DEVELOPING VERY SHORT COMPUTERISED EARLY VOCABULARY ASSESSMENTS IN A MULTILINGUAL SETTING

Jun Ho CHAI

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

in Psychology

September 2021

Abstract

The MacArthur-Bates Communicative Development Inventories (CDIs) are one of the most widely adapted sets of parent-report instruments for assessing young children's early vocabulary acquisition (Fenson et al., 2007; Frank, Braginsky, Marchman, & Yurovsky, 2021). The MacArthur-Bates Communicative Development Inventories (CDIs) are valid and reliable not only with children who are developing typically (Fenson et al., 2007, 1993; Pan, Rowe, Spier, & Tamis-Lemonda, 2004; Rescorla, Ratner, Jusczyk, & Jusczyk, 2005; Law & Roy, 2008), but also with children developing atypically (Galeote, Checa, Sánchez-Palacios, Sebastián, & Soto, 2016; Luyster, Lopez, & Lord, 2007; Mayne, Yoshinaga-Itano, Sedey, & Carey, 1998; Mayne, Yoshinaga-Itano, & Sedey, 1999; D. Thal, DesJardin, & Eisenberg, 2007). The CDIs provide extensive insights into children's vocabulary sizes, yet they rely on parents' knowledge of their children. Furthermore, the completion of each CDI is time-consuming as the parent has to assess their child's word knowledge item-by-item, often amounting to 600 words or more. Although previous efforts were made to develop 100-item short-form versions, it is time-consuming to develop short-forms for each language, and their administrations remain a considerable time investment for parents when completing such forms. The process of completing CDIs is made even more laborious, when having to fill out CDIs for several languages, for children exposed to more than one language, such as most children from Malaysia, a linguistically diverse country, where bi- and multilingually-exposed children constitute a significant proportion of the population.

Hence, the objective of the current project is two-fold: 1) to collect the first Malaysian early vocabulary data using an adaptation of the CDI forms and provide insights into Malaysian children's early language development; and 2) to develop CDI-based novel vocabulary assessment tools, with the aim of improving the administration of CDIs and examining the viability of a CDI-based toddler-directed word comprehension assessment tool using tablets to supplement the parental reports.

The first part of the thesis adopted the MacArthur-Bates Communicative Developmental Inventories - Malaysian version (MCDI-M), a trilingual form developed by Low (2009) to examine the early vocabulary trajectories of young Malaysian children, along with a quantification of environmental effects on the language development of these children, and examining the composition of early vocabularies, i.e., the acquisition of nouns and verbs.

The second part of the thesis introduced two novel early vocabulary assessment tools. First, a language-general, Bayesian-inspired item response theory-based framework was developed – based on prior work by Mayor and Mani (2018) and Makransky, Dale, Havmose, and Bleses (2016) (MM-IRT) in order to reduce the number of items needed to assess children's language. This framework was evaluated using data sampled from the Wordbank database in English, Danish, Mandarin and Italian. The framework was then applied to two sets of MCDI-M dataset to evaluate the possibility of using this set of innovations in the context of multilingual children in Malaysia. Second, a toddler-based version of CDIs, using tablets, was evaluated with the aim of directly assessing each child's language skills, and was evaluated empirically.

In sum, the first part of this thesis found the use of MCDI-M as an effective early vocabulary assessment tool for Malaysian children, whereas the second part of this thesis paved the way towards the application of rapid yet robust language assessments powered by MM-IRT, to be used in clinical settings, and demonstrated the viability and convergence of using tablets for direct-measure of language skills.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr Julien Mayor, Dr Tze Peng Wong and Dr Hui Min Low, for offering me the wonderful opportunity to work on the project, as well as for all the meaningful discussions, support, and guidance over the years. I would like to thank all the parents who contributed their time to filling out the vocabulary forms. I would like to thank my on-campus supervisor, Steve Stewart-Williams for the monthly meetings during the early years. I would like to thank my colleagues who provided support, feedback and discussion. I would like to thank my parents and my brother for their companionship.

Lastly, I would like to thank my wife, Joey, and my daughter, Xeni, for all the joy and accompanying me through thick and thin!

This work was supported by the MOSTI R&D Fund Projects (06-02-12-SF0351) awarded to Prof. Dr. Julien Mayor.

Contents

List of Acronyms

CDIs MacArthur-Bates Communicative Development Inventories

CDI-WG CDI-Words and Gestures

CDI-WS CDI-Words and Sentences

MCDI-M Multilingual Communicative Development Inventories – Malaysia version

SES Socio-Economic Status

LMM Linear Mixed Model

ZINBM Zero-Inflated Negative Binomial Models

CDI-SF Shorter Form of CDIs

IRT Item Response Theory

EM Expectation-Maximization

WLE Weighted Likelihood Estimation

CDI-CAT Computerised Adaptive Testing version of CDI

MM-IRT IRT version of Mayor and Mani (2018)'s approach

CCT Computerised Comprehension Task

2AFC Two-Alternative Forced Choice

GLMM Generalised Linear Mixed Model

List of Tables

List of Figures

Chapter 1

Introduction

The objective of this thesis is two-fold: 1) to collect and provide insights into Malaysian children's early language development using an adaptation of the CDIs forms (Studies 1-3); 2) to introduce a CDI-based novel vocabulary assessment tool as an improvement to the CDI and to explore the implementation of the CDI in a toddler-directed tool using a tablet (Studies 4-6). Chapter 2 serves as a background chapter, reviewing the state of knowledge about early vocabulary acquisition; the internal and external factors that modulate the process of acquiring one or two languages. Chapter 3, an experiment chapter that utilised a trilingual adaptation of the CDIs, reported the first Malaysian early vocabulary data and gave insights into Malaysian children's early language development. This chapter examined vocabulary growth, and the role of gender, language exposure, Socio-Economic Status (SES), linguistics, and sociocultural factors in early vocabulary acquisition. To improve the efficiency of CDIs, Chapter 4 introduces the development of a language-general, novel computational framework capable of producing very short versions of CDIs without compromising their efficacy. In addition, to supplement the CDIs, Chapter 4 examined the viability of a tablet-based comprehension assessment tool for early vocabulary acquisition using a two-alternative-forced-choice task.

1.1 Background

Young children possess an exceptional capacity to learn new words at a rapid pace. Even before their first birthday, *infants* (referred to as children from birth to 1-year-old) are capable of understanding common words (6 – 10-month-old; Schafer, 2005; Bergelson & Swingley, 2012; Kartushina & Mayor, 2019). Then, they typically begin to utter their first words at around 12-month of age (Bloom, 1993; Frank et al., 2021). Before the *toddler*s' (referred to children between 1- to 2-year-old) second birthday, from about 18 – 24-months of age, the rate of word acquisition undergoes a rapid increase (known as the *vocabulary spurt*; Bloom, 1973) and toddlers start to combine words into pairs (e.g., 'mommy sock'; Marchman, Martínez-Sussmann, & Dale, 2004).

The implication of children's word learning process cannot be overstated. It acts as the building block of children's future linguistic and cognitive development. In terms of linguistic development, children's early vocabulary size predicts later vocabulary growth (Bornstein, Tamis-LeMonda, & Haynes, 1999; Fenson et al., 1994); predicts their achievement on standardised tests of language during school age (Marchman & Fernald, 2008); and predicts grammatical development, in longitudinal (Dionne, Dale, Boivin, & Plomin, 2003) and cross-sectional studies (Marchman et al., 2004). In addition, early language skills predict cognitive skills, such as working memory and IQ (Marchman & Fernald, 2008), numerical knowledge (Carey, 1994), theory of mind (Schick, De Villiers, De Villiers, & Hoffmeister, 2007; Astington & Baird, 2005) and spatial skills (Gentner, Özyürek, Gürcanli, & Goldin-Meadow, 2013).

Early language development, in the form of early vocabulary acquisition, is strongly modulated by both the quantity and the quality of child-directed speech (Rowe, 2012; Hart & Risley, 1995). Interestingly, using computational modelling and mathematical derivations, Mayor and Plunkett (2010) showed that word frequency alone cannot fully explain the acceleration in the vocabulary growth observed towards the end of the second year of life (Bloom, 1973; Fenson et al., 1994) and that endogenous changes (learning capacity) in the learner are required to account for the vocabulary spurt. In addition, exogenous changes (in the caregivers' language) interact with endogenous changes, as parents are found to adjust the complexity of the language spoken to the children according to the competency of the children (Pellegrini, Brody, & Sigel, 1985), in line with the *transactional model* of early social environment proposed by Sameroff (1975).

In summary, the astonishing growth in early word knowledge is a fascinating feat driven by the influence of a multitude of factors on the learning environment of the child and on the improving cognitive skills taking place during early childhood. Further, investigations into early vocabulary acquisition using tools such as parent reports provide extensive information about the word knowledge of their young children, thus enabling insights into the broader patterns of language learning.

1.2 Outline of the Thesis

Chapter 2 introduces the process of early vocabulary acquisition and discusses internal and external factors which modulate this process. In this chapter, I will discuss findings regarding early vocabulary trajectory, i.e., vocabulary trends across age, and gender differences in children's early vocabulary based on CDI and other studies. I will discuss how external factors such as socio-economic status and differences in parental attention patterns can impact vocabulary development. I will also explore the connections between parental attention patterns, language-specific factors and the cross-cultural-linguistic differences in noun-bias. This chapter will introduce key findings that are directly related to the studies reported in Chapter 3, which utilises a trilingual version of CDI in a linguistically diverse country, Malaysia. This chapter also serves as a foundation for the studies reported in Chapter 4, which introduces two novel early vocabulary assessments. The first two studies focus on the development of a short and computerised version of the CDIs, and the third study focuses on the development of a toddler-directed, tablet-based tool that can supplement parental reports (CDIs).

Chapter 3 explores the use of a trilingual version of the CDIs in a linguistically diverse country, Malaysia. I will report the first Malaysian early vocabulary data and provide insights into Malaysian children's early language development. The first study investigates the effectiveness of the trilingual CDI in capturing developmental trends captured in previous research which used CDIs in a monolingual context. In that chapter, I evaluate vocabulary growth through development, between 8 and 36 months of age, and the role of relative language exposure and gender in modulating the vocabulary trajectory. (Study 1). It is noteworthy that the data used is cross-sectional and not longitudinal – any observed trend results from a collective of individuals and not a direct observation of change/growth within individuals. The second and third studies quantify a series of environmental effects on the language development of Malaysian children and compare them with previous studies. Specifically, I investigate the role of socio-economic status (i.e., education and income) in early vocabulary acquisition (Study 2). I also explore cross-cultural and cross-linguistic differences in the learning of verbs and nouns (Study 3). I will conclude with considerations concerning the insight we obtained from the trilingual CDIs and whether it matches previous literature on early vocabulary acquisition, as well as the role of internal and external factors documented in previous research.

Chapter 4 innovates on existing assessment techniques to create novel and cost-efficient tools that facilitate the application of rapid yet robust CDI-based early vocabulary assessments to be used in clinical settings. I will introduce a novel, shorter version of CDIs using a Bayesian-inspired item response theory-based computational modelling framework (MM-IRT model; based on Mayor & Mani, 2018; Makransky et al., 2016) and will report an assessment of its effectiveness using normative data from the Stanford Wordbank open database (Frank, Braginsky, Yurovsky, & Marchman, 2017) (Study 4). In this new computerised language assessment, the model dynamically selects words that optimally discriminate the language skills of children during the test, thus reducing the number of test items to a practical minimum. In addition, I will explore how this model can be effective on a much smaller data sample, i.e., the sample we collected from the Malaysian population (Study 5). I will also explore the viability of using tablets in assessing early word comprehension by means of implementing a two-alternative forced-choice task that is attended directly by toddlers, using items from the CDI forms (Study 6). This toddler-based assessment can potentially serve as a supplemental and convergent measure of vocabulary sizes, alongside parental reports. I will conclude with a discussion of the effectiveness and limitations of both novel early vocabulary assessment tools, in terms of the model's ability to map to full-sized vocabulary assessments (CDIs) (Study 4), the model's ability to generalise to a smaller sample, the Malaysian sample (Study 5) and the tablet assessment tool's ability to index children's vocabulary knowledge, along with a measure of how it compares with full-CDI forms.

Chapter 5 will conclude by reviewing our understanding of early vocabulary acquisition introduced in Chapter 2, in light of the empirical results introduced in Chapter 3. I will summarise the contributions of the chapters and discuss the utility of the trilingual CDIs in assessing the early vocabulary of Malaysian children, as well as providing a theoretical contribution within the greater context of research on early vocabulary acquisition. In addition, I will review the novel early vocabulary assessment tools and their effectiveness in light of the results from Chapter 4. I will discuss the contribution of these tools in paving the way towards the application of rapid yet robust language assessments to be used in clinical settings, in particular in a linguistically diverse context such as Malaysia. I will end the chapter by briefly discussing possible avenues for future work, including large-scale collection of Malaysian early vocabulary data, validating the MM-IRT model with empirical data, and integrating the MM-IRT model into the toddler-directed tablet-based word comprehension task.

1.3 Published Work

Parts of the material presented in this thesis have now been published. This concerns the following studies; Chapter 3: Study 3, Chapter 4: Study 4, and Chapter 4: Study 6. The papers are attached in the Appendix.

The research that examines potential sources of the noun-bias, i.e., language-specific features and extra-linguistic factors, is published as; Chai, Low, Wong, Onnis, and Mayor (2021) and is reported in Chapter 3: Study 3.

Chapter 4: Study 4 built on innovative ways of assessing vocabulary using CDI, by combining CDI-based computerised adaptive testing developed by Makransky et al. (2016) – which adaptively selects items based on the participant's responses, thereby reducing the number of test items, yet appropriately reflects the participant's knowledge; and a Bayesian-inspired model of vocabulary acquisition developed by Mayor and Mani (2018) – which extrapolates full-CDI scores from a subset of test items, yet closely matched with original full-CDI scores. The novel language assessment "creates an efficient and less data-hungry adaptive technique" (Kachergis, Marchman, Dale, Mankewitz, & Frank, 2021), and is published as; Chai et al. (2021).

The experiment in Chapter 4: Study 6 presented with the aim of examining the viability of using tablets in assessing early word comprehension by means of a two-alternative forced-choice task, is published as; Lo, Rosslund, Chai, Mayor, and Kartushina (2021).

In summary, this thesis contains 3 published work:

- Study 3 was peer-reviewed and published in the Journal of Cultural Cognitive Science, and was included in this thesis (section 2.3 & 3.7):

  Chai, J. H., Low, H. M., Wong, T. P., Onnis, L., & Mayor, J. (2021). Extra-linguistic modulation of the English noun-bias: evidence from Malaysian bilingual infants and toddlers. *Journal of Cultural Cognitive Science*, 5(1), 49-64, https://doi.org/10.1007/s41809-021-00078-5

  - Contributed in conceptualising the study, data collection, data analyses, interpretation of data, discussion of findings and writing.

- Study 4 was peer-reviewed and published in the Journal of Speech, Language, and Hearing Research (JSLHR), and was included in this thesis (section 4.2 & 4.5):

  Chai, J. H., Lo, C. H., & Mayor, J. (2020). A Bayesian-Inspired Item Response Theory–Based Framework to Produce Very Short Versions of MacArthur–Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research*, 1-13,

  https://doi.org/10.1044/2020_JSLHR-20-00361

  - Contributed in conceptualising the study, coded adaptations and improvements of the algorithm, data analyses, interpretation of data, discussion of findings and writing.

- Study 6 was peer-reviewed and published in the Infancy journal, and was included in this thesis (section 4.3 & 4.7):

  Lo, C. H., Rosslund, A., Chai, J. H., Mayor, J., & Kartushina, N. (2021). Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers. *Infancy*, 26(4), 596-616,

  https://doi.org/10.1111/infa.12401

  - Contributed in data analyses, interpretation of data, discussion of findings and writing.

Chapter 2

Early Vocabulary Acquisition

2.1 Variability and Consistency

Early vocabulary acquisition is influenced by both internal (e.g., age, gender and individual differences) and external factors (differences in input quantity and quality), contributing to the considerable variability in early vocabulary acquisition rate among young children (Fenson et al., 1994; Feldman et al., 2000). Differences in language acquisition have been reported between girls and boys, with a small yet significant girl advantage in receptive vocabulary and productive vocabulary, consistently observed across languages (see Bleses et al., 2008a; Bornstein, Hahn, & Haynes, 2004; Eriksson, 2017; Fenson et al., 1994; Galsworthy, Dionne, Dale, & Plomin, 2000; Schults & Tulviste, 2016; Simonsen, Kristoffersen, Bleses, Wehberg, & Jørgensen, 2014; Stolt, Haataja, Lapinleimu, & Lehtonen, 2008; Tardif, Fletcher, Liang, & Kaciroti, 2009). Other girl advantages include grammatical knowledge (Bleses et al., 2008a) and verbal skills (Leaper & Smith, 2004).

In addition, one of the internal factors that modulate language acquisition lies in stylistic differences observed among children (Fenson et al., 2007). Historically, young children can be grouped into two groups according to their language learning style, *referential* or *expressive* (E. Bates et al., 1994; Nelson, 1973). Children in the *referential* group tend to acquire more nouns than words from the other semantic categories, produce speech that is less syntactically complex, and are able to learn vocabulary at a quicker pace than children in the *expressive* group, who tend to acquire fewer nouns and produce speech that is more syntactically complex. It is argued that the referential/expressive difference is related to the analytic/holistic differences in attentional structure – children who attend to language *holistically* (focusing on syllables and phonemes or prosodic tunes) tend to be *expressive*, children who attend to language *analytically* ("focusing on smaller units then combining them") tend to be *referential* (Hoff, 2013).

Other than that, external factors such as parenting differences in language input are found to influence the vocabulary composition of their children – mothers who produce more word tokens tend to have children with vocabulary that is more varied in terms of word types (Fenson et al., 1994; Goldfield, 1987; Hampson & Nelson, 1993). For example, Tardif, Gelman, and Xu (1999) showed that 20-month-old toddlers of Mandarin-speaking mothers tend to use verbs more often, whereas toddlers of English-speaking mothers tend to use nouns more often. Similarly, English-speaking mothers also tend to use nouns more frequently than verbs when talking to their children, whereas Mandarin-speaking mothers tend to favour verbs over nouns.

Apart from the composition of language input, learning words from different parts of speech often requires different strategies. For example, when learning common nouns, children's learning may be facilitated by the mother's labelling of objects (Hoff, 2006). In contrast, verb learning relies on the repetitive use of verbs in a variety of syntactic environments (Naigles & Hoff-Ginsberg, 1998). Moreover, the context in which parent-child interaction takes place may affect word learning. When the interaction between the caregivers and the child involved a storybook, more nouns than verbs were produced by both English- and Mandarin-speaking children; when a toy was used, more verbs than nouns were produced by both (Tardif et al., 1999). Individual differences in parenting mean variability in language environments, which explains some of the variability observed in children's language development (Hoff, 2006).

Despite variability in the rate of vocabulary acquisition, cross-sectional studies have demonstrated that early vocabulary acquisition is cross-linguistically consistent in terms of overlaps (low variability) in words that children first produce (Frank et al., 2021; Schneider, Yurovsky, & Frank, 2015; Tardif, Fletcher, Liang, et al., 2008). Frank et al. (2021) found that the common first ten words consist of immediate family members (*mommy*, *daddy*, *grandma*), social routines (*hi*, *bye*, *peekaboo*), and sounds (*yum yum*, *vroom*, *woof woof*). It is noteworthy that young children have a high tendency to produce the same first words even though they speak different languages, while their vocabulary comprehension varies considerably across languages and cultures: 12 out of 36 words in production were matched in at least 10 languages, whereas only 4 words in comprehension were matched in at least 10 languages (Frank et al., 2021).

Yet, when the variability in vocabulary is examined with respect to vocabulary scores, the variability in expressive vocabulary tends to increase as vocabulary sizes increase, whereas variability in receptive vocabulary tends to stabilise from 100 words onwards (for the 416-item Oxford CDI), before ceiling effects can have an impact (Mayor & Plunkett, 2014). In other words, children tend to produce more unique words as they grow their vocabulary, whereas, in parallel, they become *general comprehenders* (Mayor & Plunkett, 2014). The finding is in line with Frank et al. (2021), who found that cross-linguistic similarity declines over the course of vocabulary acquisition.

In summary, despite the consistencies in children's first few words, they tend to develop their production towards a more individualistic direction than in comprehension, thus suggesting differences in the mechanisms that drive the development of receptive and expressive vocabulary. Hence, it is important to assess early language skills in both comprehension and production, to fully understand the language skills of each child.

2.2 Early Vocabulary Assessment

2.2.1 MacArthur-Bates Communicative Developmental Inventories

The MacArthur-Bates Communicative Developmental Inventories (CDI; Fenson et al., 2007), a parental-report form, was used as the basis for this thesis. The aim of this thesis is two-fold: 1) to collect the first Malaysian[1] early vocabulary data using an adaptation of the CDI forms and provide insights into Malaysian children's early language development; 2) to introduce CDI-based novel vocabulary assessment tools, with the aim of improving the ad-

[1]The context of Malaysia as a linguistically diverse society is introduced formally in Chapter 3 and some details of the Malaysian government's educational policy are introduced briefly in Section 2.2.3.

ministration of the parent-reported CDI and exploring the viability of a CDI-based toddler-directed word comprehension assessment tool using tablets to supplement the CDIs.

CDIs typically consist of three forms: the CDI-Words and Gestures (CDI-WG) — targeting children between 8 and 18 months of age — assesses both receptive and expressive vocabulary, as well as the production of communicative gestures; the CDI-Words and Sentences (CDI-WS) — targeting children between 16 and 30 months of age — assesses expressive vocabulary, as well as morphosyntactic skills; and the CDI-III — a short form targeting children between 30 and 37 months of age — assesses productive vocabulary, syntactic maturity, and language use (Dale, Reznick, & Thal, 1998; Fenson et al., 2007).

When administering CDIs, parents are asked to report on their child's word knowledge by responding to each word on a list, whether their child understands that word (measured through comprehension) and/or uses that word (measured through production). The judgement of *receptive/comprehensive* word knowledge relies on whether the children react appropriately to a word that their parents have used and the ability of their parents to accurately judge the reaction as appropriate. In addition, the judgement of *productive/expressive* word knowledge relies on parents' ability to accurately recognise the word that their child has produced and that the word was used to refer to the correct item. Although parents are unlikely to recall every instance that involved every word being assessed in the CDI, Frank et al. (2021) argued that assessing children's word knowledge using vocabulary lists is more reliable than asking general questions such as "Does your child know at least 50 words?", which is still commonly used in paediatric assessments.

As a set of parent-report instruments, CDIs are one of the most widely used assessment tools for assessing an extensive list of words a child understands and/or produces (Fenson et al., 2007). Originally developed in American English (Fenson et al., 1993), many adaptations have since been developed in nearly 100 languages[2], including language variations (e.g., British English, Australian English, and Singaporean English), as well as adaptations into sign languages, e.g., American Sign Language and British Sign Language. Among

[2]Including but not limited to, Czech (Markova & Smolík, 2014), Danish (Bleses et al., 2008a), French (Kern, 2003), German (Szagun, Stumper, & Schramm, 2009), Italian (Caselli et al., 1995), Korean (Pae & Kwak, 2011), Latvian (Urek et al., 2019), Mandarin (Beijing, Tardif et al., 2009), Norwegian (Simonsen et al., 2014), Portuguese (European, Cadime, Silva, Ribeiro, & Viana, 2018) and Turkish (Acarlar et al., 2008), see (see CDI Advisory Board, n.d., for a list of available adaptations)

these forms, CDIs normative data for 29 languages, established from a total of over 75,000 children, are available in Wordbank, an open repository for storing and sharing anonymized CDIs data (Frank et al., 2017). Parents are considered reliable sources of information about whether their children understand and/or produce a given word (Ring & Fenson, 2000; Fenson et al., 2007; Styles & Plunkett, 2008), providing extensive information about the word knowledge of their young children. Moreover, the parental-report measure is less time-consuming to administer and has a relatively lower drop-out rate because data collection through CDIs does not experience problems relating to children's unstable temperament and lack of ability to cooperate compared to home-based recordings, or attending an experiment, and thus a questionnaire is an efficient and inexpensive means of assessment.

