Mandarin Classifiers

The Mandarin classifier task consists of 20 items, including five bĕn items (books), five zhī items (animals), five tiáo items (long thin objects), and five zhāng items (flat objects), as described in Section 3.3.3. Each classifier item has the carrier sentence ‘Please point to that (classifier [CL])’ (请指出那[CL] ‘qing zhichu na [CL]’). The classifiers used in this study are in Table 5.15 (copied from Table 3.7). The analysis for the classifier task will be reported following the same process as the syntax tasks. The across-group results will address the first research question about the developmental patterns of classifier accuracy, and then the results will be divided into home language and most frequently used language groups in order to observe any differences based on how the groups are formed.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Meaning</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>bĕn 本</td>
<td>books/bound objects</td>
<td>book, dictionary</td>
</tr>
<tr>
<td>zhī 只</td>
<td>small to medium animals</td>
<td>dog, cat</td>
</tr>
<tr>
<td>tiáo 条</td>
<td>long, thin objects</td>
<td>river, road, fish</td>
</tr>
<tr>
<td>zhāng 张</td>
<td>flat objects</td>
<td>table, painting</td>
</tr>
</tbody>
</table>

5.5.1 Across group

Table 5.16 presents the mean accuracy on Mandarin classifiers (n=22). Figure 5.17 displays the means for each sub-condition with 95% confidence interval error bars.
Table 5.16 Mandarin Classifier Mean Percent Accuracy Across Group

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>本・ bèn</td>
<td>77.27*** (42.10)</td>
</tr>
<tr>
<td>条・ tiáo</td>
<td>59.09*** (49.39)</td>
</tr>
<tr>
<td>张・ zhāng</td>
<td>48.18** (50.20)</td>
</tr>
<tr>
<td>只・ zhī</td>
<td>60.91*** (49.02)</td>
</tr>
<tr>
<td>Total</td>
<td>61.36 (48.75)</td>
</tr>
</tbody>
</table>

Figure 5.17 Mandarin Classifier Mean Percent Accuracy Across Group

The order of accuracy for Mandarin classifiers is bèn > zhī > tiáo > zhāng. This is in line with Erbaugh’s (1992) theoretical hierarchy and experimental findings from a narrative task. T-tests from chance run on participant-averaged data show that accuracy is significantly better than chance on all classifier items, bèn: $t(21)=7.81, p<.001, d=1.66$; tiáo: $t(21)=4.98, p<.001, d=1.06$; zhāng: $t(21)=3.17, p<.01, d=0.68$; zhī: $t(21)=4.75, p<.001, d=1.01$. 
5.5.2 Home language

Scores for mean accuracy based on the home language designation of Mandarin (n=12) or Cantonese (n=10) are in Table 5.17, below. Figure 5.18 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure show, both groups have the highest accuracy on bĕn, and the lowest accuracy on zhăng. The Mandarin home language group mean across condition is higher than the Cantonese language group mean, and also the range of highest to lowest performance is greater for the Mandarin home language group, showing greater differentiation of the classifiers.

Table 5.17 Mandarin Classifier Mean Percent Accuracy by Home Language

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Home Language</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mandarin</td>
<td>Cantonese</td>
<td></td>
</tr>
<tr>
<td>本・bĕn</td>
<td>88.00</td>
<td>68.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(32.83)</td>
<td>(46.91)</td>
<td></td>
</tr>
<tr>
<td>条・tiáo</td>
<td>58.00</td>
<td>60.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(49.86)</td>
<td>(49.40)</td>
<td></td>
</tr>
<tr>
<td>张・zhăng</td>
<td>46.00</td>
<td>50.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(50.35)</td>
<td>(50.42)</td>
<td></td>
</tr>
<tr>
<td>只・zhī</td>
<td>68.00</td>
<td>55.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(47.12)</td>
<td>(50.17)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>65.00</td>
<td>58.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(47.82)</td>
<td>(49.40)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.18 Mandarin Classifier Mean Percent Accuracy by Home Language
A 2 x 4 mixed ANOVA shows that the main effect of group is not significant, $F(1,20)=0.98, p=.33, \eta^2=.02$. The main effect of condition is significant, $F(3,60)=6.84, p<.001, \eta^2=.17$. Pairwise t-tests show that participants are significantly ($p<.01$) more accurate on běn items than zhāng items, as indicated by the solid line in Figure 5.19, below. The dotted line indicates a higher accuracy on běn items than tiáo items to a degree approaching significance, $p=.09$. There are no other significant relationships between sub-conditions. Additionally, the interaction between group and condition is not significant, $F(3,60)=1.46, p=.23, \eta^2=.04$.

**Figure 5.19 Mandarin Classifier Pairwise Comparisons Across Group**

5.5.3 Most Frequent Language

The data are next separated into groups based on the parents’ report of most frequently used language: Mandarin (n=8), Cantonese (n=8), and English (n=6). Scores for mean accuracy are in Figure 5.19, below. Figure 5.20 displays the means for each sub-condition with 95% confidence interval error bars. As the table and figure show, participants who most frequently speak Mandarin have the highest mean average, the greatest differentiation between most accurate and least accurate sub-condition, and the same hierarchy of scores as the across-group condition. Participants who most frequently speak English have the lowest mean average, and nearly the same accuracy hierarchy as the across-group condition. The mean average for the participants who most frequently speak Cantonese falls in between the other two groups, and the hierarchy of accuracy is different (běn > tiáo > zhī = zhāng), with accuracy on tiáo items falling higher in the ranking than for the Mandarin or English groups.
Table 5.18 Mandarin Classifier Mean Percent Accuracy by Most Frequently Used Language

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Mandarin</th>
<th>English</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>本 • běn</td>
<td>87.50</td>
<td>70.00</td>
<td>72.50</td>
</tr>
<tr>
<td></td>
<td>(18.32)</td>
<td>(32.86)</td>
<td>(28.16)</td>
</tr>
<tr>
<td>条 • tiáo</td>
<td>65.00</td>
<td>43.33</td>
<td>65.00</td>
</tr>
<tr>
<td></td>
<td>(25.63)</td>
<td>(15.06)</td>
<td>(25.63)</td>
</tr>
<tr>
<td>张 • zhāng</td>
<td>47.50</td>
<td>43.33</td>
<td>52.50</td>
</tr>
<tr>
<td></td>
<td>(23.75)</td>
<td>(15.06)</td>
<td>(26.05)</td>
</tr>
<tr>
<td>只 • zhī</td>
<td>72.50</td>
<td>56.67</td>
<td>52.50</td>
</tr>
<tr>
<td></td>
<td>(30.12)</td>
<td>(15.06)</td>
<td>(30.12)</td>
</tr>
<tr>
<td>Total</td>
<td>68.13</td>
<td>53.33</td>
<td>60.63</td>
</tr>
<tr>
<td></td>
<td>(27.76)</td>
<td>(22.59)</td>
<td>(27.58)</td>
</tr>
</tbody>
</table>

Figure 5.20 Mandarin Classifier Mean Percent Accuracy by Most Frequently Used Language

A 3 x 4 mixed ANOVA shows that the main effect of group is not significant, $F(2, 19) = 1.62$, $p = .22$, $\eta^2 = .06$. The main effect of condition is significant, $F(3, 57) = 5.99$, $p < .01$, $\eta^2 = .16$, and was described in Section 5.5.2. The interaction between group and condition is not significant, $F(6, 57) = 0.77$, $p = .59$, $\eta^2 = .05$.

5.5.4 Summary of Classifier Results

In summary, there is a significant effect of condition across both instantiations of groups for Mandarin classifiers, with participants having significantly better accuracy on běn items than zhāng items,
and better performance on ɓęn items than ʈiǎọ items to a degree approaching significance. T-tests from
chance run on participant-averaged data show that all classifier sub-conditions are significantly different
from chance in the across-group comparison.

Furthermore, although the trends show that Mandarin home language children have a higher
across-condition average, a more differentiated performance by sub-condition, and higher accuracy on
ɓęn and ʈiǎọ than their Cantonese home language counterparts, there is no significant difference between
the groups, and no significant interaction between home language group and condition. Similarly, despite
the trend that children who most frequently speak Mandarin have a higher across-condition average than
children who most frequently speak Cantonese and English, there is no significant difference between the
groups, and no significant interaction between dominance and condition.

5.6 Sociolinguistic Variable Analysis

This section reports results for parent answers to the Home Language Questionnaire and
presents an analysis of the relationship between three sociolinguistic variables (socioeconomic status,
mother's English proficiency, and amount of time a child spent in China) and performance on the
Mandarin and English syntax tasks.

5.6.1 Child Results for Home Language Questionnaire

This section reports the results of the 17 child-centered questions of the Home Language
Questionnaire complete by parents of the 24 child participants in this study. Basic demographic data is
presented, followed by the measure of each child’s language input based on the Chinese variety spoken
in the home, and parent report of each child’s most frequently spoken language.

A profile of the child participants can be seen in Figure 5.21, below. Child participants for this
study are all four years old, ranging from four years, zero months to four years, nine months. The average
age is four years, four months, with a median of four years, three and a half months. 14 children are
female and 10 children are male. None of the children who participated in the study are reported to have
hearing loss. Two out of the 24 children are reported to use the services of a Speech-Language
Pathologist, but for both cases the parents specified that the reason for those services is to help the child
speak stronger English and catch up to monolingual English-speaking peers. Additionally, one child is reported to have selective mutism. The parents and teachers were eager for the child to participate in the study, and the child agreed to participate in the study by nodding.

**Figure 5.21 Child Participant Profile**

Ten of the 24 children have spent one year in a preschool setting prior to testing, 13 children have not spent a year in a preschool prior to testing, and results for one child are not reported. Parents reported whether the child has lived abroad and if so, where. The questionnaire prompted parents to indicate whether the child has spent a significant amount of time in a country outside the U.S. If the child has
spent time outside the U.S., parents were asked to circle one from among the options “less than six months”, “six months to a year” or “over a year”, as well as write in the country and the city. 14 children are reported not to have spent a significant amount of time in a country outside the U.S. Of the children who have spent a significant amount of time in a country outside the U.S., five spent under six months in China. Three of them spent that time in Fujian, one spent time in a region where Cantonese is spoken, and one spent time in a region which is neither predominantly Mandarin-speaking nor Cantonese-speaking. No children are reported to have spent between six months and a year living abroad. Five children are reported to have spent a year or more living outside the U.S. Two of them spent time in Taishan (a Cantonese-speaking area) and three spent time in other cities which are in neither predominantly Mandarin-speaking nor Cantonese-speaking regions.