2.2.2 Validity and Reliability of the CDIs

The CDIs are valid and reliable not only with children who are developing typically (Fenson et al., 2007, 1993; Pan et al., 2004; Rescorla et al., 2005; Law & Roy, 2008), but also with children developing atypically (Galeote et al., 2016; Luyster et al., 2007; Mayne et al., 1998, 1999; D. Thal et al., 2007).

Previous studies have provided support for CDIs' reliability despite being an indirect measure of lexical knowledge (Fenson et al., 1994; Friend & Keplinger, 2003; Marchman & Martínez-Sussmann, 2002). Marchman and Martínez-Sussmann (2002) found a significant correlation between CDI scores and empirical measures of language skills, which included real-object naming tasks and spontaneous language use during free play in children learning both English and Spanish. Besides, Friend and Keplinger (2003) also found a correlation between children's performance in a tablet-based target recognition task and item difficulty that is classified using normative CDI data. Specifically, Friend and Keplinger (2003) found that infants could identify 65% of easy words (defined as comprehended by more than 66% of 16-month-olds in the normative data) in contrast to 42% of moderate words (known by 33%–66% of infants) and 37% of difficult words (known by less than 33% of infants). This demonstrated that normative data collection using CDIs can index children's word knowledge in terms of recognition. CDIs have also been reported to possess moderate to high concurrent validity with typically developing monolingually-exposed children (Fenson et al., 1994; Pérez-Pereira & Resches, 2011; Szagun, Steinbrink, Franik, & Stumper, 2006), with bilingually-exposed children (Marchman & Martínez-Sussmann, 2002), as well as with children diagnosed with Down syndrome (Galeote et al., 2016) and with autism spectrum disorder (Nordahl-Hansen, Kaale, & Ulvund, 2014).

2.2.3 Assumptions and limitations of CDIs

Despite the advantages of using CDIs to collect vocabulary knowledge, they also have their assumptions and limitations. It is assumed that, for typically developing children, parents can manage to track their receptive vocabulary up to about 16 - 18 months of age, after which vocabulary sizes will be too large to be reliably monitored and will typically reach the ceiling of the vocabulary list (Frank et al., 2021). In contrast, productive vocabulary can be monitored up to about three years of age, when it eventually reaches the ceiling of the vocabulary list. It is noteworthy that these assumptions and limitations are commonly applied to monolingual children. Yet, for children exposed to more than one language, CDIs may be administered at older ages, since children exposed to multiple languages typically have lower vocabulary sizes for each language, when compared to their monolingual peers (Bedore, Peña, García, & Cortez, 2005).

Being a parental report measure, the validity of the CDI relies on the parents' willingness to complete it faithfully and on their ability to judge their children's word knowledge accurately. Parental reports are prone to over- and under-estimation, but the accuracy of the measure can be improved under three main conditions: 1) evaluation is limited to current behaviour, 2) assessment is focused on emergent behaviours (i.e., identifying objects given their names or saying object names given the objects), and 3) assessment is conducted based on recognition (in contrast to recalling), thus relieving demands on the respondent's memory (Frank et al., 2021).

In addition, the American English CDI-WG, targeting infants between 8- and 18-month-old, includes 396 words, whereas the American English CDI-Words and Sentences, for toddlers (16- to 30-month-old), includes 680 words. Even in its shorter form, completing a list of 396 words still involves significant time involvement and patience from parents. A shorter version of the form would save time and effort. Yet the shorter form still has to be as representative as the full-CDI for vocabulary assessment. Previous efforts at shortening the CDIs (Fenson et al., 2000; Frota et al., 2016; Jackson-Maldonado, Marchman, & Fernald, 2013; Pérez-Pereira & Resches, 2011; Soli, Zheng, Meng, & Li, 2012) have led to the development of short forms for monolingual use. Moreover, when the assessment needs to be carried out in multiple languages (e.g., in Malaysia), short forms are even more desperately needed, as the number of items is multiplied by the number of languages the child learns. Often, CDIs that were developed on the basis of monolingually-exposed children are used in studies involving bilingually-exposed children (Core, Hoff, Rumiche, & Señor, 2013; De Houwer, Bornstein, & Putnick, 2014; Marchman & Martínez-Sussmann, 2002), which can be problematic when they are used in other countries with a different culture (e.g., Malaysia). For example, vocabulary items in the American English CDI such as *gas station*, *bug*, *mad* and *candy* are uncommon words in the Malaysian context (e.g. commonly: petrol station, fly/mosquito, angry and sweet), signalling the need for adaptation in the Malaysian context.

2.3 Early Vocabulary Development

2.3.1 Age

Age is an important factor when assessing a child's language skills. Typically developing children are expected to acquire larger vocabulary sizes as they grow. However, the rate of acquisition varies across age – it is typically slow at age between 6 and 18-month-old, in which a rapid spurt occurs around 18 and 24-month-old. For children aged 8 - 30 months, Fenson et al. (1994) reported a significant positive correlation between age and vocabulary (in production and in comprehension). More recently, Feldman et al. (2000) compared the production and comprehension vocabulary scores among children between 10- and 13-month-old, as well as the production scores among children between 22- and 25-month-old. On a monthly basis, all comparisons showed progressive growth in vocabulary sizes.

Among older children, too, Garcia et al. (2014) reported a significant development trend with an increase in the production vocabulary in Basque from 30 months to 42 months of age, which stabilised after 42 months of age. The stabilisation of vocabulary sizes is not due to children stopping to learn new words, but rather a constraint limited by the number of items tested in the CDI. Despite that, a clear age-vocabulary relationship was found using CDIs in many countries, across a range of languages (see CDI Advisory Board, n.d., for a list of available adaptations). However, age is not the sole predictor of the development of children's vocabulary knowledge. Fenson et al. (1994) found that age only explained 36% of the variance for children between 8 and 16 months and 22% for children between 16 and 30 months of age. According to Fenson et al. (1994), the remaining variance is driven by, but not limited to, individual differences and respondent inconsistencies – parents who had pride in their child might lead to overestimates or parents who are frustrated with slower learners might lead to underestimates (Frank et al., 2021).

In contrast, the use of CDIs revealed consistencies in early vocabulary acquisition. Studies using CDIs have found cross-linguistic similarities in early vocabulary trajectories among children speaking different languages (Bleses et al., 2008c; Braginsky, Yurovsky, Marchman, & Frank, 2019; Frank et al., 2021). The evidence suggests that most children produce their first words between 12 and 20 months of age (Bleses et al., 2008c; Devescovi et al., 2005; Fernald et al., 2001) and that their vocabulary acquisition rate increases rapidly after 18 months of age (E. Bates & Goodman, 2001; Fernald et al., 2001; Fernald, Perfors, & Marchman, 2006). In addition, a strong relationship between lexical and grammatical development has been reported in CDI-based studies, signalling the role of lexical knowledge in child language development (e.g. Caselli, Casadio, & Bates, 1999; Stolt, Haataja, Lapinleimu, & Lehtonen, 2009; Conboy & Thal, 2006; E. Bates & Goodman, 1997; Devescovi et al., 2005; Marjanovič-Umek, Fekonja-Peklaj, & Podlesek, 2013). Furthermore, CDIs have been used as an additional criteria for identifying Late Language Emergence. For example, starting from 24 months of age, a child is typically considered to be a late talker or a late language learner if he/she has an expressive vocabulary at or below the 10th percentile on the CDI (Weismer, 2017; Dale, Price, Bishop, & Plomin, 2003; Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008; Rescorla & Dale, 2013). The benefit of an early diagnosis cannot be overstated – evidence suggests that some 24-month-old toddlers with language delay remain delayed 2 to 3 years later when left untreated (Dale, 1991; Fischel, Whitehurst, Caulfield, & DeBaryshe, 1989; Rescorla, 1989; Thal & Bates, 1988).

2.3.2 Gender Differences

The rate of early vocabulary acquisition was found to differ between genders – Fenson et al. (1994) discovered that girls scored higher than boys, yet the difference only represented 1% to 2% of the variance. This finding is consistent with a meta-analysis study conducted by Hyde (1981), which involved 27 verbal skill studies in child- and adulthood. It was found that gender represented approximately 1% of the variance in verbal skills. Moreover, a study conducted by Eriksson (2017) with a CDI-III designed for Swedish children aged between 2 years 6 months and 4 years reported that 2.6% variance was accounted for gender, in which girls (M = 67.89) had an average of 9.45% more words in production than boys (M = 62.03). In a Danish CDI study conducted with children aged between 8 and 36 months, Bleses et al. (2008a) also found a small yet significant gender difference in both word comprehension and production scores, where girls scored higher than boys. Conversely, in an American English CDI study, Feldman et al. (2000) did not detect a girl advantage in comprehension, but it was present in vocabulary production, and in terms of phrases understood and gesture score. In all, a small yet significant girl advantage in vocabulary development was consistently observed across languages (see Bleses et al., 2008a; Bornstein et al., 2004; Eriksson, 2017; Fenson et al., 1994; Frank et al., 2017; Galsworthy et al., 2000; Schults & Tulviste, 2016; Simonsen et al., 2014; Stolt et al., 2008; Tardif et al., 2009).

The gender difference is more pronounced when comparing the use of early grammar such as genitives, plural nouns, prepositions, or past tenses. In a Danish study, Bleses et al. (2008a) reported that girls were considerably better at applying early grammar into their expression than boys – while the magnitude of difference in mean production score was by a factor of approximately 1.3, the mean number of complex forms was higher by a factor of two at 26-month-old. One potential explanation is that girls are more socially interactive and, thus, have had more practice in language use. Girls tend to engage more in verbal interactions during play activities and are more talkative than boys (Leaper & Smith, 2004). This parallels the difference in the development pace whereby girls generally develop socio-cognitive skills, e.g., theory-of-mind, earlier than boys (Barbu, Cabanes, & Le Maner-Idrissi, 2011). Barbu et al. (2011) argued that these gender differences in the development of socio-cognitive skills facilitated gender differences during play activities and influenced the subsequent girl advantage in grammar skills. However, early language learning, like any other type of learning, often relies on external factors such as availability of input, especially in a linguistically rich society such as Malaysia.

Despite the use of CDI to assess the language skills of both monolingually-exposed and bilingually-exposed children, these groups of children need to be treated differently due to the distinction in the pace and process of language acquisition. One source of variation is the distribution of attention across languages in bilingual acquisition, as opposed to monolingual learners who focus solely on one language. Besides, larger relative exposure in one language affects real-time processing efficiency of that language, which in turn facilitates vocabulary acquisition in that language, as shown in a longitudinal study by Hurtado, Grüter, Marchman, and Fernald (2014). As a result, the language that received less attention invariably leads to slower acquisition.

2.3.3 Language Exposure and Dual Language Learning

Under the age of three years, the *amount* of exposure to a language is found to be positively correlated to the vocabulary size in that language, in dual language children learners (Côté, Gonzalez, & Byers-Heinlein, 2020; David & Wei, 2008; Hoff et al., 2012; Parra, Hoff, & Core, 2011; Pearson, Fernandez, Lewedeg, & Oller, 1997; Thordardottir, 2011). For instance, Pearson et al. (1997) reported that the amount of time spent with speakers of a language is directly proportional to the number of words that the Spanish-English dual language children learners know in that language. In addition, the language that had a dominant exposure in the environment naturally led to a reduction in exposure to the other language in bilingually-exposed children. As a result, the word learning of the other language(s) can be negatively affected. Indeed, Parra et al. (2011) studied American Spanish families with language exposure between the range of 10% English and 90% Spanish to 90% English and 10% Spanish for their 2-year-old children and found that the influence of language exposure on language development is language specific. In other words, home language exposure to the English language positively predicts English vocabulary but negatively predicts Spanish vocabulary.

In addition, simply *being exposed* to two languages appears to have negative consequences on dual-language children learners' overall vocabulary size. Oller and Eilers (2002) reported that bilingually-exposed children acquired a lower number of vocabulary items in each language when compared with monolingually-exposed children learning just one of those languages. This vocabulary gap between monolingually-exposed and bilingually-exposed children learners can be reduced when an optimal amount of exposure to a language is achieved. Cattani et al. (2014) found that a 60% English exposure is sufficient for bilingually-exposed children learners to perform at a level on par with English monolinguals. In contrast, Thordardottir (2011) found no vocabulary gap between bilingually-exposed children learners of French and English and monolingually-exposed children in receptive vocabulary, despite spreading their attention across two languages. Thus, the findings seem to suggest a facilitative effect on early vocabulary acquisition when learning linguistically similar languages. Indeed, other studies have shown the effect of form similarities between words on vocabulary acquisition across both languages, such as between Spanish and Catalan (Bosch & Ramon-Casas, 2014), English and Japanese *katakana* words (Kutsuki, 2021), Italian and German, and French and German (Persici, Vihman, Burro, & Majorano, 2019).

2.3.4 Socio-Economic Status

Early vocabulary is modulated by socio-economic status (SES), among other factors such as age, gender, and individual differences. Studies have shown that SES modulates the linguistic skills of the child, especially lexical development and language processing (see Arriaga, Fenson, Cronan, & Pethick, 1998; Bornstein, Haynes, & Painter, 1998; Fenson et al., 1994; Fernald, Marchman, & Weisleder, 2013; Hart & Risley, 1995; Hoff-Ginsberg, 1991, 1998). Strikingly, Fernald et al. (2013) found a 6-month gap in vocabulary scores and real-time language processing abilities between lower-SES and higher-SES families – 24-month-old toddlers from lower-SES families had similar vocabulary scores and language processing efficiency as 18-month-old children from higher-SES families. As early life is a critical time to establish the building blocks of language, a six-month-gap is huge – as if a quarter of the lower-SES children's lives were lost in terms of language development. Some studies measured an even wider lag, in which a 2-year-gap among 5-year-old lower-SES children was found compared to higher-SES children when they entered school (Greenwood et al., 2017). It is noteworthy that SES is a composite index and the influence of SES over child language development is multifaceted. Family characteristics such as income, occupation, household possessions, family structure, and parental educational attainment are found as reliable indicators of SES (Brese & Mirazchiyski, 2013). The current project focuses on household income, in which higher income indicates better access to learning resources, and parental education attainment, which is related to children's academic achievement (Davis-Kean, 2005; Rowe, 2008).

2.3.4.1 Household Income

In parallel, differences in household income have been associated with a vocabulary difference – children from higher income families had a larger vocabulary size compared to children from lower income families, with evidence from a range of countries: US - Arriaga et al. (1998) and Layzer and Price (2008); UK - Blanden and Machin (2010); Australia - Taylor, Christensen, Lawrence, Mitrou, and Zubrick (2013). Using the Peabody Picture Vocabulary Test for vocabulary comprehension, Layzer and Price (2008) found a difference of nine months for 36-month-old American children whose family income was below the poverty line (as defined by their government). In a CDI-based study, Arriaga et al. (1998) compared the vocabulary in production of American children (*M* = 23.49 months, *SD* = 4.23) in low-income families with the normative sample from Fenson et al. (1993)'s study and found that the children in low-income families had 30% less vocabulary in production than their middle-income peers. For comparisons between children from low- and high-income families, Hart and Risley (1995) found that, at the age of 3 years, children from low-income households produced twice as few words as their peers from high-income households.

The vocabulary differences persist at a later age – Taylor et al. (2013)'s study on Australian children observed an 8.5 month vocabulary size difference in comprehension at 4 years old between children in low- ($600 per week) and high-income families ($2000 per week). The gap remained when the children's vocabulary knowledge was reassessed at 8 years old. Worse, other studies have shown a widened gap at a later age – in a UK study conducted by Blanden and Machin (2010), a 10-month vocabulary size difference in production was found between children in low- and high-income families at age 3 and the gap widened to 15 months at age 5. Household income is important to children's vocabulary acquisition as it reflects the ability of the parents to provide language learning materials that contribute to building a language-rich environment for their children, alongside fulfilling the basic needs of the family.

2.3.4.2 Household Income and Dual Language Learning

In parallel, in the context of dual language learning, evidence indicating a relationship between income and vocabulary is limited and partial (Castro, Páez, Dickinson, & Frede, 2011; Oller & Eilers, 2002). While Castro et al. (2011)'s found a negative effect of low income, the majority of bilingual households in their sample belong to the low-income group, indicating that their findings reflect both economic level and bilingualism. The study compared bilingually-exposed children learners' English vocabulary skills to monolingually-exposed peers from a high-income group, with a lack of focus on their heritage language, which is not the dominant language in the country they currently live in. As a result, it is unclear whether income level had the same effect on their heritage language. In a study conducted in Hong Kong, Cheung and Wong (2020) found a negative effect of low income on children's English language skills, a foreign language (which is associated with a better social status in Hong Kong), but their Mandarin language skill and Cantonese oral language skills were unknown (both are the national language and regional dialect of Hong Kong, respectively). Nonetheless, in a bilingual context, there is some evidence that suggests a positive relationship between parental income and a child's word knowledge. While children from higher income families had more access to educational resources (Davis-Kean, 2005; Rowe, 2008), in the context of bilingualism, it is unclear whether this advantage translates into a larger vocabulary size in both languages. In addition, it is unclear whether the social status of languages modulates this income effect. These are important questions that we will attempt to address in Study 2.

2.3.4.3 Parental Education

Parents' educational attainment has been shown to contribute to the disparities in word learning of children across culture (Hart & Risley, 1995; Hoff, Burridge, Ribot, & Giguere, 2018; Rowe, 2008; Wanless, McClelland, Acock, Chen, & Chen, 2011). For instance, Rowe (2008) measured the vocabulary comprehension of 30-month-old children (mostly European Americans) and found a positive correlation between English-speaking parents' educational level and the children's comprehension scores, which was mediated by the amount of child-directed speech. Similar findings were found in the Asian context – Wanless et al. (2011) examined Taiwanese children aged between three and half- to four and a half-year-old and found that a mother's education was positively related to her child's Mandarin vocabulary comprehension. Hoff et al. (2018)'s study suggested language specificity in the influence of maternal education on children's word learning process – they found maternal educational level in Spanish was positively related to their children's productive language in Spanish but not to English and that maternal educational level in English was positively related to their children's English vocabulary but not to Spanish. Parents' education is crucial to their children's learning as it shapes parents' knowledge of child development and thus encourages them to provide a rich language environment that facilitates learning (Hoff et al., 2018).

2.3.4.4 Parental Education and Dual Language Learning

The relationship between SES and children's word learning is more complicated in bilingual families due to the distributed attention to each language. In a study involving dual language children learners aged between 4 and 6 years old in Singapore, Sun, Ng, O'Brien, and Fritzsche (2020) did not find any significant relationships between mothers' educational level and children's vocabulary comprehension in their heritage language. This is likely due to a huge shift in the use of English, as a result of Singaporean government policies. (Cavallaro & Ng, 2014; Gopinathan, Ho, & Saravanan, 2004; Tupas, 2011). Ever since the 1960s, English has enjoyed widespread use from government to business to the media and is the medium of instruction (MOI) for schools in Singapore (Sun et al., 2020). This affects Singaporean parents' ability to provide a fluent and facilitating environment for their children to acquire their ethnic language (EL) [3]. In addition, Saravanan (2001) found that Singaporean parents from higher SES families tended to prefer the use of the English language at home, which resulted in poorer proficiency of their children in EL. Yet, the sociopolitical climate in Malaysia, a historically-intertwined neighbour, is a stark contrast to Singapore's embrace of the English language.

2.3.5 The Malaysian Context

Malaysia – the country where we collected early vocabulary data reported in Chapter 3 – is a multi-ethnic country that consisted 69.6% of Malay-ethnic people, 22.6% of Chinese-ethnic people, and 6.8% of Indian-ethnic people, with 1.0% categorised as "Others" (Department of Statistics Malaysia, 2020). Twenty-three percent of the population are children aged between 0 and 14 years of age, with a total fertility rate of 1.8 babies born to every woman in the reproductive period (15-49 years). People from all three ethnicities have typically retained their ethnic languages, with Malay-ethnic people who speak Malay, Chinese-ethnic people who speak Mandarin and Indian-ethnic people who speak languages of Indian origin (e.g., Malayalam, Hindi, and Punjabi) (Albury, 2017). Communication between ethnic groups relied on the use of English or Malay – a legacy of British colonial rule (Jenkins, Cogo, & Dewey, 2011).

In Malaysia, EL is still being used as the MOI in public primary and secondary schools. The EL discussed in this context refers to standard languages such as Malay and Mandarin, which are widely used in formal education.

For primary education, Malay is used in the National Schools, whereas Mandarin and

[3] The term 'ethnic language' (EL) refers specifically to the language belongs to parents' ethnicity.

Tamil are used in the national-type schools[4]. The Malay and English languages are compulsory subjects in all public primary schools (Hall, 2015; Hashim, 2009).

For secondary education, Malay is the MOI and English is a compulsory subject in the National Secondary School, whereas parallel systems such as the Chinese independent secondary schools[5], and the English-medium international secondary schools (Hall, 2015) use Mandarin and English respectively.

One of the caveats is that, for public primary and secondary schools, a separate language policy was implemented for the MOI of Mathematics and Sciences (Math-and-Science) and it has gone through two major changes – a change from EL to English in 2005 by the implementation of the *Teaching and Learning of Science and Mathematics in English* (PPSMI) then replaced by the *Dual Language program* (DLP)[6] in 2012 to allow schools to offer an option between the Malay and the English language.

In terms of tertiary education, public universities have used English as the MOI for only science and technology courses since 1993, and private universities use English as the MOI for all courses (Gill, 2004).

This complex historical backdrop thus motivates an in-depth investigation into the influence of parents' education level on their children's vocabulary acquisition when learning more than one language, in Study 2. In summary, for parents who had a university education (taught in English), they would have more experience in the use of the English language, as opposed to parents who did not continue their education beyond secondary/primary school (primarily taught in EL).

2.3.6 Early Interventions in Malaysia

As discussed, SES factors contribute to disparities in children's educational outcomes. Yet, children from low-SES families can benefit from early intervention. For example, Head Start, an Early Childhood Care and Education (ECCE) program, was able to improve the English-based educational outcomes of children from low-income households in the United

[4]Non-Malay primary schools that follow the national curriculum.
[5]Private secondary schools without government funding.
[6]This second change would not affect the parents reported in this thesis, due to their age.

States of America (USA; Bauer & Schanzenbach, 2016; Deming, 2009). The Head Start program aims at improving children's readiness for the transition from preschool to elementary school and to help them meet the expectations of the school, by providing early childhood education and health supports (McWayne, Cheung, Wright, & Hahs-Vaughn, 2012). Such a program is beneficial to poor families (Park, Gurel, Oh, Bettini, & Leite, 2015), as well as fosters the educational outcomes of their children (W. Lee & Pring, 2016), and may be used as a basis to inform policy making in Malaysia. However, it is noteworthy that, while the Head Start program emphasizes English-based education due to the dominant use of the English language in the USA, the socio-cultural background in Malaysia is richer and ethnic languages still share a significant position in society alongside the English language.

Efforts were made by the Malaysian government, launching a nationwide screening program for language, numeracy, and literacy difficulties among primary school students under the LINUS initiative in 2010 (Ministry of Education Malaysia, 2013). However, the LINUS program was intended for primary school students, not for *pre-school* children, which is often considered relatively late for intervention since children had already entered school. As documented in the Malaysia Education Blueprint 2013-2025, the Ministry of Education Malaysia aims to improve the Malay and English language proficiency of children nationwide (Ministry of Education Malaysia, 2013). It is crucial that effective assessment tools are available to attain the goals paved out in the blueprint, by detecting and then providing early interventions to children with language delay. Early intervention typically targets children from birth to 3-year-old but may continue older than 3 (ASHA, 2020). Early intervention involves both families and professionals (e.g., speech-language therapists, audiologists) and helps children develop skills in the cognitive, communication, motor-sensory, social, and self-help domains, depending on the child's needs and the family's priorities (Ramey & Ramey, 1998; ASHA, 2020).

Early detection and intervention of language delays and impairments are also essential to a child's future achievement (Dale, 1991; Fischel et al., 1989; Preston et al., 2010; Rescorla, 1989; Thal & Bates, 1988) and well-being (Johnson, Beitchman, & Brownlie, 2010). Many children with language delays still suffer from learning difficulties in schools due to lack of intervention during the early stages of life (Preston et al., 2010). Evidence showed that toddlers with language delay remained delayed 2 to 3 years later when left untreated (Dale, 1991; Fischel et al., 1989; Rescorla, 1989; Thal & Bates, 1988). Moreover, children who have language delays or are considered at-risk (due to low language proficiency of the family or, etc.) often struggle academically and may have problem socialising (Johnson et al., 2010). Early intervention is especially important for children who need special support, such as children with learning disabilities, dyslexia, hearing impairment, autism spectrum disorder, and attention deficit hyperactive disorder (Bowyer-Crane et al., 2008; Dawson & Bernier, 2013). Therefore, early diagnosis and intervention are crucial. Previous research indicated an improvement in language learning later in life when an intervention was provided to children who were younger than 3 to 5-year-old (Fuchs & Fuchs, 2009; Mack, Smith, & Straight, 2010).

Although the cost of nationwide-scale language screening is considerable, the benefits cannot be overstated. In the USA, for example, economists calculated that every dollar invested in early diagnosis and intervention of developmental disabilities saves up to 7 dollars in the long run (Keating & Hertzman, 2000). The identification of language delays should take place as early as possible, preferably during the preschool years, to ensure that potential interventions can achieve maximum efficacy and be sure that all children enter primary school with similar starting points. By using parental reports such as CDIs, children at risk can be identified at a younger age, during infancy and toddler-hood.

Moving on, while external factors such as SES influence children's early vocabulary acquisition in terms of vocabulary *size*, other external factors, such as the parents' attentional pattern and the linguistic properties of the language also play a role in shaping the vocabulary *composition* of the children, the focus of the next chapter.

2.4 Noun-Verb-Bias and Vocabulary Compositions

Over the past decades, researchers have documented that young children tend to acquire nouns before verbs across a range of languages, such as English (Gentner, 1982), German, Italian (Caselli & Casadio, 1995) or Spanish (Jackson-Maldonado, Thal, Marchman, Bates,

& Gutierrez-Clellen, 1993) and that they are able to learn novel nouns at a quicker pace than novel verbs, in English (Childers & Tomasello, 2002). Gentner (1982) offered two hypotheses to explain the presence of a noun bias during early language learning: the natural partitions hypothesis and relational relativity hypothesis. Although the rationales behind these hypotheses are somewhat different, both lead to the same expected outcome: the superiority of noun learning over verbs during infancy. The natural partitions hypothesis suggests that it is easier for infants to label concrete objects than to learn verbs, which describe relations between objects. In order to acquire the meaning of a verb, the learner has to first understand what/who the actor is and, often, the receiver of the action as well. Hence, the acquisition of nouns is a precursor to verb learning.

Yet, numerous studies have shown that noun learning does not always outperform verb learning. While multiple languages are considered "noun-friendly" (e.g., English, Spanish) – children's early vocabularies contain more nouns than verbs – studies on the acquisition of Korean, Cantonese and Mandarin as first languages have suggested these languages to be "verb-friendly" (Arunachalam, Leddon, Song, Lee, & Waxman, 2013; Chen et al., 2015; Choi & Gopnik, 1995; Tardif et al., 1999; Tardif & Fletcher, 2008) – with nouns and verbs equally prevalent in early vocabularies (Choi & Gopnik, 1995; Tardif, 1996; Chen et al., 2015; Tardif et al., 1999; Tardif, Shatz, & Naigles, 1997) – and with more verbs than among noun-friendly languages (Choi & Gopnik, 1995; Tardif, Fletcher, Liang, et al., 2008; Tardif et al., 1997). These findings suggest that the strength of the noun bias varies across languages and cultures, rather than being a language universal phenomenon (Lavin, Hall, & Waxman, 2006).

2.4.1 Extra-Linguistic Factors

Researchers have suggested that attentional patterns differ across cultures, hence impacting the composition of early vocabulary: Westerners tend to focus their attention *analytically* on objects–typically referred to as nouns–whereas Asians tend to focus *holistically* on the relationship between objects–often described using verbs (Nisbett & Miyamoto, 2005). Additional studies provided further demonstration of culture-dependent attentional structures (see Nisbett, 2004, for a collection of studies). For instance, Ji, Peng, and Nisbett (2000) found that Americans scored higher than East Asians on a Rod-and-Frame test, a test used to measure visual field independence. A higher score reflects a greater ability to differentiate an object from the field. In contrast to Asians, Ji et al. (2000)'s findings suggest Americans were more accustomed to analysing and directing themselves with respect to a focal object than to their surroundings. Cultural differences in attentional patterns seem to affect parent-child interactions. Tardif et al. (1999) showed that 20-month-old toddlers of Mandarin-speaking mothers tend to use verbs more often, whereas toddlers of English-speaking mothers tend to use nouns more often. Correspondingly, English-speaking mothers also tend to use nouns more frequently than verbs when talking to their children, whereas Mandarin-speaking mothers tend to favour verbs over nouns.

Children acquired the attentional pattern specific to their culture too – Waxman et al. (2016) discovered that when a video was being shown, 24-month-old Chinese infants attended to more action-related elements whereas American infants focused more on the objects involved. American infants displayed more interest when a new object was featured in a familiar scene depicting the same actor (e.g., a girl) performing the same action (e.g., watering) on a new object (e.g., a plant), whereas Chinese infants displayed more interest when a new action was featured (Waxman et al., 2016). This indicates that Chinese and American infants deploy their attention differently. Such differences in the attentional pattern can also modulate children's abilities to learn words. In a habituation-switch experiment conducted by Chan et al. (2011), they showed word-to-scene pairings to Mandarin-learning and English-learning infants during the habituation phase and tested whether the infants had formed the correct pairings during the test trial by switching one of the labels (either the noun or the verb). It is expected that infants should elicit longer looking times if they notice a novel label-to-scene pairing during a test, thus suggesting the formation of correct word-to-scene pairings during habituation. Mandarin-learning infants were better at associating novel words with actions, whereas English-learning infants did better at mapping novel words to objects (Chan et al., 2011). Based on these studies, we suggest that differing attentional patterns across cultures can manifest as a learner's bias that may be modulating infant lexical development. Such differences are external to the linguistic properties of a language – we will refer to them as *extra-linguistic* factors.

Other extra-linguistic factors may also be driving differences in early lexical composition across languages and cultures. Using the "Human Simulation Paradigm" (Gillette, Gleitman, Gleitman, & Lederer, 1999), Snedeker, Li, and Yuan (2003) investigated the ability of adults to correctly guess the target words (equal proportion of nouns and verbs) from silent videos depicting a play session between a mother and an infant. The use of silent videos removed language-related verbal cues and participants have to rely on extra-linguistic cues to guess the target words. They showed that both English- and Mandarin-speaking adults were better at identifying nouns than verbs from silent videos of English infant-directed speech. Yet, when exposed to silent videos of Mandarin infant-directed speech, both Mandarin and English speakers had similar performances when identifying nouns and verbs. Snedeker et al. (2003) results suggest that extra-linguistic information may account for the presence of a noun-bias in the vocabulary of English-learning children, while the early vocabulary of Mandarin-learning children would be more balanced. While this correspondence suggests infants exposed to a noun-friendly language acquire an early lexicon rich in nouns, it does not explain why some languages are noun-friendlier than others.

2.4.2 Language-Intrinsic Factors

In parallel, other researchers have argued that factors *intrinsic* to a language can also impact the early word acquisition process (e.g., morphological transparency – the extent to which words in a language change their morphological form through inflections; pronoun-dropping parameter – the tendency to drop pronouns, thus reducing the proportion of nouns; and word order; Tardif et al., 1997, 1999; Tardif, 1996). For example, Mandarin (but not English) is a pronoun-dropping language that allows omission of pronouns, making verbs more likely to appear at the salient front or end of a sentence depending on whether the subject or object was omitted (Tardif et al., 1997), in turn putting the emphasis on the verb.

Interestingly, a language could have its pronouns dropped when used but still be a noun-friendly language at the same time. An example of that would be the Italian language. In terms of morphological transparency, its rich verb inflections (frequently marked for person, number, tense and gender) relative to nouns favour the learning of nouns over verbs (see, Tardif et al., 1997, for linguistic example). English, with limited noun inflections and richer verb inflections (see, Gentner, 1982, for linguistic example) can also be seen as favouring the acquisition of nouns, making English a noun-friendly language. In contrast, Mandarin hardly has any verb or noun inflections, thus reducing the asymmetry between noun and verb inflection complexity (see, Tardif et al., 1997, for linguistic example). Additionally, verbs in Mandarin are often enhanced by participles, further enhancing the notion of verb in the sentence, in turn making Mandarin a verb-friendly language (see, Tardif et al., 1997, for linguistic example).

Our brief review of the literature suggests that both language-intrinsic features and socio-cultural influences may impact children's acquisition patterns of nouns and verbs. Crucially, these language differences appear to correlate with the attentional structure displayed across cultures. Hence, a direct comparison of participant's vocabulary across languages usually does not allow for an assessment of the relative contribution of language-intrinsic factors (syntax, morphology) and language-extrinsic factors (attentional patterns, extra-linguistic context) to the noun- (or verb-) bias.

2.4.3 Noun-Verb-Bias and Dual Language Learning

One promising avenue is to assess bilingually-exposed children, as this allows researchers to evaluate potential differences in the noun-bias of two languages within a single learning environment. Xuan and Dollaghan (2013) collected parental reports on Mandarin-English bilingually-exposed children raised in the USA. They found that the expressive lexicon in Mandarin contained more verbs than in English, while an analysis of the 50 most frequent words in English contained more nouns than in Mandarin. They concluded that the noun-bias is language-specific, as other potentially confounding factors such as socio-economic status did not vary within the sample. Yet, as bilingually-exposed children in the study may have learned their languages from two different speakers (or in two different learning contexts, e.g., Mandarin at home, and English in a day-care), they may have been provided with distinct extra-linguistic cues when learning both languages. Both language and differences in parenting context may play a role in creating differences in the lexicons of children across language groups.

To dissociate language specificity from extra-linguistic cues, bilingually-exposed children should ideally be learning both languages from the same caregiver. Chan and Nicoladis (2010) addressed this issue, as they followed longitudinally two Mandarin-English bilingually-exposed children who were learning both languages from the same person, thereby providing enhanced control over extra-linguistic cues when assessing language-specific factors in the noun bias. They found both children to use more nouns than expected from the analysis of the salient positions in their utterances, for both of their languages. Chan and Nicoladis (2010) suggested that the parents, immigrants to Canada, were acculturating to a western style of communication, hence potentially reducing the highlighting of verbs in their non-verbal behaviour that would otherwise be observed in monolingual Mandarin-speaking parents (Snedeker et al., 2003). Other than a correspondence between the input children were exposed to and their developing lexicon, Chan and Nicoladis (2010)'s results also suggested that the social-cultural contexts in which children are being raised (in that case, Chinese immigrants acculturating to a western culture) modulate the noun-bias in languages they are learning.

While previous studies with bilingual children have taken the approach of assessing differences across their languages learnt within a *common learning environment* (e.g., Chan & Nicoladis, 2010; Xuan & Dollaghan, 2013), we suggest a complementary approach of examining the differences in a *common language* learnt from two distinct socio-cultural environments (Study 3). Specifically, we focus on the English learnt by two groups of bilingual children from two distinct sociocultural environments – raised in Chinese-ethnic families and in Malay-ethnic families in Malaysia. To measure the vocabulary acquisition of the children, we relied on the MacArthur-Bates Communicative Developmental Inventories - Malaysian version (MCDI-M), an adaptation of the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007) developed by Low (2010). As a trilingual CDI, the Malaysian version comprises vocabulary in Malay, Mandarin and English language, thus covering some of the most prevalent standard languages in Malaysia

(H. W. Lim, Wells, & Howard, 2015), excluding Tamil (a candidate for future version of the form) and excluding regional dialects. Every conceptual item in the Malaysian CDI is presented with translation equivalents in all three languages. The trilingual CDI consists of $1,800$ vocabulary items – 600 conceptual items presented in three languages – in contrast to 680 words in the original American English CDI-WS.

2.5 Efficiency of CDI Assessments

The trilingual CDI is an exhausting list – it contains $1,800$ vocabulary items. Hence, the applicability of CDIs becomes limited when it is crucial for a rapid assessment, i.e., when multiple assessments are being conducted in a lab session, or when assessing multiple languages of non-monolingually exposed children. These are limitations for monolingual CDI too, which often consisted for 600 items and above. There were some efforts to develop short-form versions of monolingual CDI forms to address these limitations (Fenson et al., 2000; Frank et al., 2017). These short forms have demonstrated high validity and reliability, as evidenced by high correlation with the full forms, thus making them a useful alternative when time is limited (Fenson et al., 2000). However, due to CDI-SF's reduced length, these forms reach the ceiling earlier than the full CDIs – the short-form CDI-WS suffers from a ceiling effect after 27 months, in contrast to 36 months in the full CDIs. Furthermore, the development of such forms is language-specific, time-consuming and labour-intensive, that is, to reduce the length of the CDIs while maintaining a balance of items from different semantic categories and with different difficulty levels.

2.5.1 Parental Reports

2.5.1.1 Item Response Theory-based Computerised Adaptive Test

Recently, in an effort to develop shorter versions of CDI assessments that are tailored to the ability of each child, Makransky et al. (2016) applied Computerised Adaptive Testing (CAT; van der Linden & Glas, 2010), whose principles are based on Item Response Theory (IRT). In their approach (hereafter referred to as the Computerised Adaptive Testing version of CDI (CDI-CAT)), items in the American English CDI-WS normative sample[7] are fitted to an IRT model. In IRT, items may differ in discrimination value, which determines how well each item discriminates the level of knowledge across individuals (Fraley, Waller, & Brennan, 2000). During CAT, 10 items with maximal item information[8] are initially sampled at random from the full CDI, to initialise a stable ability parameter. The algorithm then selects subsequent items that reflect the ability parameter of the child that is estimated at each point (i.e., item) in the test. Based on the results obtained from CDI-CAT simulations with 5, 10, 25, 50, 100, 200, 400, and 680 (the full form) items, it was found that at 50 items and above, CDI-CAT performed well, with correlations above .95 with the full CDI, average *SE*s below .20, and reliability coefficients above .96 –Makransky et al. (2016)'s threshold for test acceptability. While this may be a viable solution to reducing the length of the full CDI, the performance of CDI-CAT with novel empirical data (in terms of correlation with the full CDI), as pointed out by Makransky et al. (2016), may be lower due to systematic or random errors, when compared to their simulation study. It is also possible that the respondents would respond differently as items in CDI-CAT are presented in a semantically unstructured order as opposed to the semantic grouping adopted in CDIs.

2.5.1.2 Bayesian-Inspired Early Vocabulary Model

With the objective of developing a short-form version of CDI-WS that parallels the accuracy and precision of the full form, Mayor and Mani (2018) presented a language-general framework that takes advantage of the richness of data from the Stanford Wordbank (Frank et al., 2017). Their approach is able to estimate full-CDI scores with only a subset of items drawn randomly from the full forms, by utilising normative CDI data sampled from language-, gender-, and age-matching children on Wordbank. In order to examine the effectiveness of the framework, Mayor and Mani (2018) conducted a series of real-data simulations using CDI-WS normative data of American English (Fenson et al., 2007), German (Szagun, Stumper, & Schramm, 2014), and Norwegian (Simonsen et al., 2014)

[7] The American English CDI-WS normative sample consisted of 1,461 children between 16 and 30 months of age.

[8] As measured by items' capacity to discriminate between children based on composites of item discrimination and difficulty level estimated using the IRT model.

retrieved from Wordbank. The results revealed that at 50 items, correlations reached .97, with average *SE*s of .05, and reliability coefficients of .99, suggesting that their approach, which takes into account children's age and gender, outperforms the CDI-CAT reported in (Makransky et al., 2016). In addition, empirical validation with 25- and 50-item word lists administered to the parents of German-speaking children showed similar performance to the real-data simulations, with correlations of .96, average *SE*s of .14, and a reliability of .98, even when parents showed inconsistencies (about 10-15% of responses) between responding in the full and short forms. However, due to the large variation in vocabulary acquisition (i.e., within- and between-age variations, and gender differences; Fenson et al., 2007), Mayor and Mani (2018)'s approach requires that the sample size on Wordbank be sufficiently large to be able to capture the norm — the German CDI-WS dataset, being the smallest dataset of the three CDIs they selected, had over 70 children at each month-age group. Hence, it is unclear whether their approach would perform well with a smaller sample, e.g., in languages having fewer normative data available on Wordbank and in the current Malaysian sample.

Another obvious limitation to this approach is that items are *randomly* selected during the test. Consequently, the sampled items may be minimally informative, i.e., items that are either too easy (e.g., *cat* is produced by over 95% of 30-month-old in American English) or too difficult (e.g., *snowman* is produced by just 1% of 16-month-olds), and thus may inform less about a child's language ability. This limitation can be remedied by the application of the IRT introduced by Embretson and Reise (2000); Makransky et al. (2016). For instance, weaker or younger children may be randomly assigned items with a high difficulty level. While a difficult item may be able to differentiate between two children having a high degree of knowledge, this is unlikely for weaker or younger children. To address these issues, our approach aims to combine Mayor and Mani (2018) and (Makransky et al., 2016)'s approaches (we termed MM-IRT) to produce language-general, short-form versions of CDIs in which items are selected to be maximally informative, with their performance on the normative sample assessed in Section 4.2: Study 4 and their application to the current Malaysian sample reported in Section 4.3: Study 5.

Yet, the current application of the MM-IRT is restricted to parent reports of comprehension and production using the CDIs. There have been concerns about relying solely on parental reports, especially when it comes to assessing comprehension, because parents can only infer comprehension based on infants' and toddlers' non-verbal responses to language (Feldman et al., 2000; Houston-Price, Mather, & Sakkalou, 2007; Tomasello & Mervis, 1994).

2.5.2 Direct Measures

2.5.2.1 Convergence between Parental Reports and Direct Measures

While, on a general level, previous research has found moderate to strong correlations between parental reports on the CDI and direct measures of infants' and toddlers' word knowledge (Fernald & Marchman, 2012; Fernald et al., 2006; Friend, Schmitt, & Simpson, 2012; Hurtado, Marchman, & Fernald, 2008), the evidence is inconsistent on an item-level. For instance, Houston-Price et al. (2007) revealed underestimation of parental reports when compared with child comprehension operationalized as visual gaze preference utilising indirect, eye-tracking measures. When infants, aged between 1 and 2-year-old, heard the target's label, the preference for the target image increased significantly from *baseline* (before the words were heard and no significant preferences for either images) on both name-known and name-unknown (as reported by their parents) trials – suggesting an underestimation of parental reports (Houston-Price et al., 2007). However, in other visual gaze preferential studies, significant alignments were found between parental reports and infant comprehension (Styles & Plunkett, 2008; Syrnyk & Meints, 2017). Infants' preference for the target image increased significantly from baseline on words that were reported as known, but did not change significant for unknown words (17-18-month-old infants: Styles and Plunkett (2008); 9-month-old infants: Syrnyk and Meints (2017)).

Inconsistencies between parental reports and direct measurements of comprehension may be a result of the immaturity of children's early lexical-semantic representations, making it difficult for parents to assess whether their child understands a particular word. Previous research has demonstrated that infants and toddlers employ a range of clues to disambiguate words, rather than a one-to-one word–object mapping, thus indicating that their early word representations are (semantically) *coarse*. For instance, infants fail to disambiguate semantically/functionally related items at 6 months of age (Bergelson & Aslin, 2017a), and at 8 months, they struggle to disambiguate frequency matched items in child-directed speech (Kartushina & Mayor, 2019). Although word–object mappings develop over time and greater semantic specificity is acquired by the age of 18–to-20-month-olds (Bergelson & Aslin, 2017b), early word representations remain imprecise by the end of the second year (Arias-Trejo & Plunkett, 2010). Arias-Trejo and Plunkett (2010) found that 18–to-24-month-olds had difficulty distinguishing between items that were both perceptually and semantically related (e.g., an apple and an orange), as opposed to items that were semantically related but perceptually dissimilar (e.g., an apple and a banana). This implies that the presence of a perceptually similar distractor increases the burden of visual discrimination due to feature overlap for semantically related objects.

In addition, studies that used direct measures of comprehension (operationalized as a touch response) reported moderate item-level agreement (Friend et al., 2012; Friend & Zesiger, 2011). Friend and Zesiger (2011) tested 16-19-month-old infants with a word list consisted of an equal proportion of easy words (understood by more than 66% of 16-month-old children), moderately difficult words (understood by 33% -66%) and difficult words (understood by less than 33%) defined a priori based on data from parental reports in a given language. The response profile according to the difficulty of the words moderately conforms to what was expected, with mixed results – the children were more accurate in selecting targets for easy words than for difficult words, with mixed results for moderately difficult words, in English-speaking children (results from all three levels were consistent with the general profile), French-speaking children (similar performance in easy and moderately difficult words) and Spanish-speaking children (performed the worst in moderately difficult words) (Friend & Zesiger, 2011).

In summary, the imprecision of parental reports discourages sole reliance on them when utilised as a basis for decisions in a clinical setting (Yoder, Warren, & Biggar, 1997). As a result, alternative measures to parental reports are recommended (Dale et al., 2003; Fenson et al., 1993), and their validity must be further examined.