The Home Language Questionnaire included inquiries about the languages spoken by and to the child participants. The questions were multiple choice, with the options of Mandarin, Cantonese, Fujianese, Shanghainese, Taiwanese, English, Other Chinese Variety (please specify) and Other Language (please specify). Parents indicated which individuals live with the child, and which languages those individuals speak to the child. From these answers, children were grouped by the Chinese variety spoken in the home, which is laid out as home language in Table 5.19 There was a nearly even divide between children with Cantonese and Mandarin as a home language. Additionally, parents reported which language the child speaks most frequently. There is an even distribution of children who most frequently speak Mandarin, Cantonese, and English, also reported in Table 5.19. Two children were reported to most frequently speak both Cantonese and English. For the analysis of most frequently spoken language, the results for these two children were compared to the Cantonese and English groups, and were included with the Cantonese group for the Mandarin classifier task and the English syntax test, and with the English group for the Mandarin syntax test.

---

8 Almost all children live with both parents and two-thirds of participants live with grandparents. 13 out of the 24 participants live with older siblings, nine live with younger siblings, and six children do not live with siblings.
9 Some families also speak some English in the home. Of the 11 children with a Mandarin home language background, five families are reported to speak Mandarin only in the home, and six report speaking Mandarin and English. Of the 13 children with a Cantonese home language background, six families report speaking Cantonese only, and seven report speaking Cantonese and English.
Table 5.19 Child Language Information

<table>
<thead>
<tr>
<th>Language Background</th>
<th>Mandarin</th>
<th>Cantonese</th>
<th>Mandarin &amp; English</th>
<th>Cantonese &amp; English</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most Frequently Spoken Language</td>
<td>11</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

5.6.2 Parent Results for HLQ

This section reports the results of the 10 parent-centered questions of the Home Language Questionnaire. Parent data collected includes socioeconomic status (education level and occupation), hometown, native language, and parents’ self-reported level of English. Relevant factors to the analysis are explained along with the justification for including these factors.

The socioeconomic status information collected in the Home Language Questionnaire includes mother’s level of education and occupation, and father’s level of education and occupation. Mother’s level of education has typically been used as the key socioeconomic status indicator (Hollingshead, 1975; Bornstein, Hahn, & Suwalsky, 2014), and is used the same way in this study. On the multiple choice questionnaire, the question about each parent’s level of education has six corresponding options: junior high school, high school, associate degree, bachelor’s degree, master’s degree or PhD. Results are shown in Table 5.20 below. A socioeconomic status (SES) measure is calculated based on mother’s level of education. The majority of participants are assigned to the middle socioeconomic status grouping. In the case of one participant, no information was collected for mother’s level of education, and that participant was assigned to the middle socioeconomic status grouping based on father’s level of education.

Table 5.20 Socioeconomic Status Results

<table>
<thead>
<tr>
<th>Education Level</th>
<th>Mother</th>
<th>Father</th>
<th>SES Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior High</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>High School</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Associate degree</td>
<td>4</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>PhD</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Occupation data was collected in the Home Language Questionnaire in short answer format (not multiple choice). Parent results constitute a wide range, from waiter and cashier at the lower socioeconomic status level, to post office worker and office assistant at the middle socioeconomic status level, and doctor and scientist at the higher socioeconomic status level. Overall, the socioeconomic status information collected in the Home Language Questionnaire showed a range of occupations and education levels, with more than half of the total participants in the middle socioeconomic status group. No answers for occupation or father’s level of education show a very different profile than mother’s education data.

Hometown and native language information for both parents was also collected on the Home Language Questionnaire. Hometown data was collected in short answer format, but fits into four groups by descending order of frequency reported: Guangdong, Fujian, other areas in China, and not China. It is not surprising that Fujian and Guangdong have the highest representation of parent hometowns within China. New York City Chinatown was established by Cantonese-speaking immigrants from Guangdong, while more recently, there has been increased immigration by individuals from Fujian.

The other measure related to parent data collected from the Home Language Questionnaire is parent level of English, reported in Table 5.21, below. On the multiple choice questionnaire, English proficiency for each parent was reported from among one of the following options: excellent, good, or limited. The majority of mothers and fathers are reported to have “good” English.

<table>
<thead>
<tr>
<th>English Proficiency</th>
<th>Mother</th>
<th>Father</th>
</tr>
</thead>
<tbody>
<tr>
<td>excellent</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>good</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>limited</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>not reported</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

5.6.3 Discussion of Child and Parent Language Results

This dissertation has emphasized the necessity of collecting specific information about a child’s home language variety in order to determine whether an assessment tool will measure structures that a child has been regularly exposed to. Because of the tendency to consider “Chinese” as a single variety, parents’ home language information was collected in addition to information about children’s language.
input and use. For the parent native language results, 15 sets of parents are reported to have Cantonese as a native language and nine sets of parents are reported to have Mandarin as a native language. However, based on the languages spoken in the household, 13 children were found to have a home language of Cantonese and 11 children with a home language background of Mandarin. This is because two sets of parents who report their hometown as Guangdong, a Cantonese-speaking region of China, and their native language as Cantonese, also report that they speak to their child in Mandarin at home. These children’s results were examined and they did not appear to be outliers within the Mandarin home language group. However, because of the prestige that is given to Mandarin over Cantonese, it does not seem unlikely that parents would intentionally or unintentionally inflate the amount of Mandarin exposure a child has because of the positive association with speaking Mandarin. The remaining 22 children are reported to have a home language that matches both parents’ native language, or for the cases where one parent was not born in China, the native language of the Chinese-speaking parent.

Of the 11 children with a home language background of Mandarin, eight are reported to most frequently speak Mandarin. This includes all five children whose family speaks Mandarin only, and also three out of the six children whose parents report that Mandarin and English are both spoken in the house. The other three children who are exposed to both Mandarin and English in the home are reported to speak English most frequently. Of the 13 children with a home language background of Cantonese, seven are reported to most frequently speak Cantonese. This includes all six of the children whose families are reported to only speak Cantonese, and one out of the seven children whose parents report that Cantonese and English are both spoken in the house. Four children whose parents report that Cantonese and English are both spoken in the house are reported to most frequently speak English. This means that of the sample of seven children reported to most frequently speak English, they are divided nearly evenly between Cantonese and Mandarin home language children. Finally, as mentioned above, two children whose families report speaking both Cantonese and English at home are reported to speak both English and Cantonese most frequently.
5.6.4 Correlations of Sociolinguistic Variables

Correlations were run to compare English and Mandarin performance on the syntax tasks overall, as well as on coordination and on RC structures separately, to see whether high performance in one language has a positive correlation with high performance in the other language. This result would be expected if knowledge of a structure in one language can facilitate understanding in another language, as was found in the comparison of first and second language syntax skills and their correlations with second language reading skills in a second-language dominant environment (Martohardjono et al., 2004, 2005; Gabriele, Troseth, Martohardjono, & Otheguy, 2009).

Results show that there is no significant correlation between English and Mandarin accuracy on RC and coordination structures combined (r(21)=.19, p=.39) or relative clause accuracy separately in English and Mandarin (r(21)=.08, p=.71). There is a medium correlation between English and Mandarin coordination accuracy that nearly approaches significance, r(21)=.35, p<.11. See Figure 5.22 below, which suggests that children who have higher accuracy on the Mandarin coordination structures also tend to have higher accuracy on English coordination structures.

**Figure 5.22 English and Mandarin Coordination Correlation**

![English and Mandarin Coordination Correlation](image)

The next series of correlations shows the relationship between key sociolinguistic variables and performance on the Mandarin syntax structures (coordination and RCs) and the English syntax structures (coordination and RCs). These variables are socioeconomic status, mother’s English proficiency, and
amount of time the child has spent in China. Correlations are run between each variable and performance on the combination of Mandarin coordination and RC structures, and between each variable and performance on the combination of English coordination and RC structures.

First, socioeconomic status is strongly correlated with accuracy on English structures ($r(22)=.54$, $p<.01$; see Figure 5.23, below) but not with accuracy on Mandarin structures ($r(21)=-.001$, $p=.99$). The result that socioeconomic status is strongly correlated with English performance has been a robust finding in previous literature (Fernald, Marchman, & Weisleder, 2013). Higher socioeconomic status has been connected with better educational performance overall, predicting a positive correlation between socioeconomic status and performance on the Mandarin syntax structures. However, Hoff & Core (2013) found that children’s home language performance is not improved by mother’s level of education because more educated mothers read to their children and encourage literacy and numeracy skills more often in English, even if they speak another language to the child. It is possible that that accounts for the lack of correlation between socioeconomic status and Mandarin syntax performance.

**Figure 5.23 Socioeconomic Status Correlation with English Accuracy**

Next, correlations show the relationship between mother’s proficiency in English and children’s performance on English coordination and RC structures combined. As expected, mothers’ self-reported English ability is significantly strongly correlated accuracy on the English syntax structures ($r(19)=.54$, $p<.05$; see Figure 5.24 below) and is not significantly correlated with accuracy on the Mandarin syntax structures ($r(18)=.17$, $p=.48$). The finding that children who have mothers with greater English proficiency
have higher accuracy on the English syntax task is expected, since the mother provides such a large amount of language input to the child (Hoff & Core, 2013). It could have been the case that mothers who are more proficient in English speak less Chinese to their children, in which case a negative correlation between mother’s English and Mandarin syntax accuracy might have appeared, but this does not seem to be the case.

**Figure 5.24 Mother’s English Proficiency Correlation with English Accuracy**

![Graph showing correlation between mother's English proficiency and English accuracy](image)

Finally, the amount of time a child spent in China nearly approaches a significant correlation with Mandarin syntax performance ($r(21) = .33$, $p = .12$, see Figure 5.25, below) and is not significantly correlated with accuracy on the English syntax task ($r(22) = -.19$, $p = .36$). These findings are also expected, since children who have spent more time in China likely have had more exposure to the home language than children who have spent their whole lives in the U.S., where the societal language is different from the home language.
This chapter presented results for child performance on Mandarin and English coordination and RC structures, and Mandarin classifiers. Children are grouped according to home language and most frequently spoken language as reported on the home language questionnaire administered to parents. Further sociolinguistic variables of mother’s English proficiency, child’s amount of time spent in China, and socioeconomic status were also described and compared to performance on the English and Mandarin syntax measures. The next chapter discusses the main findings from this investigation and offers recommendations for future studies.
6. Conclusion and Recommendations

While legislation requires home language testing in order to designate a child as having a language impairment, the dearth of home language assessments available to practitioners severely limits their ability to comply with the mandate. One of the goals of this study was to create a linguistically based assessment in a high-needs home language that could be extended into a practitioner’s tool in the future. This chapter discusses the major findings from the English and Mandarin assessments administered to children from Chinese-speaking households. The findings suggest the transfer of comprehension skills from the home language to the societally dominant language based on a set of cross-linguistically robust structures which have been attested as developmental milestones. This chapter also describes how performance on the assessments interacts with information gathered from the Home Language Questionnaire, in order to demonstrate the ways that children’s linguistic environment and language use must be considered when administering a language assessment. The chapter concludes with recommendations for implementing a home language assessment for Mandarin-English bilinguals on a larger scale, and wider implications for appropriate assessment practices in preschools.