2.5.2.2 Toddler-Directed Assessment Tools

A direct language measure can serve both as a convergent and a supplemental measure of parental reports. While many structured tests are available to assess young children's vocabulary knowledge, such as the Peabody Picture Vocabulary Test (Dunn, 2018) and the Expressive Vocabulary Test (Williams, 2018), direct measures that are appropriate for assessing children under two years of age remain scarce. It is a challenge for young children to comply with tests that require a motor response (e.g., pointing, manipulating objects), as they are unable to reliably provide manual response to stimuli (Gurteen, Horne, & Erjavec, 2011; Hendrickson & Friend, 2013). By removing the need for a volitional response (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013), looking-based measures such as the Intermodal Preferential Looking Paradigm (Golinkoff & Kerr, 1978; Hirsh-Pasek & Golinkoff, 1996) and the Looking-while-listening task (Fernald et al., 2006; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998) have been successfully used with infants as young as 4 months old. Yet, the passive and repetitive nature of such measures may quickly lead to boredom among older toddlers, due to the inherent difficulty in maintaining infants' and toddlers' interest and attention (Friend & Keplinger, 2003), thus rendering an extensive assessment impracticable. In parallel, the Computerised Comprehension Task (CCT) (CCT; Friend & Keplinger, 2003) has been shown to be effective in maintaining children's attention, circumventing the need for advanced motor skills, and improving compliance in toddlers aged 16 to 24 months (Friend & Keplinger, 2003; Friend & Zesiger, 2011; Friend et al., 2012; Hendrickson, Poulin-Dubois, Zesiger, & Friend, 2017; Poulin-Dubois, Bialystok, Blaye, Polonia, & Yott, 2013).

The CCT consists of four training trials, followed by 41 test trials and 13 reliability trials. The experimenter prompts the child to point to or touch an image in response to the target word heard (e.g., Where is the apple? Touch apple!) between a pair of images that is presented on a touchscreen. The words in the CCT, which include nouns, verbs, and adjectives, are drawn from the CDI–WG and CDI–WS and with varying difficulty based on norming data at the age of 16 months. Correct tap responses are interpreted as evidence of children's decontextualised word knowledge (Friend, Smolak, Patrucco-Nanchen, Poulin-

Dubois, & Zesiger, 2019). The reliability and validity of this measure have been well established across three languages (including bilinguals), with scores significantly correlating with CDIs (Friend & Keplinger, 2003, 2008; Friend et al., 2012; Friend & Zesiger, 2011; Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Poulin-Dubois et al., 2013). Despite the touchscreen nature of the task, CCTs are employed in laboratories and require frequent full arm movements due to screens being typically mounted on the wall or placed on a desk.

Other alternatives, such as tablet devices, In parallel, require only minimal motor movements, are easy to use even for very young children, and are more portable due to the size of the device. Toddlers aged 2-year-old were found to be capable of reliably performing both the tap and drag/slide gestures (Azah et al., 2014) and swiping the screen without assistance from an adult to perform tasks such as turning the pages of electronic books (Marsh, 2015). In addition, toddlers aged between 17- and 26-month-old have been demonstrated to be more attentive and engaged when reading electronic picture books (on a tablet) than when reading printed picture books with identical content (Strouse & Ganea, 2017). Older children aged 3- to 6-year-old were also found to be rarely discouraged when learning to draw on tablets (Couse & Chen, 2010). Hence, implementing tasks such as CCT on a tablet device provides a promising ground for novel assessment that is toddler-directed, thus providing a direct-measure assessment to supplement parent reports. We will explore the viability of a toddler-directed word recognition tool using tablets in Chapter 4, Study 6.

2.6 Summary

This chapter discussed early vocabulary acquisition and how internal and external factors modulate this process. I discussed findings from studies that have relied on the administrations of CDIs as well as other studies that have focussed on early vocabulary trajectories, i.e., vocabulary patterns across age, gender differences, and on the effect of relative language exposure on children's early vocabularies. I discussed findings that demonstrated how socioeconomic status and language use impact early language development. Together, these studies provide a basis for evaluating the Malaysian early vocabulary data collected for the present thesis, using parental reports. In addition, this chapter discussed the relationships between parental attention patterns, language-specific factors, and cross-cultural differences in shaping the composition of early vocabularies, in particular with respect to the noun-bias. We hope that, with the Malaysian data, the evaluation of the relative verb-noun distribution of monolingually- and bilingually-exposed children will shed more light on the emergence of differences in verb-noun distribution across languages and cultures.

With regards to early vocabulary assessment, this chapter highlighted the need to improve the efficiency of existing CDIs. This can be achieved by combining Mayor and Mani (2018)'s approach – to estimate full CDI scores by combining responses on a subset of test items with prior knowledge about children vocabularies and Makransky et al. (2016)'s approach – to dynamically select words that optimally discriminate the language skills of children during the test. Furthermore, inconsistencies in parental reports, as shown in previous studies, prompted the need for direct-measure assessment tools to supplement these reports. Hence, a tablet-based toddler-directed word recognition task was introduced as a suitable candidate to measure children's word comprehension.

Chapter 3

Early Vocabulary Assessment using Trilingual CDIs

3.1 Overview

One of the main goals of this thesis was to collect the first sample of early vocabulary data for infants and toddlers in Malaysia, using the Multilingual Communicative Development Inventories – Malaysia version (MCDI-M), an adaptation of the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007) developed by Low (2010). As a trilingual CDI, the Malaysian version evaluates children vocabulary knowledge and use in Malay, Mandarin and English, thus covering some of the most prevalent standard languages in Malaysia (H. W. Lim et al., 2015), excluding Tamil (a candidate for future version of the form) and regional dialects. Every conceptual item in the Malaysian CDI is presented with translation equivalents in all three languages. The trilingual CDI consists of $1,800$ vocabulary items – 600 conceptual items presented in three languages – in contrast to 680 words in the original American English CDI-WS.

Study 1 investigated the effectiveness of the trilingual CDI at capturing developmental trends described in previous research. Specifically, we examine the early vocabulary development trajectory across age, of young Malaysian children as it is important that the trilingual CDI captures overall patterns of vocabulary growth, characterised by an increase in vocabulary size as age increases. In addition, we will also examine whether there were any potential gender differences in early vocabulary acquisition, along with language differences due to social statuses of the languages.

Study 2 investigated the effect of income and education on both languages of Malaysian children from bilingual families, as EL (Malay for Malay-ethnic and Mandarin for Chinese-ethnic) hold an important position in society and functionally co-exist with the English language in some domains of education, business and communication. The objectives of this study were to investigate:

1. Whether SES, in terms of educational and income level, affects language preference between EL (i.e., Malay or Mandarin) and the English language, in bilingual families.

2. Whether vocabulary differences in comprehension (in both languages) associated with differences in SES.

3. Whether vocabulary differences in production (in both languages) associated with differences in SES.

4. Whether relative language exposure mediates the relationship between SES and vocabulary size in comprehension and production (if present).

Study 3 was conducted with the aim of further examining the role of sociocultural influences on language acquisition patterns, in particular on the *noun bias*. While previous studies with bilingual children have taken the approach of assessing differences across their languages learnt within a single learning environment (e.g., Chan & Nicoladis, 2010; Xuan & Dollaghan, 2013), we adopted the distinct, and complementary, approach of considering the impact of sociocultural differences on their learning environment when learning a common language. Specifically, we compared the English verb-noun ratio of Mandarin-English bilingually-exposed children with the English of Malay-English children – any differences in the English verb-noun ratio would suggest that the noun bias is modulated by factors external to that language.

In sum, across these 3 first studies, we analyse the vocabulary acquisition of young monolingual and bilingual learners in Malaysia, where the native language is often the dominant language in their immediate social environment, but other languages (e.g. English) are needed when communicating with people from different ethnic-language groups. Thus, the objective of this chapter was to investigate whether the MCDI-M can capture these influences of age, gender, language exposure, and SES on the development of word production and comprehension in young children between 8- and 36-month-old.

3.2 Methods

3.2.1 Participants

For this PhD project, I collected 257 samples using the online version of the MCDI-M. This sample, aged between 6- and 46-month-old (*M* = 23.70, *SD* = 9.20), consisted of 119 girls and 138 boys. In addition, this thesis draws on unpublished samples collected by my supervisors, Dr Julien Mayor, Dr Tze Peng Wong and Dr Hui Min Low, using the paper version of the MCDI-M. This group of samples consisted of 308 children (150 girls and 158 boys), aged between 12 and 46-month-old (*M* = 23.90, *SD* = 5.77).

The current chapter reports the use of MCDI-M in collecting the first Malaysian early vocabulary data and provides insights into Malaysian children's early language development, in which three research questions were formulated. Three subsets of the sample were extracted to address each research question.

Study 1 aimed to examine the early vocabulary trajectories of Malaysian children using the MCDI-M. The objective is to investigate whether the MCDI-M can capture the influences of age, gender, and language exposure on the development of word production and comprehension in young children, as shown in previous CDI studies (See Chapter 2). Children aged between 8- and 36-month-old were extracted as the CDIs are most often designed and applied to children of this age group. Children who were identified with developmental delay, as reported by the parents in the CDI forms were excluded. This subset of the sample consisted of 240 children from the online forms and 301 children from the paper forms.

Study 2 quantified the environmental effects on vocabulary acquisition of young children from bilingual families. Only the online sample was used, given the information about household income and educational attainment was available from the online sample but not the paper sample. A subset of 155 typically developing children aged between 8 and 36 months old (M = 21.85, SD = 8.16) were extracted for data analysis, with Malay ethnic (N = 113, exposed to the Malay and English language) and Chinese ethnic (N = 42, exposed to the Mandarin and English language) children. Bilingually exposed children with exposure ranging from 5% to 95% in both languages were included to investigate parents' language preferences (in percentage, on a continuous scale) from various SES groups and their mediation effects on children's vocabularies.

Study 3 examined the role of sociocultural influences on language acquisition patterns. A joint sample from the online and paper forms was used. Since the focus of the study is not on bilingualism per se, but rather on comparing sociocultural influences on language structure, we adopted a generous inclusion criterion; infants classified as bilinguals were defined as having non-zero exposure to two languages. Relative exposure to English, as reported by their parents, was similar across bilingual groups, $t(38.77) = 1.59$, $p = .12$. Hence, a subset of 514 Malaysian infants and toddlers (248 females and 266 males) were extracted for this study.

Among this subset of sample, all Malay learners were ethnically Malay and all Mandarin learners were ethnically Chinese. 117 children were exposed to Malay only, with an age range between 7- and 45-month-old ($M = 24.80$, $SD = 8.43$); 22 were exposed to Mandarin only, with an age range between 15- and 45-month-old ($M = 23.00$, $SD = 5.15$); 297 were exposed to both Malay and English, with ages ranging from 6- to 48-month-old ($M = 24.00$, $SD = 8.43$); and 78 were exposed to both Mandarin and English with an age range of 7- to 45-month-old ($M = 22.50$, $SD = 8.43$).

3.2.2 Demographics

As reported in Table 3.1, the gender distributions of both paper-based and online-based sample closely matched those of the population. Some differences were observed for the ethnic distribution – the paper-based sample had much more Malay-ethnic group than other ethnic groups. Yet, it is noteworthy that the demographic composition of the online-based sample was a closer match with the demographic of Malaysia (Department of Statistics Malaysia, 2020) when compared with the paper-based sample, suggesting an advantage of online sampling.

Table 3.1. Gender and ethnicity from the Malaysian CDI data (paper and online) in comparison with the Malaysian population based on statistics by Department of Statistics Malaysia (2020).

	Online CDI		Paper CDI		Population
	n	%	n	%	%
Gender					
Girls	120	46.0	150	48.7	48.4
Boys	141	54.0	158	51.3	51.6
Ethnicity					
Malay	161	61.7	261	84.7	67.4
Chinese	70	26.8	43	14.0	24.6
Indian	18	6.9	4	1.3	7.3
Others:	12	4.6	-	-	0.7

Furthermore, the compositions of language exposures from the paper-based and online-based sample were compared in Table 3.2. Again, there were some differences in the distribution of language exposure groups – the paper-based had a majority of Malay monolinguals and Malay-English bilinguals (more than 80%). In contrast, the online sample had a more evenly distributed language exposure groups, with 58.5% of Malay speakers from monolingual and bilingual groups.

In terms of the time spent to complete the online CDI form, Figure 3-1A reports that 231 of these forms were completed in less than 24 hours (*Min* = 13 minutes, *Max* = 23 hours 36 minutes, *Median* = 57 minutes, *Median Absolute Deviation (MAD)* = 37 minutes). Among them, a majority (81%) were completed in the range of 25 minutes to 2 hours and 30 minutes; 30 parents spent more than 5 hours to complete the forms – it is likely that these parents took a break or were interrupted but decided to continue the forms later on.

Table 3.2. Language exposures from the Malaysian CDI data, a comparison between paper and online sample.

	Online CDI		Paper CDI	
Language exposure	n	%	n	%
Monolingual [a]				
Malay	34	13.0	139	45.1
Chinese	17	6.5	23	7.5
English	18	6.9	-	-
Balanced bilingual [b]				
Malay-English	21	8.0	8	2.6
Chinese-English	10	3.8	-	-
Unbalanced bilingual [c]				
Malay majority - English minority	98	37.5	110	35.7
Chinese majority - English minority	24	9.2	19	6.2
English majority - Malay minority	19	7.3	4	1.3
English majority - Chinese minority	12	4.6	5	1.6

Note. The language exposure groups were classified with reference to Thordardottir (2011).

[a] It is defined as having 95% and above exposure to a particular language, as 100% monolingual exposure is very rare in the Malaysia.

[b] It is defined as being exposed to two languages, each between 40% and 60% exposure.

[c] It is defined as being exposed to a dominant language (between 61% and 94%) and a minority language (between 6% and 39%).

Figure 3-1B reports that 9 of the forms were completed in more than 24 hours: 8 of these parents completed the forms within a week; one form was completed after 38 days and is excluded from the analysis given that vocabulary size can change a lot over a month.

3.2.3 The Malaysian Trilingual Adaptation of the CDI

The MCDI-M is available in both paper-based and online, digital version. The first batch of parents were recruited to complete the paper-based form in 2016 and the second batch of parents completed the online version between 2018 and early 2019. The paper-based forms were distributed to Malaysian parents with the help of undergraduate students from the University of Nottingham Malaysia and Universiti Sains Malaysia. The website link to the online version was distributed to parents over the internet, via parental groups and pages on Facebook, messaging applications (e.g. WhatsApp and Messenger) and recruitment flyers at childcare centres. Similar to the paper-based forms, the online version is presented in three languages, English, Malay and Mandarin.

The parents were asked to assess the word knowledge of their child, word-by-word, by picking between three options – 'Does not understand', 'Understands' or 'Understands and speaks'. The online version was expected to be less cognitively taxing and more time-efficient as only words from the language(s) that the child is exposed to as reported by their

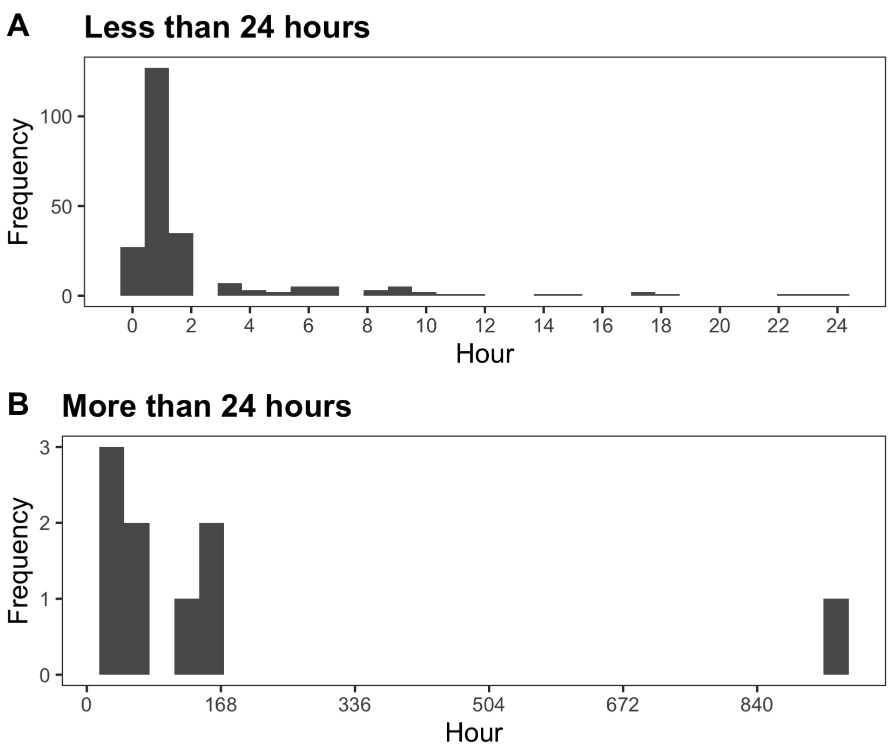


Figure 3-1: *The histogram of the duration in hours parents took to complete the online CDI forms in:* ***A)*** *Less than 24 Hours;* ***B)*** *More than 24 Hours.*

parents are displayed, rather than the full list of 1,800 items from the paper-based version. The comprehension count for the children is calculated by summing the words that were either labelled as 'Understands' or 'Understands and speaks' whereas the production count is calculated by summing the words that were labelled as 'Understands and speaks'. Demographics such as ethnicity, language exposure, parental educational level and household income were asked in the survey. (See Appendix for the complete list of demographical questions and vocabulary items on the list.)

The study was approved by the Science & Engineering Research Ethics Committee (SEREC, JHC250917). The parents were reminded that their sensitive information would be kept confidential and that no identifiable information would be made public during data collection, analysis and publication of results. This was emphasised due to the sensitive nature of the information collected, i.e. the child's gender, ethnicity, birthday, history of health, mother's name and occupation, household income, parents' highest education level, family's health history and the state of residence.

3.2.4 Comparison between the MCDI-M, American English CDI-WS and Beijing Mandarin CDI-WS

A comparison of the composition of semantic categories between conceptual items of the MCDI-M (consisted of vocabulary in Malay, English and Mandarin, but not Tamil) and total vocabulary items of the American English CDI-WS (the original CDI forms which the adaptation of the trilingual version was based on, and English is one of the languages included in the trilingual CDI; Fenson et al., 2007) and the Beijing Mandarin CDI-WS (Mandarin is one of the languages included in the trilingual CDI; Tardif & Fletcher, 2008) is presented in Table 3.3. Both the American English CDI-WS and Beijing Mandarin CDI-WS were extracted from the Stanford Wordbank[1] (Frank et al., 2017). The term *conceptual* is defined as word knowledge in terms of conceptual units, e.g., the word *ball* and the word *bola* are treated as one conceptual unit representing the concept of a ball. When assessed, children reported using the word *ball* and the word *bola* will only receive a score of 1 due to

[1] There were no Malay or Tamil CDIs to extract for comparisons.

Table 3.3. Comparison of the number of items in each category in the vocabulary lists of Malaysian, English (American) and Mandarin (Beijing) CDIs.

	Malaysian	English (American)	Mandarin (Beijing)
No. of items*	600	680	710
Nominals			
Animals	52	43	51
Vehicles	14	14	14
Toys	12	18	24
Food & drinks	42	68	82
Clothings	23	28	27
Body parts	26	27	39
Household items	48	50	63
Furnitures & rooms	27	33	33
People	30	29	32
Places to go	-	22	19
Outside things	49**	31	35
Sound effects & animal sounds	-	12	-
Total	309	375	419
Non-nominals			
Games & routines	24	25	30
Action words	101	103	124
Descriptive words	63	63	75
Words about time	13	12	13
Pronouns	17	25	19
Question words	7	7	9
Prepositions & locations	22	26	-
Quantifiers & articles	16	17	9
Helping verbs	8	21	6
Connecting words	6	6	6
Total	291	305	291

*Conceptual items are reported for the trilingual Malaysian CDI. Full list of items are reported for the other CDIs.
**19 out of 49 items in this category are translation equivalents of items from the *Places to go* category in other CDIs.

both words being translation equivalents. Conceptual vocabulary is often used as an index of language ability in bilingually-exposed children(Anaya, Peña, & Bedore, 2018; Bedore et al., 2005; Core et al., 2013; Gross, Buac, & Kaushanskaya, 2014; Mancilla-Martinez, Pan, & Vagh, 2011).

In terms of conceptual items, the Malaysian version has 11.8% fewer items compared to the original American English CDI-WS and 15.5% fewer items than the Beijing Mandarin CDI-WS. Most changes were made in the nominal group of word categories, with the most reductions in the food & drinks category. In contrast, non-nominal items remained relatively unchanged, with most reductions in the Helping verbs category – a 60% decrease from the American version but similar to the Mandarin version. This effectively reduced the total items needed to be assessed in the Malaysian CDI from $2,040$ items ($680*3$: as if the 680 items in the American English CDI were directly translated into three languages in the Malaysian trilingual version) down to $1,800$ items. Among the three versions of CDIs compared, the MCDI-M contained the least number of conceptual items to minimise the total vocabulary items presented to the parents. Although the number of conceptual items in the MCDI-M was reduced from 680 in the original American English CDI to 600, the proportion of words among different semantic categories was essentially maintained. We hope the reduction of total vocabulary items presented to parents could improve data collection efficiency by reducing time spent on completing the forms and drop-out rates, yet without compromising the quality of the assessments (by preserving the proportion of semantic distributions).

The need for adapting CDIs into different cultures cannot be overstated. Yet, the alteration of CDI also poses a threat to comparisons across different CDIs. Hence, to improve generalisability, the replacement words have to resemble the semantic meaning as closely as possible with the original. A closer inspection of the MCDI-M adapted by Low (2010) revealed cultural differences in word choice with the American English CDI. Words such as *gas station*, *mad* and *candy* were replaced with semantically similar words that are more culturally-relevant in Malaysia – *petrol station*, *angry* and *sweet* in the trilingual adaptation of the CDI (Low, 2010). Words that are more general in meaning were replaced and broken down into lower hierarchical words, such as *mosquito* and *housefly* from *bug* and *corn flakes* from *cereal* to fit the Malaysian context. Furthermore, while both CDIs included *church*, the Malaysian CDI included two more religious places – *mosque* and *temple*, reflecting the multi-religious aspect of Malaysian culture.

3.2.5 Growth Curve based on the CDIs data

When measuring vocabulary sizes, it is typical to observe a large variation in terms of total scores as each child, even of the same age, develops at a largely different pace due to individual differences (Fenson et al., 1994). It is important, however, in cross-sectional CDI studies, that these variations follow a trend – the vocabulary size increases as age increases. Figure 3-2 and 3-3 depict the growth curve[2] in the normative sample (retrieved from Wordbank) of American children aged between 8 and 18 months old (in comprehension) and between 16 and 36 months old (in production), with an increase in vocabulary size by month of age, across all percentile groups.

When comparing MCDI-M data with the American English CDI-WG data in Figure 3-4, a similar vocabulary growth was observed in the current MCDI-M sample of infant data (aged between 8- and 16-month-old), with the exception at 75th and 90th percentile. We argued the lower growth at 75th and 90th percentile in the American English CDI-WG was due to a ceiling effect (400-item list with a maximum vocabulary score of 400) rather than a real delay in development. In addition, the growth curve of the current MCDI-M sample of toddler data (aged between 16 and 36) was identical to that of the American Wordbank data set (see Figure 3-5 for MCDI-M; Figure 3-3 for American English CDI-WS; except with a visibly slower growth rate at 10th percentile).

3.2.6 Children's First Words

Table 3.4 reports the first 10 words for children who can say 1-10 words on MCDI-M and includes Malay vocabulary from Malay-ethnic children, Mandarin vocabulary from

[2]The curves were computed using the *gcrq* function in *R* (Muggeo, Sciandra, Tomasello, & Calvo, 2013). They are estimated via B-splines of degree 3, with a L1 penalty on the spline coefficient differences, and are restricted to monotonously increasing to obtain curves resembling growth. A growth curve typically begins with a low increase rate, followed by exponential growth, and then enters a stationary phase.

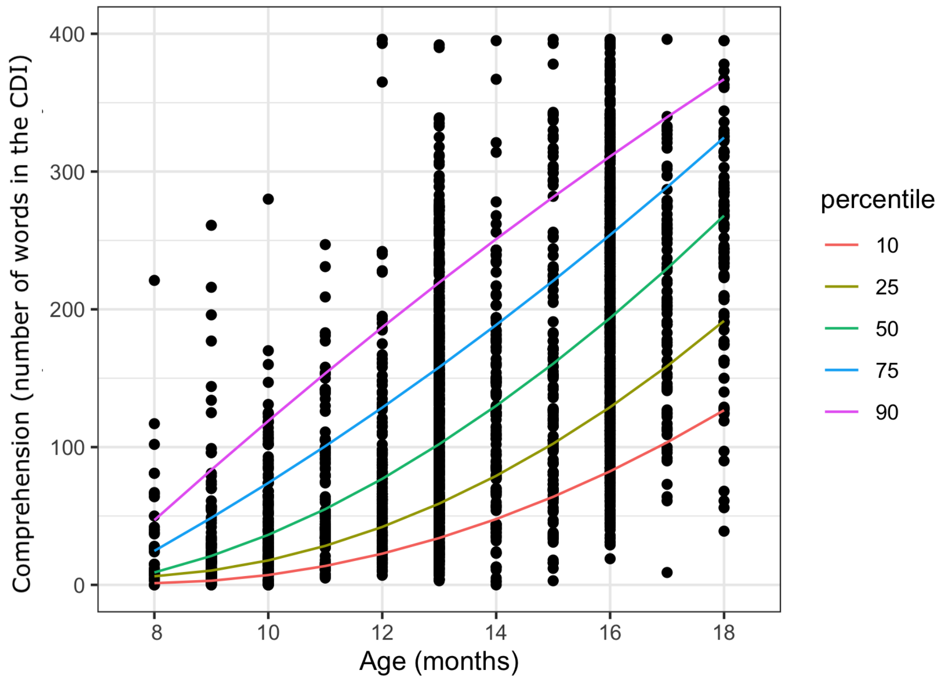


Figure 3-2: *Growth curve of American English CDI for comprehension of child aged between 8 and 18 month, by percentiles (10th, 25th, 50th, 75th, 90th).*

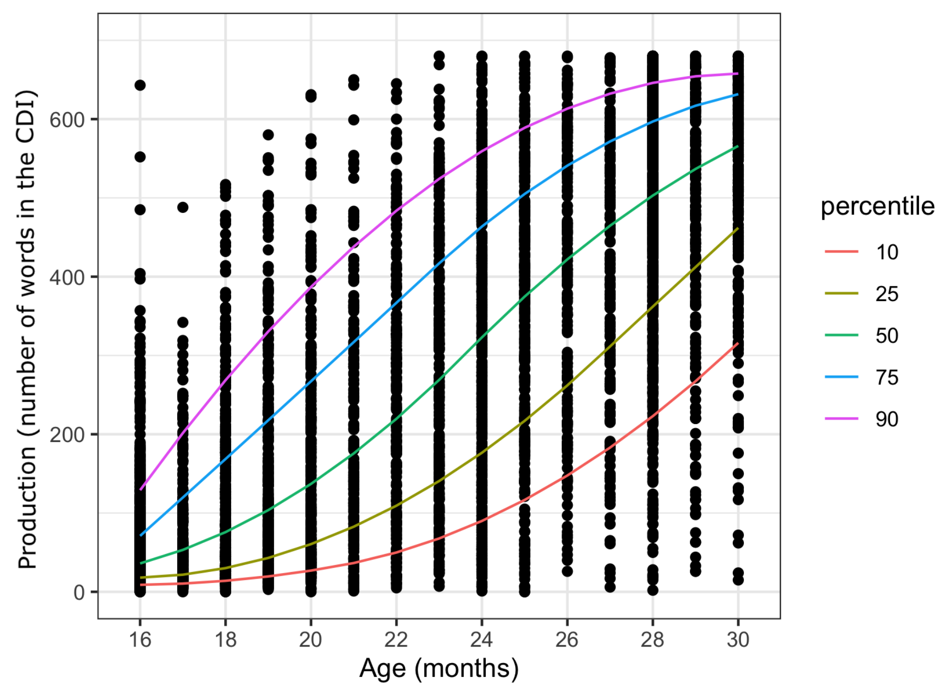


Figure 3-3: *Growth curve of American English CDI for production of child aged between 16 and 30 month, by percentiles (10th, 25th, 50th, 75th, 90th).*

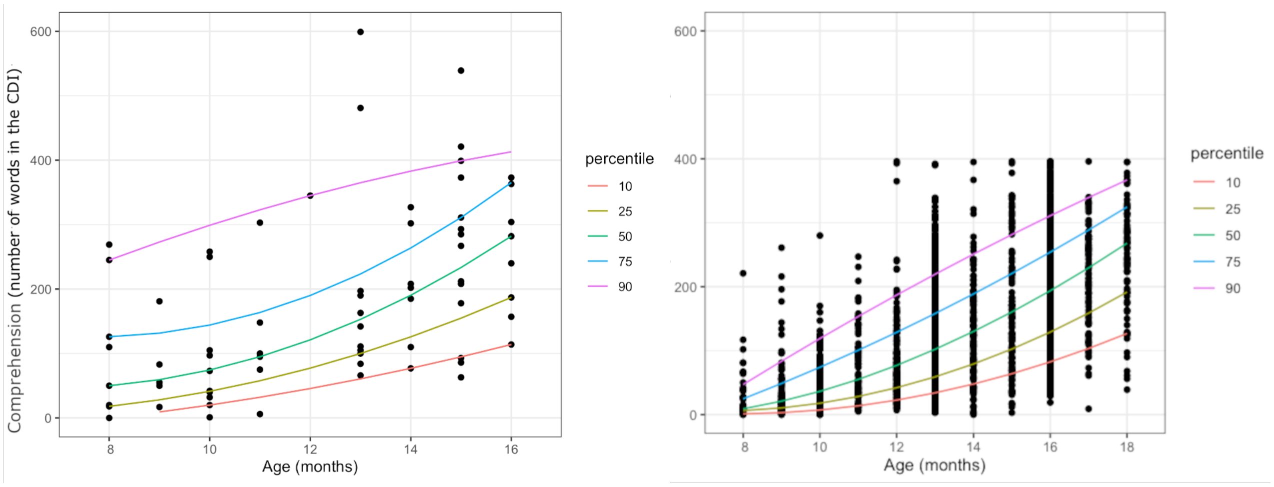


Figure 3-4: *Growth curve of Malaysian CDI for conceptual comprehension of child aged between 8 and 16 months, by percentiles (10th, 25th, 50th, 75th, 90th) with a scaled Figure 3-2 for comparison.*

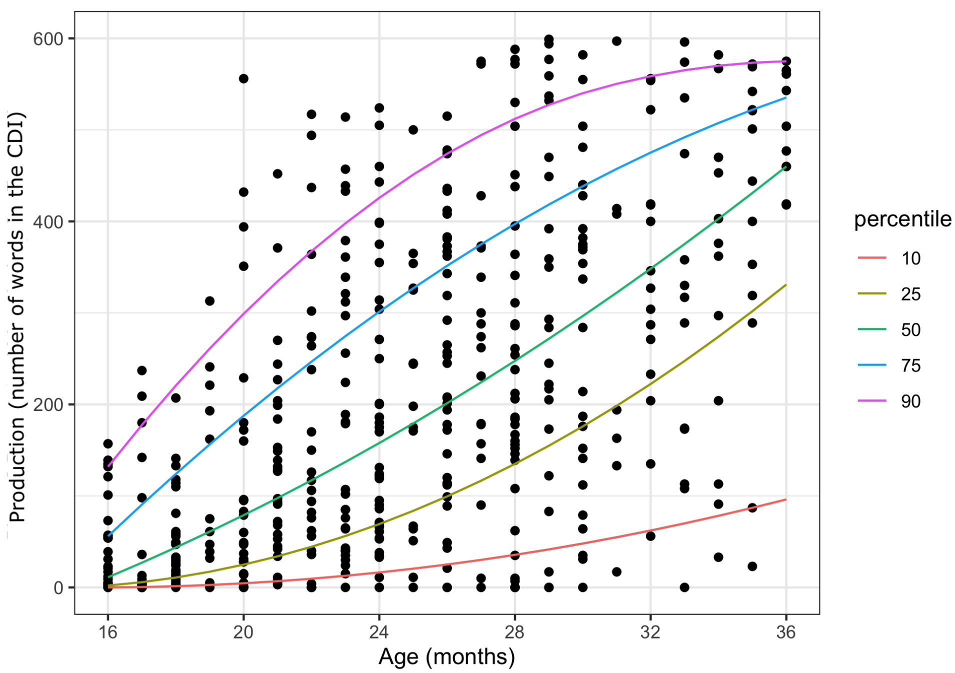


Figure 3-5: *Growth curve of Malaysian CDI for conceptual production of child aged between 16 and 36 months, by percentiles (10th, 25th, 50th, 75th, 90th).*

Chinese-ethnic children, and English vocabulary from bilingually-exposed children; Malay-ethnic children exposed to Malay and English, and Chinese-ethnic children exposed to Mandarin and English. Only vocabulary data in production was analysed, but not compre-

Table 3.4. The first 10 words (ordered by frequency) for children who can say 1-10 words on CDI, by language. Note that English (MalE) represents the English vocabulary of Malay-English exposed Malay children, whereas English (ManE) represents the English vocabulary of Mandarin-English exposed Chinese children.

Malay	Mandarin	English (MalE)	English (ManE)
bye-bye	older sister	baby	bye-bye
daddy	daddy	ball	(peek-a-) boo
mummy	mummy	cat	bird
milk	older brother	bye-bye	dog
shhh	no	fish - animals	bath
cat	maternal grandmother	daddy	hello
water - food	paternal grandfather	mummy	night-night
(peek-a-) boo	shhh	fish - food	no
hello	hit	hello	mummy
older sister	bye-bye	car	daddy

hension data, as the majority of children in our sample were reported to understand more than 10 words. There are both similarities and differences in the words that children were reported to produce as their first 10 words in Malay, Mandarin and English languages. Notably, similarities across languages, such as *mummy* and *daddy* – the informal term for father and mother – were among the first 10 words produced by children across languages, as well as other terms referring to people (*older sister*), household pets (*dog*, *cat*) and social routines (*hello*, *bye-bye*, *shhh*), most of which are also common first ten words acquired by children across other languages, as reported in Frank et al. (2021).

Despite commonality, comparisons of Malay, Mandarin and English revealed clear cultural differences. Six out of the first 10 Mandarin words consisted of kinship terms (e.g., parents, grandparents, older siblings) compared to 3 out of 10 in Malay and 2 out of 10 in English for Malay-ethnic and Chinese-ethnic children. Chinese-ethnic children, in our sample, were more likely to use kinship terms in Mandarin, as reported by their parents, than in other languages. However, it is noteworthy that, as a Confucian culture that stressed

Table 3.5. The 10 most frequent words produced by Malay, Mandarin and English learners (% in parentheses). English (MalE) represents English vocabulary of Malay-English exposed Malay children whereas English (ManE) represents English vocabulary of Mandarin-English exposed Chinese children. Note: These items were not controlled for age and lingualism status.

Malay	Mandarin	English (MalE)	English (ManE)
mummy (74)	daddy (80)	ball (34)	bye-bye (50)
bye-bye (73)	mummy (73)	baby (33)	baby (46)
daddy (72)	older sister (70)	cat (32)	hello (45)
ball (67)	older brother (68)	bye-bye (30)	ball (45)
want to (67)	auntie (64)	fish - animals (29)	mummy (42)
cat (67)	carry (64)	no (28)	car (42)
milk (67)	bye-bye (61)	fish - food (28)	bird (42)
water - food (65)	no (61)	hello (27)	apple (40)
chicken - animals (65)	water - food (61)	car (26)	banana (40)
water - outside (64)	dog (59)	shhh (24)	dog (40)

order between senior and junior (T. Y. Lee, Yu, & Nah, 2011), the most common Chinese addressing practice is to address both kin and non-kin using kinship terms such as *a1 yi2* (auntie), *shu1 shu* (uncle), *jie3 jie3* (elder sister in Mandarin) and *ge1 ge1* (elder brother in Mandarin; Wu, 1990; S. Lee & Shanmuganathan, 2020). It is unclear whether the extension of kinship terms to non-kin is common in the Malay community due to a lack of literature. Limited evidence suggests the common use of the word *aunt/aunty* in both the Chinese and Malay communities to address older females across different languages and regardless of kinship, as shown in a discourse study conducted by S. Lee and Shanmuganathan (2020).

Table 3.5 reports the 10 most frequent words produced by children whose vocabulary grows beyond the first 10 words in Malay, Mandarin and English. Comparisons of Table 3.4 and Table 3.5 revealed similarities within each language. Overall, 6 out of 10 Malay words in Table 3.4 were present in Table 3.5, in that language; 6 out of 10 Mandarin words in both tables, in that language; 8 out of 10 English words of Malay children in both tables, in that language; and 5 out of 10 English words of Chinese children in both tables, in that language. This suggests that, the first words most children learn to produce (see Table 3.4) are also the words most children produce, even after they grow beyond their first 10 words (see Table 3.5).

Comparing vocabulary composition, Chinese children were reported to produce the verb *hit* among the first 10 words in Mandarin, with no verbs reported in other languages. Similarly, the verb *carry* was among the 10 most frequent words produced in Mandarin, yet no verbs were present in other languages. The higher prevalence of verbs in Mandarin (a verb-friendly language) relative to English (a noun-friendly language) and Malay (unknown[3]) in early vocabulary acquisition is examined in terms of verb-to-noun ratio in Study 2.

3.3 Study 1: Developmental Trends, Gender Differences & Language Exposure.

3.3.1 Purpose

Study 1 examined the early vocabulary acquisition trajectory of Malaysian young children using the MCDI-M and gender differences were investigated. With a particular interest in the modulation of vocabulary acquisition by relative language exposure, linear models were also fitted with language exposure.

We investigated the presence of developmental trends (age in months), the effect of language exposure (in percentage) and gender, as well as language differences (Malay, Mandarin, and English), and sampling methods (to reveal any potential systematic differences between paper and online collection of data), and their interactions in predicting vocabulary sizes in comprehension and in production, respectively. We expected a positive relationship between age and vocabulary size, signalling vocabulary growth. We expected a positive relationship between language exposure and vocabulary size, as the relative amount of language input received by the children shapes their language experience, thus influencing the

[3]We aim to postulate the language status of Malay relative to Mandarin and English, in Study 2.

acquisition of the languages. We also expected an interaction between age and language exposure, as younger children might not acquire any words regardless of exposure, and the differences in vocabulary sizes due to language exposure will become more pronounced as children grow older. We also expect gender differences – girls will outperform boys in their vocabulary sizes. In terms of language differences, we expected larger vocabulary sizes in ethnic languages (e.g., Malay and Mandarin).

3.3.2 Results

Two Linear Mixed Model (LMM)s were fitted maximally with age, language exposure, gender, language type and sampling type as fixed effects, and their interactions in predicting vocabulary sizes in comprehension and in production, respectively. The subject was included as a random effect in the model to control for individual differences. The LMMs were fitted using the *lmer* function from Kuznetsova, Brockhoff, and Christensen (2017)'s *lmerTest* package in *R*. *Marginal* R^2 and *conditional* R^2 were computed using the *r.squaredGLMM* function from Barton and Barton (2015)'s *MuMIn* package in *R*. *Marginal* R^2 represents variance explained by *only fixed effects* whereas *conditional* R^2 represents variance explained by *both fixed and random effects* (Nakagawa & Schielzeth, 2013).

3.3.2.1 Comprehension

Two evaluation tests were conducted for the LMM. The evaluation of the subject in the random effect model suggests that the addition of the factor did improve the model fit, $p < .001$, when compared to a basic linear model (with intercept and fixed effects only) in a Likelihood Ratio Test (LRT). The evaluation of the fixed effects revealed a significant contribution compared to a base LMM (with only intercept and random effect), $\chi^2(47) = 936.73$, $p < .001$. Together, the evaluations suggest that the LMM is an appropriate model. The fixed effects in the model accounted for 48.77% of the variance, whereas the random effect accounted for 11.89% of the variance with the remaining 39.33% of the variance unaccounted for by the model.

The output of the Analysis of Deviance using Type II Wald Chi-square tests of the model revealed significant interaction effects between exposure and age ($p < .001$), and between exposure and language type ($p = .024$). There were also significant main effects of exposure, of age, of language type and of sampling type (all *p*s < .001).

The interaction effect between language exposure and age was found to account for 30.14% of the *marginal* variance; the interaction effect between language exposure and language type for 3.65%; the main effect of sampling type for 0.34%; main effect of exposure for 1.82%; main effect of age for 3.92%; main effect of language type for 1.74%.

In order to visualise the effect of exposure differences (estimates plotted at 33%, 66% and 99% exposures) on the relationship between age and comprehension scores, a graph[4] of estimated *marginal* means (hereby referred to as estimates) was plotted. Figure 3-6 shows a positive growth in comprehension scores across age. It is noteworthy that the growth rate was modulated by language exposure – toddlers with higher relative exposure to a language tended to acquire more words, more quickly in that language.

[4]This graph, along with subsequent figures that visualise interaction effects (by calculating their respective estimated *marginal* means from their respective models) were created using the *ggeffect* package (Lüdecke, 2018) in *R*.

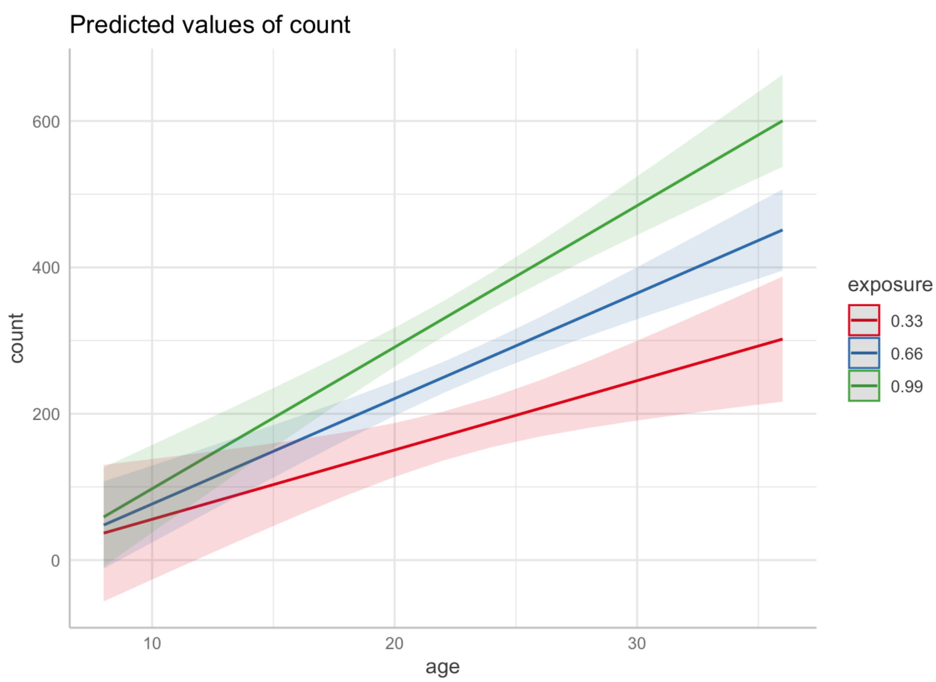


Figure 3-6: *Estimates plot for the interaction effect between proportion of language exposure (continuous variable categorised into groups of 33%, 66%, 99% exposure) and age in predicting the comprehension scores.*

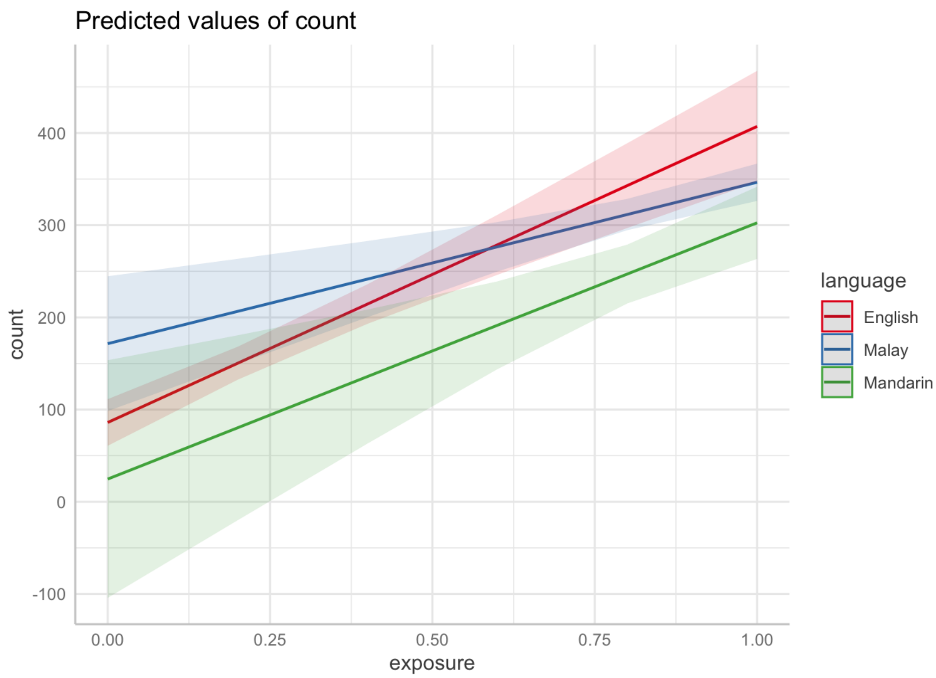


Figure 3-7: *Estimates plot for the interaction effect between proportion of language exposure and language type in predicting the comprehension scores.*

A graph of estimates was plotted to visualise the effect of language differences on the relationship between language exposure and comprehension scores. Figure 3-7 shows an overall increase in comprehension scores as language exposure increases for all languages. However, the slopes differed across languages – the slope of increase in Mandarin and English was similar, but Malay was the slowest among the language groups. In other words, toddlers with an increased amount of relative exposure to a language tended to have larger vocabulary sizes in that language, but the language exposure in Malay tended to have a weaker effect on the vocabulary size. It is noteworthy that all three languages had intercepts at non-zero points. This suggests that children acquired words in these languages despite their parents reporting no exposure.

The pairwise comparison showed that the comprehension scores from the online sample (*emmean* = 306, *SE* = 10.2) was significantly higher than the paper sample (*emmean* = 205, *SE* = 20.0), $p < .001$. Toddlers recruited via the internet, as reported by their parents, are more advanced in word comprehension than toddlers recruited via physical recruitment.

3.3.2.2 Production

Again, two evaluation tests were conducted to evaluate the significance of the LMM. The addition of the subject as a random effect significantly improved the model ($\chi^2(1) = 67.17$, $p < .001$). The evaluation of the fixed effect predictors and interactions also revealed a significant contribution ($\chi^2(47) = 889.38$, $p < .001$), when compared to the base LMM. Hence, a better model is achieved with the addition of the random effect, all fixed effects predictors, and interactions. The fixed effects in the model accounted for 45.18% of the variance, whereas the random effects accounted for 27.78% of the variance, with the remaining 27.05% of the variance unaccounted for by the model.

The output of the Analysis of Deviance using Type II Wald Chi-square tests revealed significant interaction effects between exposure and age, and between age and language type (both *p*s < .001), as well as between exposure and language type ($p = .010$), and between language type and sampling type ($p = .014$). There were also significant main effects of exposure, of age, of language (all *p*s < .001), and of gender ($p = .031$).