The findings from this investigation broadly support results reported in the literature for coordination and RC hierarchies based on previous research conducted with English- and Mandarin-speaking children. This indicates a positive result for research question one, which asked whether children from Chinese-speaking homes exhibit the same asymmetries in acquisition hierarchies as established in the literature. Additionally, the highest accuracy was on coordination in both languages, and performance on coordination was better than performance on RCs in both languages, pointing not only to support for development of coordination before subordination, but also indicating the potential for transfer of comprehension such that comprehension of syntactic structures in the home language facilitates comprehension of similar structures in the school language. The role of transfer can be seen by the structural similarity of English and Mandarin coordination, and the similar results in those languages. RCs, in addition to being more syntactically complex than coordination, are typologically different across those two languages in that the RC follows the head noun in English but precedes it in Mandarin. Accuracy on English RCs was at chance while accuracy on Mandarin RCs showed trends toward the cross-linguistic subject gap / object-gap asymmetry pattern present coss-linguistically with higher
performance on subject-gap structures than object-gap structures. It is not possible to say for sure
whether the lower performance on the English RCs is due to less exposure, or the limited role of transfer
due to the structural RC differences, or a combination of the two. Classifier, coordination, and RC
hierarchies are each described in detail below.

The across-group hierarchy of classifiers (bĕn > zhī > tiáo > zhāng) suggests support for
Erbaugh’s (1992) theoretical hierarchy and experimental findings from a narrative task. The theoretical
hierarchy and experimental findings showed that children first acquire specific, common classifiers, such
as the one indicating books (bĕn). Next, classifiers that represent important (to the child) and animate
objects, such as the classifier for animals (zhī) are acquired. Within the shape categories, the long thin
object classifier (tiáo) is acquired before the flat object classifier (zhāng), though no reason is offered as
to why this might be the case. In this investigation, scores for the book classifier bĕn are significantly
higher than for the flat object classifier zhāng, and higher than long thin object classifier tiáo to a degree
approaching significance. Across group, accuracy is significantly higher than chance, though the average
scores are not at ceiling, showing that children are in the process of acquiring these structures. Children
who most frequently speak Mandarin have the same order of performance as was found across group;
however means for the group of participants who most frequently speaks Cantonese show a different
hierarchy of classifiers: bĕn > tiáo > zhī = zhāng. This may be due to differences in pronunciation of the
classifiers in Cantonese. No significant difference was found between home language groups or groups
by most frequently used language, and so a larger sample is needed to explore whether there is a
significant difference between the groups, and if so, what that difference is. With a larger sample, further
investigation into classifiers could test whether the difference between Cantonese and Mandarin speakers
is based on the phonetic differences between classifiers in each variety.

Children responded to IP coordination stimuli with higher mean accuracy than other types of
coordination in both Mandarin and English, supporting previous findings from Lust, Chien & Flynn’s
(1987) elicited imitation task. Furthermore, IP coordination is the only coordination sub-condition to have
a score significantly different from chance for the participant groups who most frequently use Mandarin
and Cantonese on the English RMST, which indicates that among coordination, IP structures emerge
earlier than the other structures. Scores for IP coordination are significantly higher than for verb
coordination on the Mandarin syntax task and higher than for verb coordination on the English syntax task to a degree approaching significance. Verb coordination also has a lower mean accuracy than the other coordination types, and is the only coordination sub-condition not to have a score significantly different from chance for the Cantonese home language group on the Mandarin RMST. (The Mandarin home language group scored above chance on all coordination structures.) Taken together, these results suggest a hierarchy of IP > NP > VP coordination.

The final structures explored in the assessment are RC constructions. As expected, children were less accurate on subordination than coordination. The RC subject-gap / object-gap asymmetry found in previous studies was replicated on the Mandarin task, with children having higher accuracy on subject-gap RCs than object-gap RCs. This offers further evidence for the findings from Hu et al.’s (2016) Mandarin picture-point comprehension task, which found higher success rates for subject-gap structures over object-gap structures for young children. Hu’s task had only the subject-gap/object-gap dichotomy, and this investigation shows that the subject-gap/object-gap asymmetry seems to hold even when embeddedness is manipulated. In addition to subject-gap structures having higher mean accuracy than object-gap structures in this investigation, only subject-gap structures, and not object-gap structures, have scores significantly different from chance for the Mandarin home language group and the group of participants to most frequently use Mandarin. The same was not replicated for the Cantonese home language group, where accuracy is not highly differentiated by RC sub-condition, and the means for the subject-gap structures are closer to the means for the object-gap structures than for the Mandarin home language group. This is not surprising: Cantonese children were not expected to perform similarly to Mandarin home language children on complex structures given the pronunciation differences between Mandarin and Cantonese (the RCs are structured the same in both varieties).