The interaction effect between exposure and age was found to account for 32.99% of the *marginal* variance; the interaction effect between age and language type for 6.53%; the interaction effect between exposure and language type for 2.75%; the interaction effect between language type and sampling type for 2.38%; the main effect of exposure for 11.68%; the main effect of age for 0.28%; the main effect of language for 3.28%; the main effect of gender for 1.20%.

In production, a graph of estimates was plotted to visualise the interaction effect between language exposure and age on prediction of vocabulary sizes. Figure 3-8 shows a positive growth in production scores across age and the growth rate was again modulated by language exposure. Although a major proportion of the '.66' and '.99' regression lines before the intersection (at 14-month-old) fall below zero count, this does not imply that children have a negative vocabulary count, but rather reflects the limitations of the current linear model[5], as Figure 3-9 depicts zero (and close to zero) vocabulary counts prior to the age of 14-month-old, thus demonstrating children's limited ability to verbalize words prior

[5] Although it is more ideal to fit a growth curve (introduced in Section 2.4.4), it was not attempted during the analysis, due to technical constrains.

to that age.

A graph of estimates was plotted to visualise the interaction effect between language exposure and language type. Similarly, Figure 3-10 shows an increase in comprehension scores as the language exposure increases for all languages. Like comprehension, the slope of increase in Mandarin and English was similar, but the Mandarin language had a lower starting point[6] than English. Although Malay had the highest starting point, the slope of increase was the slowest among the language groups, suggesting that the exposure was less influential on the Malay vocabulary.

A graph of estimates was plotted to visualise the interaction effect between age and language type. Figure 3-11 shows that, overall, production scores increase as age increases, but the slope of increase in Mandarin is slower than in English and Malay. While no interaction between age, exposure and language type was found, an extra graph of estimates was plotted to visualise the relationship between the variables to provide a full picture of

[6] Again, while a portion of the regression line for Mandarin fall below zero count, this does not imply that children have a negative vocabulary count, but rather reflects the limitations of the current linear model.

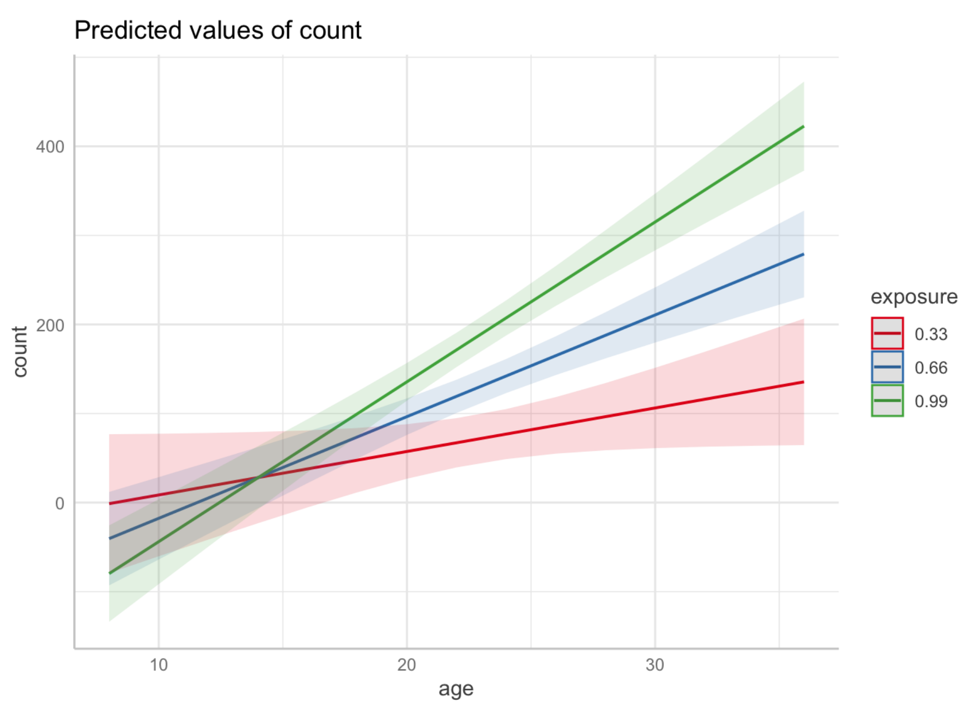


Figure 3-8: *Estimates plot for the interaction effect between proportion of language exposure (continuous variable categorised into groups of 33%, 66%, 99% exposure) and age in predicting the production scores.*

the interactions. Figure 3-12 shows that vocabulary size in English and Malay increases with age and exposure, but not in Mandarin. Surprisingly, for Mandarin, the vocabulary sizes stay at 33% of relative exposure, with a slower rate of increase at 66% of relative exposure than English and Malay.

The pairwise comparison for the main effect of gender showed that the expressive vocabulary of females (*emmean* = 146, *SE* = 12.5) was significantly higher than males (*emmean* = 103, *SE* = 15.1), $p = .028$. Female toddlers are more advanced than male toddlers in production.

The pairwise comparison for the interaction between sampling type and language type shows that, in the online sample, no significant differences were found between Malay (*emmean* = 129.5, *SE* = 11.3) and Mandarin (*emmean* = 90.8, *SE* = 15.9), $p = .097$, but English (*emmean* = 177.9, *SE* = 17.6) was significantly higher than Malay, $p = .001$, and Mandarin, $p < .001$. In paper sample, Malay (*emmean* = 152.4, *SE* = 17.3) was significantly higher than Mandarin (*emmean* = 46.9, *SE* = 31.4), $p = .007$, but showed no significant differences with English (*emmean* = 152.0, *SE* = 26.1), $p = 1.000$, and English was significantly

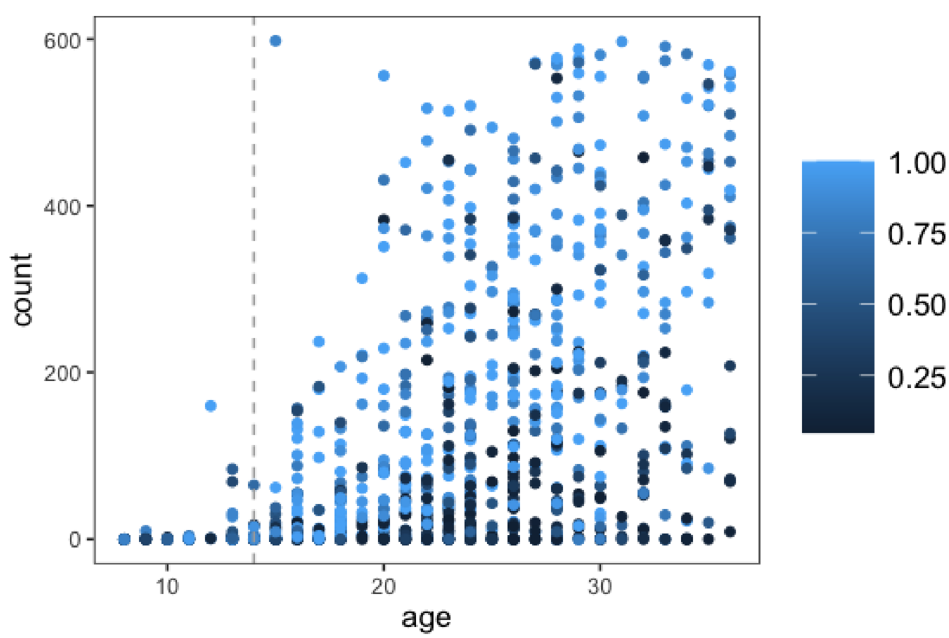


Figure 3-9: *Scatter plot of production scores across month-old groups, colour coded with exposure (0 to 1, i.e., from 0% to 100%).*

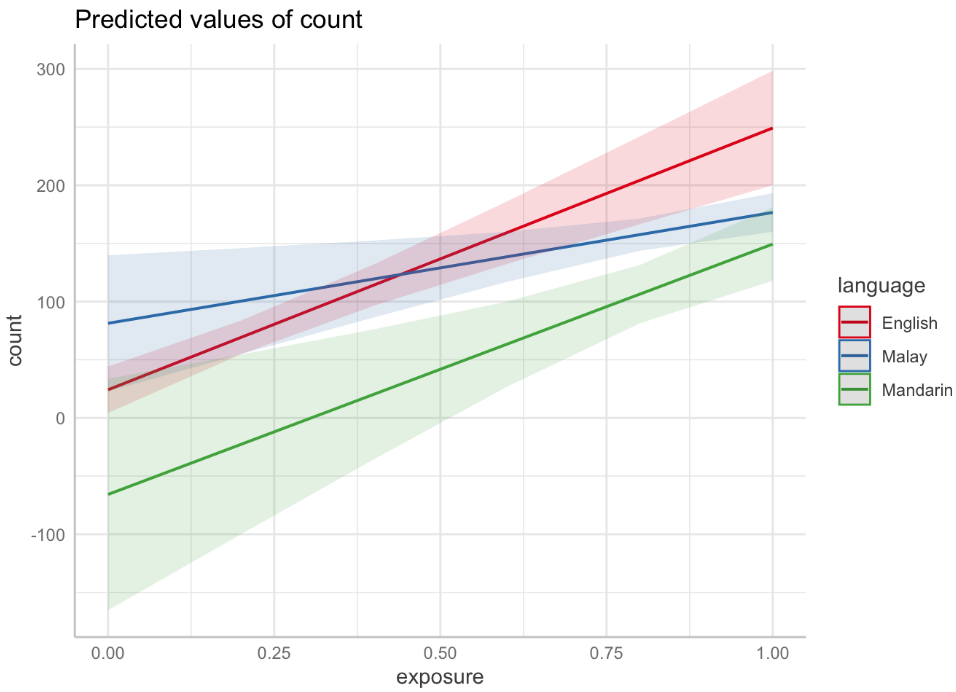


Figure 3-10: *Estimates plot for the interaction effect between proportion of language exposure and language type in predicting the production scores.*

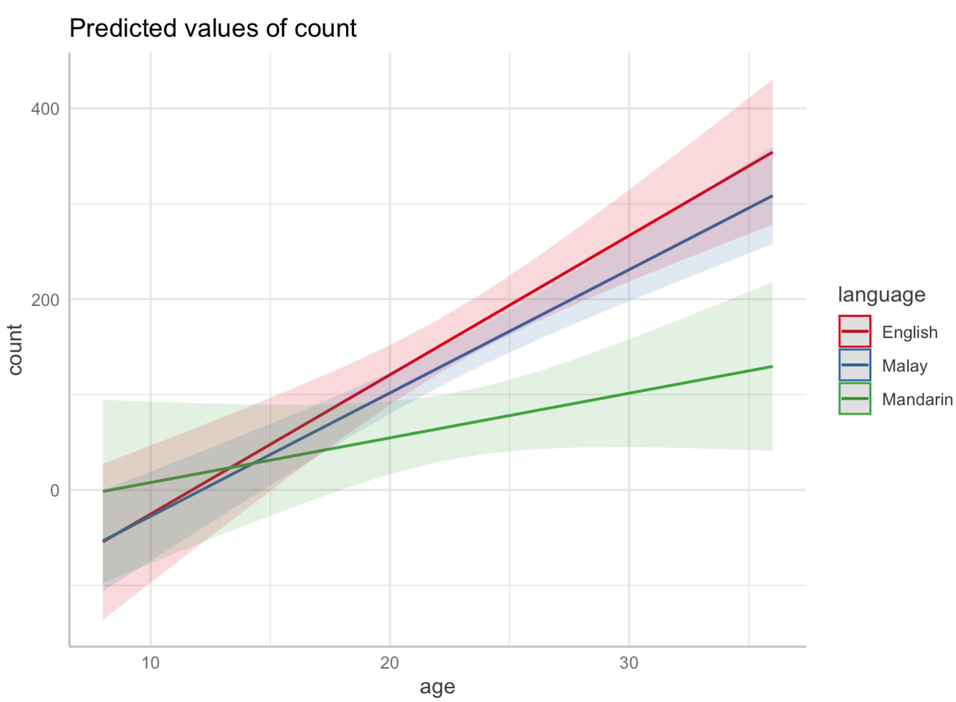


Figure 3-11: *Estimates plot for the interaction effect between language type and age in predicting the production scores.*

higher than Mandarin, $p = .012$.

3.3.3 Discussion

This study introduced early vocabulary data collected using a Malaysian-based CDI and provided insights into the early vocabulary acquisition of infants and toddlers learning Malay, or Malay and English, or Mandarin, or Mandarin and English, or English. The current study examined vocabulary growth – characterised by an increase in vocabulary sizes as age increases, capturing the effects of language exposure – characterised by a positive relationship with vocabulary sizes, and potential gender differences (i.e., girls outperform boys). The current study also examined the potential differences in language (Malay, Mandarin and English), which Malay and Mandarin belong to the Malay-ethnic and Chinese-ethnic groups. The current study utilised two means of data collection – traditional paper-based form, and online digital form, thus enabling an evaluation of any systematic differences in vocabulary assessment across sampling modes.

In terms of vocabulary growth, a positive relationship was found between age and vocabulary size – youngest children had the smallest vocabulary sizes, while older children had larger vocabulary sizes. This finding is in line with what researchers have previously found (e.g., Fenson et al., 1994; Feldman et al., 2000; Garcia et al., 2014). Language

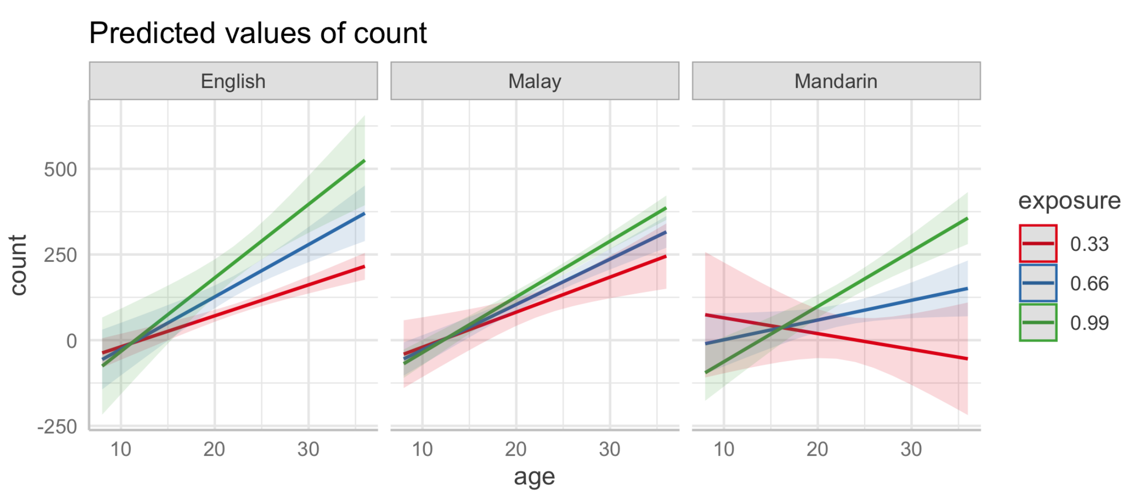


Figure 3-12: *Estimates plot for the interaction between proportion of age, language exposure and language type in predicting the production scores.*

exposure was found to be a significant factor in predicting the vocabulary size for both vocabulary production and comprehension. The findings showed that a typically developing child with higher language exposure outperformed other children who had less exposure to a language. This reflects the linguistically diverse context of Malaysian society, in which the presence of a dominant language (i.e. higher exposure) facilitated the acquisition of that language to a larger extent than a secondary language. More, this demonstrated the effectiveness of CDI in capturing the modulation of language exposure on children's vocabulary sizes. Importantly, the findings showed an interaction between language exposure and age in modulating early vocabulary acquisition — while vocabulary sizes increased with age, children under 20-month-old tended to have similar vocabulary sizes across exposure groups, and then the vocabulary gaps between exposure groups widened as age increased, in both comprehension and production. At an earlier stage of vocabulary acquisition, children's knowledge of words is limited and the effect of exposure is less pronounced. As experience accumulates (after 20-month-old), toddlers with more exposure acquire a larger vocabulary than toddlers with less exposure. The implication of exposure not being an important indicator of vocabulary sizes before the age of 20-month-old established a threshold for the effect of exposure and should be examined further in future work. Together, these findings provide important insights into the early vocabulary acquisition trajectory of Malaysian children.

Comparing vocabulary sizes between the genders, while we found no differences in vocabulary comprehension, the girls performed better in vocabulary production than boys. This finding in production is consistent with a collection of studies that showed a girl's advantage in language development (e.g., Fenson et al., 1994; Hyde, 1981; Eriksson, 2017; Bleses et al., 2008a). The lack of gender differences observed in comprehension could be due to the sensitivity of the study design (i.e., parent-report measure). Frank et al. (2021) found that the estimate of the standardised effect size for gender in the American English CDIs was twice as low as the estimate found in Bornstein and Putnick (2012), a longitudinal study using standardised and parent-report measures.

There have been concerns about collecting CDI data via the internet as only a subset of the population has access to the internet and that the sample might be skewed towards families with higher SES (Kristoffersen et al., 2013). Given that the current study involved two samples, one recruited physically and another via the internet, we were able to examine whether there were any sampling differences in vocabulary sizes. A comparison of SES was not possible due to a lack of information regarding SES [7] in the paper form. Besides, another major difference between sampling methods was the language exposure – the online sample contained more bilinguals than the paper sample (cf Table 3.2). Toddlers recruited via the internet were reported to be more advanced in comprehension than toddlers who were recruited physically. In production, however, a more complex case was found – the vocabulary size in English was larger than in Malay and Mandarin in the online sample, whereas English and Malay were larger than Mandarin in the paper sample. Together, these findings suggest that, not only were toddlers recruited via the internet more advanced in vocabulary acquisition, they tended to be more advanced in a particular language, English. The overall advantage in English vocabulary in the online sample is related to the exposure group – the online sample had 5 times as many children who were exposed to English dominantly (above 61% exposure), as the paper sample did (cf Table 3.2). It is unclear whether the overall differences in vocabulary sizes between online and paper samples were the result of an overall sampling effect, or were caused by SES. As previously mentioned, while it would be ideal to compare SES between the sampling types, this is not possible in the current study. We anticipate that a comparison within the online sample in the next study – Study 2 may shed some light on this by examining whether SES (in terms of income and educational level) impacts the pattern of language exposure in the online sample and its effects on vocabulary sizes.

In sum, we showed the effectiveness of the Malaysian CDI in demonstrating common attributes of the CDIs, that is, capturing the growth of vocabulary sizes over age among infants and toddlers. Unlike many other CDIs, the Malaysian version also included a language exposure inventory and vocabulary in three languages, which enables the investigation of modulation of exposure on vocabulary acquisition. The traditional monolingual CDI forms that are widely used in many countries are not suitable for bilingual or multilingual Malaysians. Speech-language practitioners and therapists should take note of Malaysian

[7] Nonetheless the SES of the online-based sample will be examined in Study 2.

infants and toddlers' language learning experience before comparing them with their peers to avoid misestimation of language competence. As indicated by the current findings, poor vocabulary knowledge in one language is not a definite indication of language delay, but might be due to low exposure to that language and higher exposure to other language(s). This imbalance of language exposure resulted in more knowledge of one language than the other. It is noteworthy that improvement is needed for the MCDI-M to become more inclusive, as it lacks vocabulary items in the language of Indian origin, especially Tamil – the most widely spoken Indian language in Malaysia.

To conclude, the MCDI-M is capable of capturing developmental milestones of early language acquisition and is sensitive to differences in language exposures. The following chapter will report further analyses with a focus on the socio-economic status effect that was often observed in other literature.

3.4 Study 2: Socio-economic status and Early Vocabulary Acquisition

3.4.1 Purpose

The current study examined the effect of socio-economic status (SES) on parental language use and vocabulary acquisition using data from the online version of MCDI-M. Specifically, two components of SES – parental education level and household income level – were examined.

We hypothesised that education attainment, but not income level, had an effect on language preference. As mentioned in Section 2.3.5., parents in the current sample received mathematics and science education in English during their primary and secondary school years and learned other subjects in their own ethnic languages (EL) respectively. For parents who had a university education (taught in English), they would have more experience in the use of the English language, as opposed to parents who did not continue their education after secondary/primary school (primarily taught in EL). Hence, we expected a higher relative exposure to English for children from families with better education than for children from less educated families. As a result, we expected that bilingually-exposed children from less educated families would acquire more EL vocabulary than English vocabulary when compared to children from university-graduated families.

In parallel, as prefaced in Section 2.3.4.2, there are a limited number of studies that relate income and vocabulary in the context of bilingualism, with most studies focusing on English vocabulary in an English-dominant society. The current study thus aimed to examine the effects of income on both languages in bilingually-exposed children in a non-English-dominant society. We expect that children from high-income families will acquire larger vocabulary than children from low-income families, due to better access to learning resources (Davis-Kean, 2005; Rowe, 2008).

In terms of interaction between income and parental education level, while we expect children from less educated families to acquire a larger receptive and productive vocabulary in their EL than children from more educated families, we expect children from less educated, high-income families to acquire the largest vocabulary sizes in EL. Whereas in English, while we expect children from better educated families to acquire more vocabulary in English than children from less educated families, we expect children from better educated, high-income families to acquire the largest vocabulary in English.

3.4.2 Classification of Income Levels

The sample was grouped into three household income groups – the low-income group included incomes of RM4,000 and below; the middle-income group with an income between RM4,001 and RM8,500 and a high-income group with an income of RM8,500 and above. These groups followed the classification of the Department of Statistics Malaysia (2017)'s definition of the bottom 40% (B40), middle 40% (M40) and top 20% (T20) income groups in 2016, which reported a median household income of RM13,148 for T20, RM6,275 for M40 and RM3,000 for B4. The income distribution of the current sample, shown in Table 3.6, was similar to the Department of Statistics Malaysia (2017)'s classification, with a composition of 42% (cf. B40) of the low-income group, 33% (cf. M40) of the middle-income group and 25% (cf. T20) of the high-income group. Income distributions varied

Table 3.6. Gender distribution of children and socio-economic status of their families between language exposure group.

	Malay-English bilingual, N = 113 N (%)	Mandarin-English bilingual, N = 42 N (%)	Total N = 155 N (%)
Gender			
Females	49 (43)	25 (60)	74 (48)
Males	64 (57)	17 (40)	81 (52)
Household income			
High-income	17 (15)	22 (52)	39 (25)
Middle-income	36 (32)	15 (36)	51 (33)
Low-income	60 (53)	5 (12)	65 (42)
Highest education			
Pre-university	26 (23)	6 (14)	32 (21)
University (Undergraduate)	66 (58)	29 (69)	95 (61)
University (Postgraduate)	21 (19)	7 (17)	28 (18)

between ethnic groups, with a greater proportion of Malay-English bilinguals belonging to the low-income group than the high-income group, in comparison to Chinese-English bilinguals, who were more likely to be in the high-income group than the low-income group.

3.4.3 Classification of Education Levels

The educational attainment of both parents was compared and the highest educational level was used as the indicator of the household educational level, as reported in Table 3.6. A pre-university group was defined as having the highest educational attainment of either secondary school, college or polytechnics; a university (undergraduate) group was defined as having an undergraduate degree from a university as their highest educational attainment; and a university (postgraduate) group was defined as having a postgraduate (Master or PhD) degree from a university as their highest educational attainment. With this classification, a large proportion of families were in the undergraduate group – more than 50% of families had an undergraduate degree as their highest educational attainment, while the remaining families were evenly distributed between pre-university and postgraduate levels.