For the English syntax task, accuracy on the RC sub-conditions is not significantly different from chance except for OS RCs. Because most sub-conditions were at chance, further differences between English RC structures were not explored. In the case of English RCs, the hierarchy of acquisition and the subject-gap / object-gap asymmetry is not explicitly supported, but since the only structure with a score above chance was OS (a subject-gap structure), the trends suggest development in a similar direction. The development of OS RCs before SS RCs follows Slobin’s (1971) findings that object-embedded RCs
are acquired earlier than subject-embedded RCs in English. Scores at chance suggest that the participants may not yet have acquired the English RCs, which would be expected if as a group, the children have not yet had enough exposure to English to develop the English RCs to the same level as the monolinguals who are shown to have the asymmetry at age four. With increased input and frequency of use, and general development, the asymmetry will likely emerge, followed in time by command of all four RC types.

The first research question was therefore broadly supported by the data. Children on this assessment had similar hierarchies as those established in the literature for English and Mandarin coordination, Mandarin classifiers, and Mandarin RCs. While there was not evidence of a hierarchy for the English RCs, there is also no evidence against the hierarchy, and since the highest performance was on OS RCs, the trend suggests development in the same direction.

The second research question asked what different patterns emerge when performance data are analyzed by group according to a child’s home language or most frequently used language. In order to explore this question, participant data were analyzed by home language and by most frequently spoken language to explore group differences.

As discussed in Chapter 3, “Chinese” is used broadly in the U.S. as a catch-all term that includes a large number of very different language varieties. Although Mandarin is the official language and the language of education in the P.R.C, not all “Chinese” speakers use Mandarin at home. Four-year old children in the P.R.C and the U.S. will not necessarily have been exposed to Mandarin, and there is no reason to expect that children who grow up hearing a non-Mandarin Sinitic variety in the home would perform similarly to children who grow up in Mandarin-speaking homes on a Mandarin language assessment. During recruitment for this study, participants from “Chinese-speaking” homes were recruited in order to get a representative sample of the linguistic population who together make up 13% of English Language Learners in New York City public schools. The Home Language Questionnaire administered to parents served to collect more nuanced information about children’s language input and use. Parents of children in this sample reported speaking either Mandarin or Cantonese in the home. In addition, some parents reported also speaking other Sinitic varieties, English, and in a few cases, Romance languages. This allowed for analysis by separating data into groups based on whether children
spoke Mandarin or Cantonese in the home. It was found that not all children from “Chinese-speaking” homes performed similarly on the Mandarin assessment. When results for the Mandarin syntax task were analyzed by home language group, the Mandarin home language group significantly outperformed the Cantonese home language group on the Mandarin coordination and RC structures. The Mandarin home language group average accuracy was also generally higher than Cantonese home language accuracy on the Mandarin language structures. The decision to include children who do not speak Mandarin in the sample reflects the necessity to quantifiably demonstrate how inappropriate the Mandarin task is for children from a non-Mandarin speaking household. In the case that a future version of this assessment tool becomes available for practitioners it is our hope that the evidence showing the difference in home language group performance will prevent the Mandarin language assessment being used inappropriately by practitioners. In order to implement appropriate home-language assessment practices, it is critical to have an accurate understanding of what a child’s home language is in order to ensure that there is not a mismatch between the actual home language variety and the variety being tested.

While it was expected that the variety of Chinese spoken in the home would predict performance on the Mandarin assessments, it was not expected that the Chinese variety a child spoke would predict performance on the English syntax task. As predicted, children from Mandarin and Cantonese home language groups performed similarly to each other on the English assessment. There was no significant difference of home language group on the English coordination structures. (RC structures were not examined by group since the sub-conditions were mostly at chance.)

In order to examine performance on the English syntax task, participant data were divided according to which variety each child was reported to most frequently used (Mandarin, Cantonese, or English). Participants who most frequently speak English significantly outperformed participants who most frequently speak Mandarin or Cantonese on the English coordination structures. (Again, English RCs are not reported by group because performance on three out of four sub-conditions were not different from chance.) On the Mandarin syntax tasks, most frequently used language was not significant for coordination. For the Mandarin RCs, participants who most frequently speak Mandarin did significantly outperform participants who most frequently speak English, though there were no significant differences from either group reported for participants who most frequently speak Cantonese.
Taken together, the results provide further evidence that children who are broadly grouped by schools and practitioners as “Chinese” speakers are a heterogeneous group in terms of both variety of Chinese spoken in the home, and also in terms of whether they most frequently use a variety of Chinese or English. Although the sample size for this study is rather small, it is interesting to note that participants were not recruited to be evenly subdivided by Mandarin / Cantonese home language, or by most frequently spoken language; and yet the distribution was quite even. 11 out of the 24 participants are reported to speak Mandarin in the home, with the remaining 13 speaking Cantonese. Within each of these groups, the number of households that use a variety of Chinese exclusively (5 Mandarin households and 6 Cantonese households) is similar to the number of households that use English in addition to a variety of Chinese (6 Mandarin households and 7 Cantonese households). Finally, the distribution of most frequently spoken language is also similar, with 8 children reported to most frequently speak Mandarin, 7 children reported to most frequently speak Cantonese, and children reported to most frequently speak English, with 2 children reported to most frequently speak both Cantonese and English. It is therefore critical that home language differences and frequency of use differences be considered when administering language assessments to the population of “Chinese-speakers”, as the population is heterogeneous and these factors make a difference on assessment performance.

In addition to home language and most frequently used language, extralinguistic factors that influence performance on English and Mandarin assessments were also considered. The Home Language Questionnaire collected information about socioeconomic status, mother’s English proficiency, and amount of time children spent in China. Correlations were run to explore relationships between these factors and performance on English and Mandarin syntax tasks. Socioeconomic status, which for this investigation was calculated based on mother’s level of education, has long been known to be a factor that influences performance on English language assessments for monolinguals as well as bilinguals. A positive correlation between socioeconomic status and performance on the English coordination/RC structures showed the same pattern. However, no relationship was observed between socioeconomic status and performance on the Mandarin coordination/RC structures. This finding follows previous research such as Hoff & Core (2013), which found that children’s home language performance is not improved by mother’s level of education. According to Hoff & Core, this is because more educated
mothers read to their children and encourage literacy and numeracy skills more often in English, even if they speak another language to the child.

For the other two extralinguistic variables that were considered, mother’s English proficiency was expected to have a positive relationship with English syntax performance, and time in China was expected to have a positive relationship with Mandarin syntax performance. In both cases, the predictions were supported by the data, though the relationship between time in China and Mandarin syntax performance only approached significance. Correlations were also run to see if there might be a negative relationship between time in China and English performance, or mother’s English proficiency and Mandarin performance, which would be expected if increased input in one language were to somehow negate input in the other language. As language acquisition is not a zero-sum game, and following the expectation that knowledge of one language is not detrimental to the acquisition of another, we did not expect such findings and indeed the correlations showed no support for those negative relationships.

Finally, correlations were run to explore any relationships between performance on Mandarin and English syntax structures. Correlations showing a positive relationship between performance on Mandarin syntax tasks and English syntax tasks would suggest that the development of coordination or RCs in one language would facilitate development in the other language. There is evidence for this syntactic transfer from home language to societally dominant language, such as the positive correlation Martohardjono et al. (2005) found between Spanish and English RC accuracy on a similar syntax task in kindergarten students. This study showed weak support for syntactic transfer of comprehension with a nearly significant positive correlation between performance on coordination structures in English and Mandarin. The weakness of the correlation may be due to the overall heterogeneity of the home language and most frequently used languages in the sample. Additionally, no relationship was found between performance on the English and Mandarin RC structures, likely because the English RCs were mostly at chance. We do not have evidence as to whether the lower performance on English RCs is due to children having less exposure to English in the home, or because RCs are typologically different in Mandarin and English, or a combination of both of those.

For both the Mandarin and English syntax tasks, means on the coordination condition were higher than means on the RC condition, both for the across sub-condition averages and also for the
trends within the sub-conditions. This is expected, since coordination develops before subordination. It’s interesting to note that while the across-group average for coordination is similar for the English coordination task and Mandarin coordination task (58% for each); the across-group average for RCs are higher on the Mandarin RC task (50%) than the English RC task (39%). This shows later development of complex structures in English, likely due to less frequent or less long-term exposure to English.

The purpose of this investigation was to serve as a pilot study for an assessment that could be administered to children from Chinese-speaking homes in preschools. Some of the results from this study have immediately applicable implications. First, the investigation provides further evidence that home language testing is essential for bilingual students, as across-group results showed higher accuracy rates on Mandarin RCs than English RCs despite the Home Language Questionnaire findings that English is spoken in more than half of the homes, and that a third of the participants speak English most frequently. However, home language testing is not enough; it is vital that the home variety be specified so that children are not held to unreasonable standards in Mandarin when that is not the language they are exposed to in the home.

In future versions of this research, it is recommended to incorporate additional information into the Home Language Questionnaire used for this study. The ALEQ (Paradis, Emmerzael, & Sorenson Duncan, 2010; Paradis, 2011) measure used by Jia & Paradis (2018) asks parents about the ratio of each language each member of the family speaks to the child, as well as the ratio of each language the child speaks to each member of the family. Collecting this information for each of the interlocutors the child spends significant time with can allow for a more accurate measure of language input and use than were collected in this study. Furthermore, Byers-Heinlein et al. (2018) have found that whenever possible, parent report of children’s language exposure and use should be administered through structured interviews. Using the MAPLE approach outlined by Byers-Heinlein et al. can increase reliability of parent report.

Some modifications to the assessment are recommended before it could be distributed for widespread use. First, while children generally enjoyed using the tablets and were familiar with their use, there were cases where children would at times touch one response picture, but swipe with a motion that released on a different picture, causing the incorrect picture to be recorded. The experiments were coded
with E-Prime, which only has the functionality to record a response on the picture indicated by the release. For future use of the tablet test, it is recommended to try different methods to ensure greater accuracy by recording responses at the picture point instead of the release. Furthermore, one teacher at a participating school suggested that actions like kicking and punching be eliminated from the stimuli since learning to treat other people kindly is a skill emphasized and practiced in preschool. Such examples were chosen because of the need to use active verbs that can be clearly portrayed in pictures. However, further conversations with teachers during the development may prevent the inclusion of stimuli that teachers would find inappropriate. Buy-in from educators is important if the assessment is to actually be used in the schools. Future development must also consider whether to include Cantonese structures as part of the assessment, or to create an assessment entirely in Cantonese, or whether the assessment should exclude children from homes where a non-Mandarin variety is spoken altogether. Finally, in order for a home language assessment to have any validity, it must undergo an extensive norming process on a large number of Chinese-English bilingual children who are representative of the children who will use the assessment.