Using the *eDataBank*, an electronic data bank of the Department of Statistics Malaysia (DOSM), we extracted the census by the highest level of education attained by the Malaysian population aged between 20 and 55 years of age ($13,272,936$ individuals; Department of Statistics Malaysia, 2010)[8]. The Department of Statistics Malaysia (2010) reported that 2.4% of the sampled population had the highest level of education at a primary or lower secondary school level; with 48.8% at an upper secondary, pre-university, post-secondary non-tertiary education and first stage tertiary education at certificate/diploma level, with 8.4% at first stage tertiary education at degree/advanced diploma/master level; with the remaining 22.4% as unknown. When compared with the sampled population from the data bank, not surprisingly, our sample is biased towards a population with higher educational attainment – the proportion of undergraduates in our sample (see Table 3.6) is larger by a factor of 7.26 relative to the population. Yet, it is noteworthy that the population data (Department of Statistics Malaysia, 2010) is outdated as it was released a decade ago, without any new data available since (meanwhile, the MyCensus 2020 (R. Lim, 2021) is still in data collection progress).

In contrast, partial statistics collected in 2019 (Ministry of Education Malaysia, 2020) reported 17.6% of the sampled population aged 25 and above (sample size is unknown) had the highest level of education at primary level, 51.5% at secondary level, 25.3% at tertiary level, and the rest are those without formal education. Again, an over-representation of the population with a higher education was observed – the proportion of tertiary level (undergraduate and postgraduate level combined) in our sample is larger by a factor of 3.12 relative to the Ministry of Education Malaysia (2020)'s sample, as well as an under-representation of the population with a lower education – the proportion of pre-university level (primary and secondary level combined) in the Ministry of Education Malaysia (2020)'s sample is larger by a factor of 3.29 relative to our sample.

Together, these findings suggest that the current sample is biased in terms of education, when compared to the population in Malaysia. The over-representation of parents with a higher education and the under-representation of parents with a lower education have

[8] New data collection from the MyCensus 2020 initiative is still ongoing and is delayed due to the COVID-19 Pandemic (R. Lim, 2021).

generally been found in other CDI studies (Bleses et al., 2008a; Fenson et al., 2007, 2000; Simonsen et al., 2014) and are critical when evaluating children's performance, particularly when applying and interpreting scores for children from lower SES families (see Section 2.2.1 and 2.2.3 for the relationship between parental education and children's vocabulary sizes).

The gender distribution of both groups of children and the socio-economic status of their families are reported in Table 3.6.

3.4.4 Results

Table 3.7 presents descriptive statistics of language exposure for the language type (ethnic language and English) across parents' highest educational level (pre-university, undergraduate, and postgraduate) and combined household income level (low, middle- and high-income). In general, although the average proportion of English exposure was lower than 50%, the standard deviation of EL and English exposure were high, signalling that the language preference of the parents was diverse. Figure 3-13 depicts the scatter plot of vocabulary counts in comprehension and production across age and are colour-coded for the proportion of language exposure. It is noteworthy that the excessive zeros in the comprehension and production subplots represent bilingually-exposed children with close to zero exposure to one language, due to a near maximum proportion of exposure to another language. Given the variation of vocabulary sizes across age, age will be entered as a random-effect predictor in subsequent LMM fitted for comprehension and production to control for its effects.

3.4.4.1 SES Effect on Early Language Environment

A linear model was fitted to examine the effects of the highest educational level, combined household income level, language type and their interactions in predicting language exposure. A Type III ANOVA was conducted to analyse the fitted model, as reported in Table 3.8. Pairwise comparisons were conducted for significant interaction effects between education and language type, and between income and language type.

The pairwise comparison analysis for the interaction effect between education and language type revealed no significant differences between educational levels in English exposure, *ps* > .20 and in EL exposure, *ps* > .21 (see Figure 3-14 for the visualisation of the emmeans). This indicated that parents with different educational attainment tended to use a similar amount of EL, as well as English, with their children. When examining language differences within each educational level, no significant differences were found between

Table 3.7. Language exposure average (standard deviation in parentheses) for ethnic and English language across income and education.

SES	EL exposure, %	English exposure, %
Household income		
High-income	51 (27)	49 (27)
Middle-income	79 (18)	21 (17)
Low-income	76 (16)	24 (16)
Highest education		
University (Postgraduate)	61 (28)	39 (28)
University (Undergraduate)	73 (22)	27 (22)
Pre-university	73 (18)	27 (18)

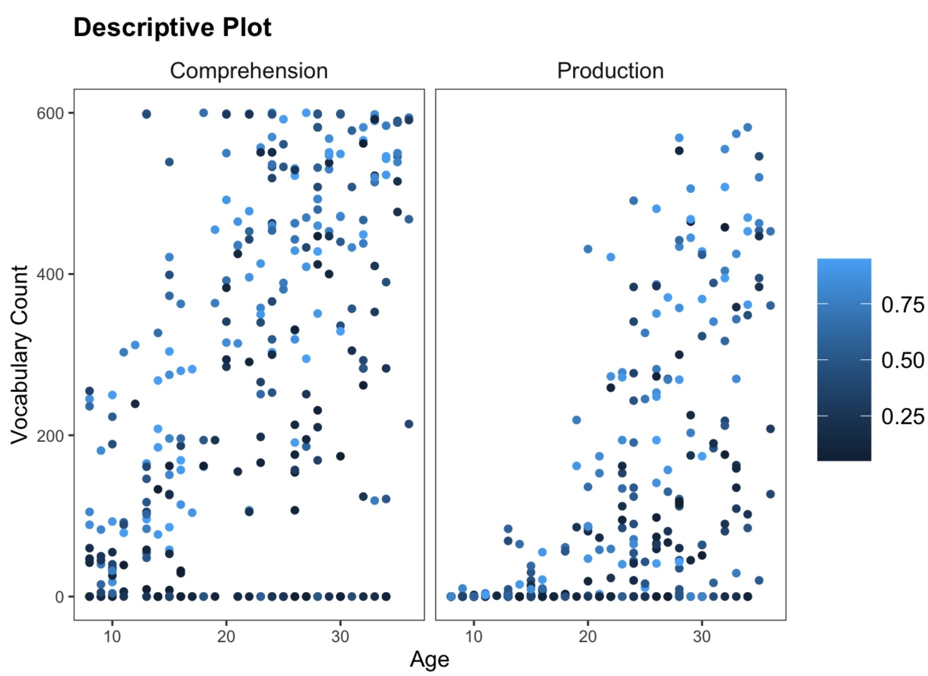


Figure 3-13: *Descriptive scatter plot of vocabulary in comprehension and production with language exposure colour-coded, across age.*

English and EL exposure in the pre-university group, $p = .086$. The English exposure was significantly lower than the EL exposure in the undergraduate group, $p < .001$. Lastly, the English exposure was also significantly lower than EL exposure in the postgraduate group, $p < .001$. Parents with undergraduate or postgraduate education tended to use more EL than English, except for parents with pre-university education, who tended to use a similar amount of both languages with their children.

Table 3.8. Type III ANOVA table for the effects of income, education and language type on vocabulary size in language exposure.

Predictors	η_p^2	**Df**	***F***	***p***
Education	.00	2	.00	1.000
Income	.00	2	.00	.999
Language	.54	1	349.75	< .001
Education:Income	.00	4	.01	1.000
Education:Language	.06	2	4.43	.013
Income:Language	.25	2	49.29	< .001
Education:Income:Language	.03	4	2.25	.064

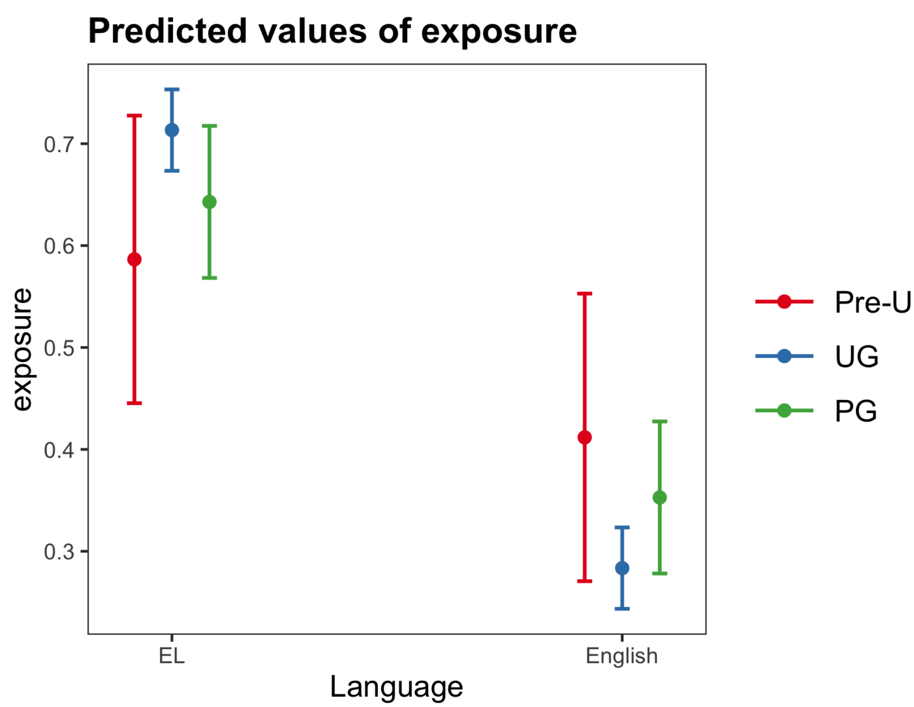


Figure 3-14: *Emmeans plot of language exposure for EL and English, across educational levels (Pre-U: pre-university, UG: undergraduate, PG: postgraduate).*

The pairwise comparison analysis for the interaction effect between income and language type revealed that, for English exposure, the high-income group was significantly higher than the low-income group and the middle-income group (*p*s $< .001$), whereas for EL exposure, the high-income group was significantly lower than both the other income groups (*p*s $< .001$). No significant differences were found between the middle and low-income groups, in English exposure, $p = .998$ and in EL exposure, $p = .992$ (see Figure 3-15 for the visualisation of the emmeans). This indicated that high-income parents tended to use more English with their children than parents with lower income and thus tended to use less EL than lower income groups. When examining language differences within each income level, children from the high-income group tended to receive a similar amount of exposure from both languages, as no significant differences between EL and English exposure were found ($p = .073$) whereas children from the middle and low-income groups tended to receive EL-dominant exposure, as EL exposures were significantly higher than English exposures (*p*s $< .001$).

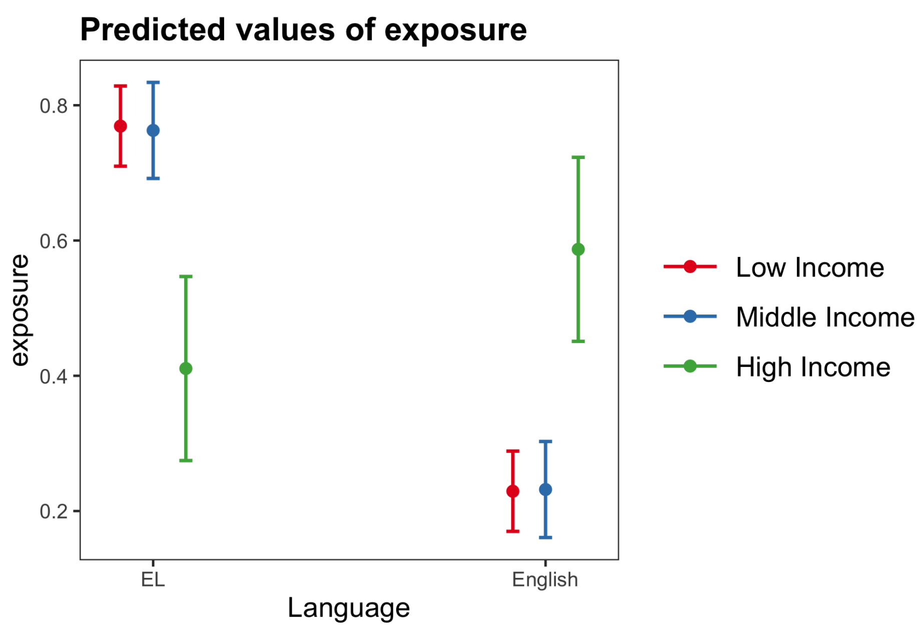


Figure 3-15: *Emmeans plot of language exposure for EL and English, across income levels.*

3.4.4.2 SES Effect on Vocabulary in Comprehension

A LMM was fitted to examine the effects of the highest educational level, combined household income level, language type and their interactions in predicting children's vocabulary comprehension. It is noteworthy that age and gender were controlled as random-effect factors. Table 3.9 reports the results of the Type III ANOVA analysis. Pairwise comparisons were conducted for the significant interaction effects between income and language type.

Table 3.9. Type III ANOVA table for the effects of income, education and language type on vocabulary size in comprehension. Age and gender were controlled as random-effect factors.

Predictors	χ^2	**Df**	p
Education	.58	2	.748
Income	.07	2	.967
Language	2.99	1	< .001
Education:Income	2.90	4	.574
Education:Language	3.80	2	.150
Income:Language	26.57	2	< .001
Education:Income:Language	1.62	4	.807

The pairwise comparison analysis for the interaction effect between income and language type revealed no significant differences in EL comprehension across income groups, all ps > .10, thus suggesting no vocabulary differences in EL comprehension due to income level (see Figure 3-16 for the visualisation of the emmeans). However, in English, the vocabulary size in the high-income group was significantly higher than the middle-income group, p = .005, and the low-income group, p = .007, with no differences between the low and middle-income group, p = .95. The findings suggest that vocabulary differences exist in English but not in EL across income groups. Specifically, children from high-income families were more advanced in English than their lower income peers but had similar EL vocabulary sizes. In addition, English comprehension was significantly lower than EL comprehension for the low and middle-income groups, ps < .001. For the high-income group, English comprehension was significantly higher than EL comprehension, p = .028. Children from the high-income group were more advanced in English word learning than in EL

(p = .028) whereas children from the low and middle-income groups were more advanced in EL than in English (ps < .001).

3.4.4.3 SES Effect on Vocabulary in Production

A LMM was fitted to examine the effects of the highest educational level, combined household income level, language type and their interactions in predicting children's vocabulary production. It is noteworthy that age and gender were controlled as random-effect factors. Table 3.10 reports the results of the Type III ANOVA analysis. Pairwise comparisons were conducted for the significant interaction effects between income and language type.

The pairwise comparison analysis for the interaction effect between income and language type revealed no significant differences in EL production across income groups, all ps > .414, thus suggesting no vocabulary differences in EL production due to income level (see Figure 3-17 for the visualisation of the emmeans). In contrast, English production in the high-income group was significantly higher than in the middle and low-income groups, ps < .001, with no differences between the low- and middle-income groups, p = .640. The findings suggest that, for production, children from high-income families were reported to

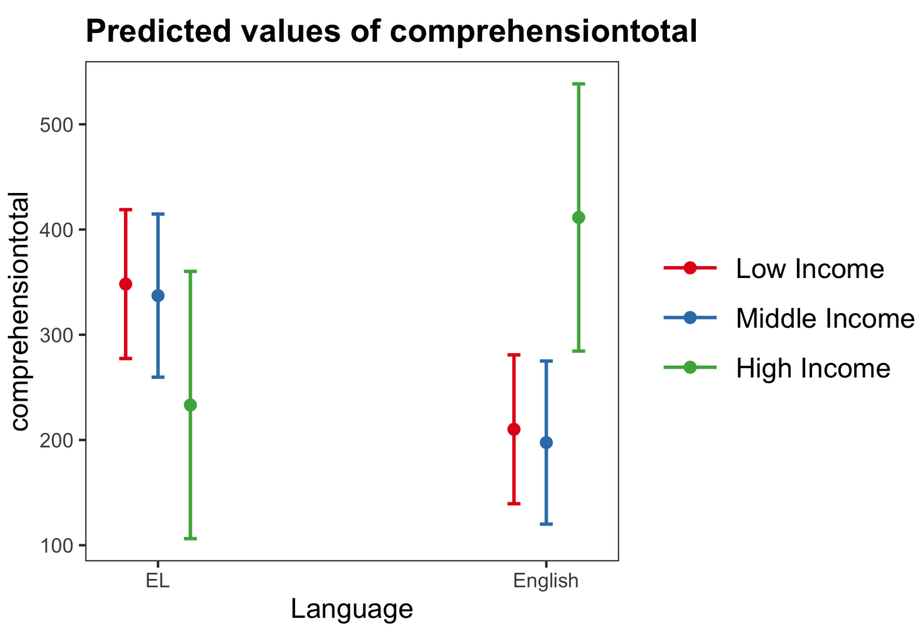


Figure 3-16: *Emmeans plot of vocabulary in comprehension for EL and English, across income levels.*

be able to produce more English vocabulary than the lower income groups but had similar EL vocabulary sizes as other income groups. In parallel, no significant differences between English and EL production were found for both the low and middle-income groups, *p*s > .055. For the high-income group, English production was significantly higher than EL production, $p < .001$. Children from high-income families were more advanced in English word learning than in EL whereas children from low and middle-income groups had similar

Table 3.10. Type III ANOVA table for the effects of income, education and language type on vocabulary size in production. Age and gender were controlled as random-effect factors.

Predictors	χ^2	**Df**	***p***
Education	1.03	2	.599
Income	9.61	2	.008
Language	1.31	1	.252
Education:Income	4.18	4	.383
Education:Language	5.65	2	.059
Income:Language	18.58	2	< .001
Education:Income:Language	6.49	4	.166

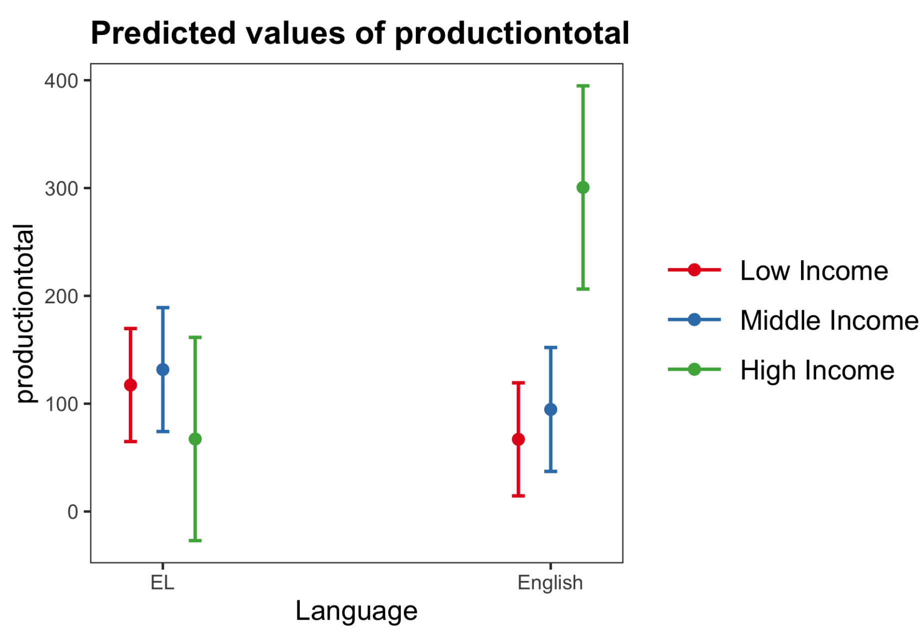


Figure 3-17: *Emmeans plot of vocabulary in production for EL and English, across income levels.*

levels of knowledge in English and EL.

3.4.4.4 Language Exposure as a Mediating Factor

We followed Baron and Kenny (1986)'s approach – to assume a three-variable system which comprised two causal paths flowing into the outcome variable: the independent variable's direct influence and the mediator's indirect impact on the outcome variable; and an additional path from the independent variable to the mediator – to investigate the role of language exposure as a mediating factor between SES (education and income) and vocabulary across languages (EL and English). To this end, we need to demonstrate the following assumptions by conducting a series of linear regression analyses:

1. A significant relationship between independent variables (SES and language type) and mediator variable (language exposure) is needed to show that the SES was indeed related to differences in language exposures (this model is the same as the previously fitted exposure model).

2. A significant relationship between independent variables (SES and language type) and outcome variable (vocabulary size) is needed to show that the SES was indeed related to differences in vocabulary sizes (this model is the same as the previously fitted comprehension/production model).

3. A significant relationship between mediator variable (language exposure) and outcome variable (vocabulary size) is needed to show that the language exposure was indeed related to differences in vocabulary sizes.

4. The final test requires a reduction in the significant relationship modelled in (2) to non-significance upon inclusion of the significant mediator variable (language exposure) in the model, to establish whether language exposure mediates the relationship between SES and vocabulary size.

In comprehension, for (1), a significant relationship between independent variables and the mediator variable was previously found, as well as for (2), a significant relationship between independent variables and the outcome variable was also previously found. For (3), a significant relationship between the mediator variable and the outcome variable was found ($b = 260.71$, $p < .001$), in which the exposure was positively related to the vocabulary sizes. For the final test, the previously significant relationship between independent variables and outcome variable was reduced to non-significance (ps > .060) upon the inclusion of the mediator exposure in the model. This indicated that language exposure indeed mediated the relationship between SES and vocabulary size.

In production, for (1), a significant relationship between independent variables and the mediator variable was previously found, as well as for (2), a significant relationship between independent variables and the outcome variable was also previously found. For (3), a significant relationship between the mediator variable and the outcome variable was found ($b = 133.27$, $p < .001$), in which the exposure was positively related to the vocabulary sizes. For the final test, most of the previously significant relationships between independent variables and outcome variable was reduced to non-significance (ps > .36) upon the inclusion of the mediator exposure in the model, except for the significant main effects of income ($p = .006$) and of language ($p < .001$). This indicated that language exposure partially mediated the relationship between SES and vocabulary size.

3.4.5 Discussion

Numerous studies have demonstrated the effect of various components of SES on children's vocabulary knowledge, for example, the effect of education: Hart and Risley (1995); Hoff et al. (2018); Wanless et al. (2011); Rowe (2008); and the effect of income: Arriaga et al. (1998); Layzer and Price (2008); Blanden and Machin (2010); Taylor et al. (2013). This study adds to the research by examining the effect of SES on both the languages of a child from bilingual families. The current study provides some insight into the language preferences of parents from different SES groups when talking to their children, as well as the ways in which bilingual children from different SES groups differ in their receptive and expressive vocabulary in both of their languages. Finally, the current study examined whether parents' language preferences mediate the relationship between SES and vocabulary in a non-English dominant society, Malaysia.

In terms of educational attainment, a number of previous studies found an influence of parental education on children's vocabulary knowledge (Hart & Risley, 1995; Hoff et al., 2018; Rowe, 2008; Wanless et al., 2011). Yet, the current study suggests that parental education seems to have no influence over children's vocabulary size, in both comprehension and production. The lack of influence of parental education has been previously reported in a few monolingual studies (Kartushina et al., 2021; Serrat-Sellabona et al., 2021; Frank et al., 2021). With respect to bilingually-exposed children, this finding is partially consistent with the study conducted by Sun et al. (2020), in which results indicated no effect of maternal education on bilingually-exposed children's receptive vocabulary in their heritage language, but they reported a difference in the English vocabulary size of their participants owing to the English-focused education policy in Singapore, whereas the current study in Malaysia found no effect of education on either the children's EL nor their English vocabulary sizes. This suggests a more nuanced role of parents' education in early vocabulary acquisition, in the context of Malaysian bilingually-exposed children. We suggest future research should consider parents' education in terms of language proficiency level and their effect on children's language development.