In order for schools to fulfill their legal obligation to conduct home language testing, they need to have home language assessments that can be administered by a monolingual educators, where the tool includes instructions and prompts in home language. The assessment presented in this study could be further developed to include additional target structures such as tense in English and aspect in Mandarin, and normed on a bilingual population to ensure validity. The tablet test could be expanded so that an English-language report could be automatically generated so that educators could understand the results and provide appropriate scaffolding. Such a tool could become be a practical solution to the very real lack of appropriate home language assessment tools and could improve the educational experience for children by giving them the opportunity to demonstrate what they do know instead of holding them to the unreasonable expectations for English development that drive the deficit model of education.
7. Appendices

7.1 Appendix 1: Detailed Description of Materials Presented to Participants

7.1.1 English RMST

The English version of the RISLUS Multilingual Syntax Test (RMST) (Klein & Martohardjono, 2013) consists of four types of constructions: coordination structures, RCs, temporal adverbials, and control structures. Participants are introduced to all of the characters in English, and then are trained on the task with five practice items. Five items are included for each of 12 sub-conditions. The 60 items are divided among 3 tests with 20 items each. The structure types for each condition, with one example item per structure, are as follows:

1. IP coordination
   ‘The bear swims and the dog walks’

2. Subject coordination
   ‘The bear and the dog chase the cat’

3. Object coordination
   ‘The cat chases the bear and the dog’

4. Verb coordination
   ‘The rabbit hits and kicks the cat’

5. Subject-subject RC
   ‘The bear, who touches the dog, dances’

6. Subject-object RC
   ‘The dog, who the bear punches, dances’

7. Object-subject RC
   ‘The cat pushes the bear, who holds the monkey’

8. Object-object RC
   ‘The bear touches the monkey, who the dog hugs’

9. Natural temporal adverbial
   ‘After swimming, the bear hugs the monkey’

10. Reverse temporal adverbial
‘The cat pushes the rabbit, after sleeping’

(11) Object control
‘The dog tells the rabbit to run, and he does’

(12) Subject control
‘The rabbit promises the dog to run, and he does’

7.1.2 Mandarin RMST

The Mandarin version of the RISLUS Multilingual Syntax Test (RMST) (Klein & Martohardjono, 2013 adapted for this dissertation) consists of three types of constructions: coordination structures, RCs, and temporal adverbials. Participants are introduced to all of the characters in Mandarin, and then are trained on the task with five practice items. Five items are included for each of nine sub-conditions. The 45 items are divided among 3 tests with 15 items each. The structure types for each condition, with one example item per structure, are as follows:

(1) IP coordination

小熊在走路而小狗在游泳
xiao xiong zai zou lu er xiao gou zai you yong
[little bear PROG walk] CNJ [little dog PROG swim]
‘The bear is walking and the dog is swimming’

(2) Subject coordination

小熊和小猫追着小狗
xiao xiong he xiao mao zhui zhe xiao gou
[little bear CNJ little cat] chase DUR little dog
‘The bear and the cat are chasing the dog’

(3) Object coordination

小狗拉着小熊和小猴
xiao gou la zhe xiao xiong he xiao hou
Little dog pull DUR [little bear CNJ little monkey]
‘The dog is pulling the bear and the monkey’

(4) Verb coordination

小狗一边跳着一边抱着小熊
xiao gou yi bian tiao zhe yi bian bao zhe xiao xiong
dog [one side jump DUR one side hug DUR little bear]
‘The dog is jumping and hugging the bear’
(5) Subject-subject RC

跳舞的 小狗 在 打 小熊
tiao wu de xiao gou zai da xiao xiong
[dancing REL little dog] PROG hit little bear
‘The dog who is dancing is hitting the bear’ / [The dancing dog is hitting the bear]

(6) Subject-object RC

小熊 摸 着 的 小猴 抱 着 小狗
xiao xiong mo zhe de xiao hou bao zhe xiao gou
[little bear touch DUR REL monkey] hold DUR little dog
‘The monkey who the bear is touching is holding the dog’ /
[The bear-touched monkey is holding the dog]

(7) Object-subject RC

小熊 抱 着 跳 的 小狗
xiao xiong bao zhe tiao de xiao gou
little bear hug DUR [jump REL little dog]
‘The bear is hugging the dog who is jumping’ / [The bear is hugging the jumping dog]

(8) Object-object RC

小猴 抱 着 小猫 推 着 的 小熊
xiao hou bao zhe xiao mao tui zhe de xiao xiong
little monkey hug DUR [little cat push DUR REL little bear]
‘The monkey is hugging the bear who the cat is pushing’ /
[The monkey is hugging the cat-pushed bear]

(9) Temporal adverbial

跑步以后，小熊 抱 小猴
pao bu yihou xiao xiong bao xiao hou
run after little bear hug little monkey
‘After running, the bear hugs the monkey’

7.1.3 Mandarin Classifiers

(1) Book classifier

请 指出 那 本
qing zhichu na ben
please indicate that CL
‘Please point to that (book)’

(2) Animal classifier

请 指出 那 只
qing zhichu na zhi
please indicate that CL
‘Please point to that (animal)’

(3)  Long thin object classifier

请指出那条
qing  zhichu  na  tiao
please indicate that  CL
‘Please point to that (long thin object)’

(4)  Flat object classifier

请指出那张
qing  zhichu  na  zhang
please indicate that  CL
‘Please point to that (flat object)’
References


https://doi.org/10.1080/02699200010017805

http://www.sciedirect.com/science/article/pii/S0022537174800642


https://doi.org/10.1017/S0305000908009185