In terms of language preference, parents across all three educational levels preferred the use of the EL over the English language. This is speculated as a parent's effort in upholding their ethnic language and protecting their ethnic identity, despite their differences in educational attainment. As an indirect support, in an interview study, S. K. Lee, Lee, Wong, and Ya'acob (2010) found that despite Malaysian undergraduate students' proficiency in English and their awareness of the advantages of English, they still experienced pressure from friends who were less experienced with English, or who perceived them as "distancing" (S. K. Lee et al., 2010), and thus preferred the use of EL to foster closeness with members of their own ethnic group. While the sample in this a study was restricted to university students, it might provide insight on parents' language use and their community's language, in general. Considering that, future research could examine the relationship between parents' perceptions of the languages, their language choice in their immediate community (e.g., relatives, daycare, friends, places frequented with their children) and their mediating effects on parents' language preference at home, thus providing a sociocultural account for the lack of educational effect (on both the language exposure and vocabulary sizes) in the current study.

Yet, children from families of different income levels differed in their bilingual exposure status – children from high-income families tended to be exposed to a similar proportion of both languages, whereas children from middle and low-income families tended to be dominantly exposed to EL. In other words, parents with high-income preferred to balance the use of English and EL with their children, whereas parents with a middle or low-income favoured the use of EL over English. In addition, the findings on the relationship between family income and children's vocabulary paralleled parents' higher preference for the use of the English language. Specifically, children from high-income families acquired an English vocabulary size twice larger in comprehension ($M_{HighIncome}$ = 411, $M_{MiddleIncome}$ = 198, $M_{LowIncome}$ = 210), and around 3-5 times larger in production ($M_{HighIncome}$ = 300, $M_{MiddleIncome}$ = 95, $M_{LowIncome}$ = 67), compared to children from low- and middle-income families, with no differences between low- and middle-income families. Indeed, the mediation analysis confirmed the mediating effect of English exposure on the relationship between income level and English vocabulary in comprehension. This showed that higher income families were willing to invest more time in using the English language with their children, which resulted in better English comprehension of their children than lower income families. Another speculation is that parents of higher income are more likely to use the English language in daily life due to their English proficiency – a metric that we did not measure in the current study. It is noteworthy that exposure to a particular language provides opportunities to practise real-time comprehension in that language, which facilitates the development of processing skills that are crucial for learning (Hurtado et al., 2014).

Nonetheless, the mediation analysis revealed only a partial mediating effect of exposure between SES and vocabulary in production. This suggests that, beyond parental exposure, there are other mediating factors that were not included in the model, e.g., exposures from other sources such as grandparents. In the USA, grandparental care is often triggered when the parents are unavailable to take care of their children, whereas in Asia, grandparental care is a norm (Baker, Silverstein, Arber, & Timonen, 2012; Bushneil, Sai, & Mullin, 1989; Croll, 2006). We postulated that a similar caregiver pattern could be observed in

Malaysia, an Asian country with a collectivist culture. As such, it is possible that children's language skills are influenced in a greater extent by interactions with other family members, such as grandparents and relatives. Besides, access to media content (e.g., TV shows, nursery rhymes, and cartoons) could play a role in shaping the language environment, too. Future research could examine a wider range of sources that contribute to the composition of language exposure. In addition, while the current study did provide an index of the quantity aspect (e.g., relative parental language exposure), the quality of the input was not measured. The quality of a speech input often involves three factors – the amount of exposure the children received from proficient speakers of a language, the richness of home learning activities, and the presence of older siblings (Rojas et al., 2016; Sun, Steinkrauss, Wieling, & De Bot, 2018). Crucially, the richness of home learning activity (in terms of access to literacy resources at home and home reading behaviours) was found to relate to income levels (Zuilkowski, McCoy, Jonason, & Dowd, 2019). Together, research could examine the potential mediation of both language quantity and quality between income and production.

In addition, similar EL vocabulary sizes were found in children from all three income groups. While this finding seems to suggest that income does not translate into an advantage over learning the EL, it is noteworthy that children from the high-income group had a reduced exposure to EL compared to middle- and low-income groups. Despite that, children from the high-income group managed to learn EL words at a similar pace as children from low- and middle-income groups, while also outpacing them in the learning of English. This finding on EL does not replicate previous studies that found children from high-SES, bilingual families had poorer EL language skills (Hoff et al., 2012); and that lower exposure leads to poorer vocabulary (Armon-Lotem & Ohana, 2016; David & Wei, 2008; Parra et al., 2011; Pearson et al., 1997; Thordardottir, 2011). Importantly, the overall linguistic contexts in the current study are distinct from the mentioned studies, which are attributed to differences with the current findings. In the current study, the EL of Malaysian children are supported in the school and community, whereas the ELs in the mentioned studies were most often used at home. In addition, we attributed the lack of an effect of income (on EL learning) to the limitation of language input measurement in the current study. We suggest that the other sources of language input (e.g., friends, relatives, grandparents) should be evaluated in greater detail, as they may mediate the effect of income level on the learning of EL. More, even overheard speech was found to facilitate acquisition of new words (Akhtar, Jipson, & Callanan, 2001).

It is noteworthy that, in this study, we relied on parental reports – while parents are adequate informers about their children's vocabulary using CDI (see sections 2.2.2 and 2.5.2.1) and about language exposure (Orena, Byers-Heinlein, & Polka, 2019). Still, conducting direct measures of language skills would shed more light on the SES effect, supplementing findings from indirect measures such as this study. One avenue is to compare the vocabulary data collected using CDI with a CDI-based word comprehension test using an application (C. H. Lo, Mani, Kartushina, Mayor, & Hermes, 2021) that was developed for assessing toddlers. Study 6 examines the viability of using such a tool (C. H. Lo et al., 2021) in early vocabulary assessment.

The findings from the current study show that parents' continued use of EL, which is interpreted as an effort to uphold their ethnic identity, is effective in fostering the EL acquisition of their children. Even though high-income parents used English more often than parents with lower income, they balanced language use between English and EL, with the observation that it did not compromise EL learning. In parallel, children from lower income families are lagging behind children from high-income families in terms of English language development due to reduced language exposure to English. The next section will look into language use and early vocabulary acquisition from a different perspective – examining the potential role of sociocultural influences on acquisition patterns, in particular on the noun bias.

3.5 Study 3: Extra-Linguistic Modulation of the Verb-Noun-Bias

3.5.1 Purpose

The aim of this study is to examine the role of language and sociocultural influences on language acquisition patterns, specifically on the noun bias.

We focus on the English learnt by two groups of bilingual children from two distinct sociocultural environments – raised in Chinese-ethnic families and in Malay-ethnic families in Malaysia. Malay is an Austronesian language with a morphological complexity between that of Mandarin and English – it relies on both noun and verb inflections of the word stem to produce complex meanings, but unlike nouns, which can be used in their bare stem forms, most verbs have to be inflected to denote actions (Tadmor, 2009). Yet, similar to Mandarin, Malay does not inflect verbs for tenses nor nouns for marking the singular-plural distinction, but relies on separate markers to do so. Consequently, we expected that language-intrinsic factors would make Malay more noun-friendly than Mandarin and more verb-friendly than English (a secondary aim of the present study being to verify this hypothesis). Crucially, differences in the ratio of verbs to nouns in the English lexicon across these two groups of children, as indexed by parental reports, would thus suggest that extra-linguistic factors can modulate the English noun bias. In summary, the independent variables of this study are language types (i.e., Malay, English and Mandarin) and bilingual groups (i.e., Malay-English learners and Mandarin-English learners) whereas the dependent variable is the verb-noun ratio. We expect the English verb-to-noun ratio in Mandarin-English to be larger (i.e., more verbs) than in Malay-English learners. We expect the Malay verb-to-noun ratio to be larger than English but smaller than Mandarin.

3.5.2 Normative data from Stanford Wordbank

To assess the noun-friendliness of Malay in comparison to other languages, we extracted additional normative data from Beijing Mandarin and American English data from Stanford Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2016). The American CDI data was

Table 3.11. Amount of items selected in each word categories (common nouns and verbs) from the MCDI-M.

Word categories	
Animals	52
Vehicles	14
Toy	12
Food and Drink	42
Clothing	23
Body parts	26
Small household items	48
Furniture and rooms	27
Outside things	49
Action words	101

collected by D. J. Thal, Marchman, and Tomblin (2013); Fenson et al. (2007). The Chinese data was collected by Tardif et al. (2009). 85 monolingual Mandarin-speaking Chinese infants and toddlers (48 males and 37 females) were sampled randomly from Wordbank. Their ages ranged from 16 to 30 months of age (M = 23.25, SD = 4.25). 85 English-speaking monolingual American infants and toddlers (45 males and 40 females) were also sampled randomly from the Wordbank with ages ranging from 16 to 30 months (M = 23.07, SD = 4.07). 85 monolingual, age-matched monolingual Malay-speaking children were extracted for the comparisons.

3.5.3 Coding of Nouns and Verbs

Nouns were counted based on the definition of Caselli et al. (1995). Nouns were stringently defined; only noun categories that represent concrete objects (animal names vehicles, toys, food and drink, clothing, body parts, small household items, furniture and rooms and outside things) were included. Nominal categories (names for people, people and locations) were excluded from analyses. Verbs were counted from the action words category. Table 3.11 reports the amount of words selected using this criteria, in each category from the MCDI-M.

When performing comparisons with data from Wordbank, we identified the common set of nouns and verbs across all three language datasets (Malay, Mandarin and English), as the exact set of words in a CDI modulates the ratio of verbs to nouns. Words from the same taxonomic level were included for comparison (e.g., candy and sweets), while words from different hierarchical levels (e.g., cereals and cornflakes) were excluded from the final sample which included 78 verbs and 156 nouns.

3.5.4 Analyses

The Malaysian infants and toddlers samples constituted four language groups: a monolingual Malay language group, a monolingual Mandarin language group, a bilingually-exposed Malay-English group, and a bilingually-exposed Mandarin-English group. The verb-to-noun ratio[9] was computed for each participant (V/(N+V)), based on parental reports. To account for a potential over-representation of null V/(N+V) ratio (young participants may not have verbs in their vocabularies yet - hence a V/(N+V) ratio of zero), Zero-Inflated Negative Binomial Models (ZINBM) was built on the V/(N+V) ratio in early lexicons, using glmmTMB (Brooks et al., 2017). The V/(N+V) ratios were converted to percentages and rounded to the nearest integers to conduct the ZINBM. Zero-inflation models essentially decompose the analysis into two parts: first on the zero-inflated part and second on the distribution of non-zero values. Since the full model with age, language group and interactions did not converge, we implemented ZINBM models with age and language group as factors for the zero-inflation (younger children are expected to know less verbs than older children; children learning a noun-friendly language are expected to know less verbs than children learning a verb-friendly language) and language group as a fixed-effect factor and age as a random-effect factor for the non-zero distribution of V/(N+V) ratios. Analyses were conducted to address the following questions:

1. Is the English V/(N+V) ratio among Malay-English bilinguals different from the English V/(N+V) ratio among Mandarin-English bilinguals? A significant difference between language groups would suggest that extra-linguistic factors can modulate the

[9]The V/(N+V) will be loosely referred as a "verb-to-noun ratio" thereafter. V/(N+V) was preferred over V/N in order to include in the analyses infants whose early vocabularies may still not include any nouns yet (and thus where V/N would not be defined).

language noun bias (in our case, in English).

2. Is the Malay V/(N+V) ratio among Malay monolinguals different from that of Malay-English bilinguals? Similarly, is the Mandarin V/(N+V) of Mandarin monolinguals different from that of Mandarin-English bilinguals? We expected no significant differences since comparisons were made in the same language by the same ethnic group.

3. Is the Malay V/(N+V) ratio among Malay learners different from that of the Mandarin V/(N+V) ratio among Mandarin learners? A significant difference between language groups would suggest that the noun-friendliness of Malay is different from the noun-friendliness of Mandarin.

In addition, the analyses will establish how noun-friendly is Malay, when compared to both Mandarin (verb-friendly) and English (noun-friendly). To this end, the current dataset was supplemented with additional data from Wordbank (Frank et al., 2016), to allow for a comparison of the noun-friendliness of Mandarin, Malay and English.

3.5.5 Results

A two-step procedure was applied to analyse the data in *R* (R Core Team, 2019). First, a test of zero-inflation was run on the distributions of verb-to-nouns using the *zero.test* function from the *vcdExtra* package (Michael, 2017). The zero-inflation test was conducted to examine whether there were excessive zeros in the verb-to-nouns ratio, contributed by having a significant proportion of samples reported with zero verb items in the CDIs. In the presence of zero-inflation, ZINBM were then used to investigate differences in verb-to-noun ratios between language groups. The *distributional part* of the model assesses the effects of language groups and age when children possess at least one verb in their vocabulary, whereas the *zero-inflation* part of the model assesses the effects of language groups and age when children possess zero verb in their vocabulary. Linear mixed models (LMM) were used when distributions were not zero-inflated.

3.5.5.1 Q1 – Does the English V/(N+V) ratio among Mandarin-English bilinguals differ from the English V/(N+V) ratio of the Malay-English bilinguals?

Our first research question was to find out whether the English V/(N+V) ratio among Mandarin-English bilingually-exposed children differed from the English V/(N+V) ratio of the Malay-English bilingually-exposed children. The tests for the presence of zero-inflation were significant for both production and comprehension ($p < .001$), hence zero-inflation models were used.

Table 3.12. Estimates of the zero-inflated negative binomial model for English verb-to-noun ratio between both language groups, Malay-English bilingually-exposed children and Mandarin-English bilingually-exposed children, in comprehension (*.ref* = reference point).

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.42	.13	26.86	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.20	.08	-2.55	.011
Age	-.01	.00	-1.44	.149
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-2.87	.77	-3.72	< .001
Language group (Mandarin-English bilinguals used as ref.)	1.61	.62	-2.55	.009
Age	-.001	.02	-.08	.935

In comprehension, the zero-inflation part of the ZINBM did not reveal an effect of age (p = .94) but an effect of language group (p = .01), as reported in Table 3.12. Malay-English bilingually-exposed children were more likely than Mandarin-English children to not possess verbs in their vocabularies yet. Crucially, the *distributional* part of the model (when children possess at least one verb in their vocabulary) did not reveal an effect of age but revealed an effect of language group on the English V/(N+V) ratio (p = .01); the English of Mandarin-English bilinguals was richer in verbs than the English of Malay-English bilinguals. The significant difference between groups, in comprehension (see Figure 3-18), suggests that extra-linguistic factors can modulate the English noun bias.

In production, the ZINBM analyses revealed that both age ($p < .001$) and language group ($p = .002$) were significant predictors of excessive zeros, as reported in Table 3.13. Younger participants and Malay-English bilinguals were more likely not to produce any verbs, yet. Once vocabularies contained verbs (as revealed by the *distributional analyses*) neither language group ($p = .25$) nor age ($p = .89$) were significant predictors of the English V/(N+V) ratio. The English of Mandarin-English bilinguals was similar in the proportion of verbs with the English of Malay-English bilinguals.

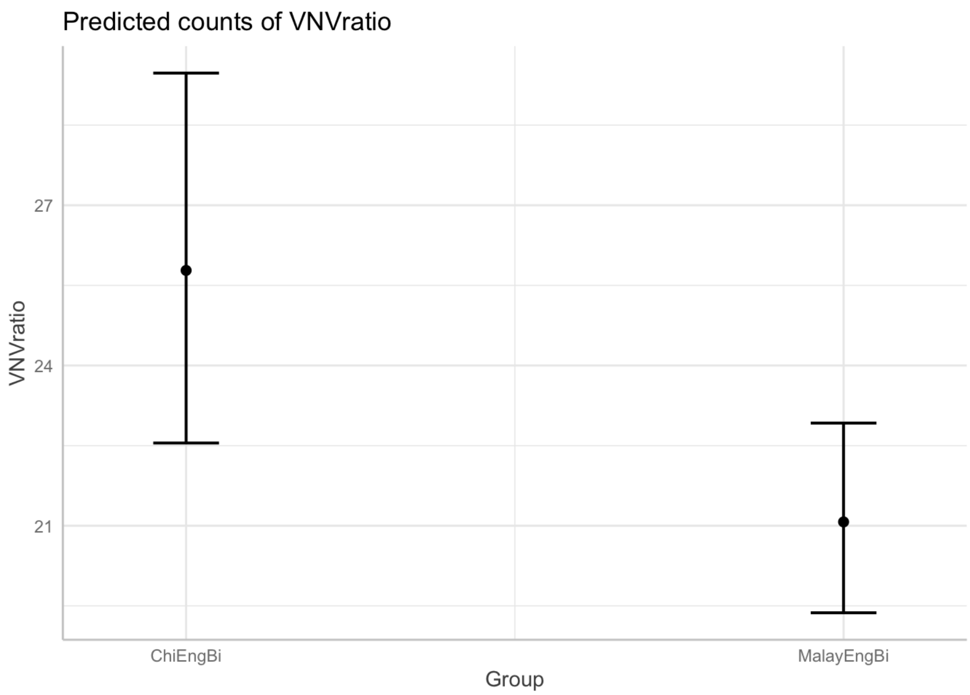


Figure 3-18: *Predicted English verb-to-noun percentage ratios in comprehension for Malay-English and Mandarin-English bilingual children using zero-inflation negative binomial model, by age and group.*

3.5.5.2 Q2 – Do verb-to-noun ratios differ between monolinguals and bilinguals?

Our second research question was whether the verb-to-noun ratios would differ between monolingually-exposed children and bilingually-exposed children. To address this question, we compared the Malay verb-to-noun ratio of Malay monolinguals to that of Malay-English bilinguals, and the Mandarin verb-to-noun ratio of Mandarin monolinguals to that of Mandarin-English bilinguals.

Table 3.13. Estimates of the zero-inflated negative binomial model for English verb-to-noun ratio between both language groups, Malay-English bilingually-exposed and Mandarin-English bilingually-exposed children, in production (*.ref* = reference point).

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.08	.24	12.71	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.13	.11	-1.15	.252
Age	-.001	.01	-.15	.885
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	.65	.75	-.87	.382
Language group (Mandarin-English bilinguals used as ref.)	1.51	.49	-3.07	.002
Age	-.11	.03	-4.04	< .001

For the Malay learners, the tests for the presence of zero-inflation were significant for both production and comprehension data ($p < .001$), hence zero-inflation models were used. In comprehension, age ($p = .003$) but not language group ($p = .823$), was a significant predictor of excessive zeros, as reported in Table 3.14. Similarly, the distributional part of the analysis revealed an effect of age ($p < .001$) but not of language group ($p = .343$). In other words, the verb-to-noun ratio of Malay, in comprehension, did not differ between monolinguals and bilinguals.

As for comprehension, in production, age ($p < .001$) but not language group ($p = .139$), was a significant predictor of excessive zeros, as reported in Table 3.15. In contrast, language group ($p = .019$) but not age ($p = .169$) was a significant predictor of the Malay verb-to-noun ratio, in production. The Malay of Malay-English bilinguals was richer in verbs than the Malay of Malay monolinguals.

For Mandarin learners, the tests for the presence of zero-inflation were highly significant for production ($p < .001$) but not for comprehension data ($p = 1$), hence a zero-inflation model was used for production data, whereas a linear model was used for comprehension

Table 3.14. Estimates of the zero-inflated negative binomial model for the Malay verb-to-noun ratio between both language groups, Malay monolinguals and Malay-English bilinguals, in comprehension (*.ref* = reference point).

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.54	.06	63.95	< .001
Language group (Malay-English bilinguals used as ref.)	-.03	.03	-.95	.343
Age	-.01	.00	-3.70	< .001
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-.18	1.03	-.17	.862
Language group (Malay-English bilinguals used as ref.)	-.18	.82	-.22	.823
Age	-.17	.06	-2.97	.003

Table 3.15. Estimates of the zero-inflated negative binomial model for the Malay verb-to-noun ratio between both language groups, Malay monolinguals and Malay-English bilinguals, in production (*.ref* = reference point).

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.15	.10	30.79	< .001
Language group (Malay-English bilinguals used as ref.)	-.12	.05	-2.35	.019
Age	.01	.004	1.38	.169
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	1.65	.71	2.31	.021
Language group (Malay-English bilinguals used as ref.)	-.58	.39	-1.48	.139
Age	-.15	.03	-4.61	< .001

data. In comprehension, age (p = .004) but not language group (p = .902), was a significant predictor of the Mandarin verb-to-noun ratio, as reported in Table 3.16. The verb-friendliness of Mandarin, in comprehension, did not differ between monolingually-exposed

Table 3.16. Estimates of the linear model for the Mandarin verb-to-noun ratio between both language groups, Mandarin monolinguals and Mandarin-English bilinguals, in comprehension (*.ref* = reference point).

	Estimate	*SE*	*t*	*p*
Intercept	45.79	4.47	10.23	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.38	3.05	-.12	.902
Age	-.54	.18	-2.99	.004

Table 3.17. Estimates of the zero-inflated negative binomial model for the Mandarin verb-to-noun ratio between both language groups, Mandarin monolinguals and Mandarin-English bilinguals, in production (*.ref* = reference point).

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.62	.26	14.04	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.24	.13	-1.87	.061
Age	-.01	.01	-.58	.563
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	1.63	2.31	.71	.481
Language group (Mandarin-English bilinguals used as ref.)	.87	.90	.96	.337
Age	-.18	.10	-1.77	.077

and bilingually-exposed children.

In production, neither age (p = .077) nor language group (p = .337), were significant predictors of excessive zeros, as reported in Table 3.17. Similarly, neither language group (p = .061) nor age (p = .563) predicted the Mandarin verb-to-noun ratio. As for comprehension, the verb-friendliness of Mandarin, in comprehension, did not differ between monolingually-exposed and bilingually-exposed children.

3.5.5.3 Q3 – How noun-friendly is Malay (in comparison to Mandarin and English)?

Our third question aimed at establishing the noun-friendliness of Malay, first in comparison to Mandarin, using the present sample of Malaysian data, then in comparison with both Mandarin and English, with data retrieved from WordBank. The common set of verbs and nouns was used when comparing the Malaysian sample, collected with the MCDI-M, with data retrieved from WordBank.

First, we compared the Malay verb-to-noun ratio of Malay monolingually-exposed children to that of the Mandarin of Mandarin monolingually-exposed children, in our sample. Due to its more complex morphological complexity, we expected that the Malay verb-to-noun ratio to be smaller than in Mandarin, in line with the proposal that richer morphological complexity of a lexical category slows down its learning (Gentner, 1982; Tardif et al., 1997). The tests for the presence of zero-inflation were significant for both production and comprehension data ($p < .001$), hence zero-inflation models were used. In comprehension, age ($p = .043$) but not language group ($p = .997$), was a significant predictor of excessive zeros – older participants were more likely to have verbs in their vocabulary. In contrast, language group ($p = .030$) but not age ($p = .215$) was a significant predictor of the verb-to-noun ratio. The Mandarin of Mandarin monolingually-exposed children was richer in verbs than the Malay of Malay monolingually-exposed children, in comprehension.

As for production, age ($p = .017$) but not language group ($p = .321$), was a significant predictor of excessive zeros. In contrast, both language group ($p = .586$) and age ($p = .252$) were not a significant predictor of the verb-to-noun ratio. The verb-to-noun ratio of Malay among monolingually-exposed children was comparable to that of the Mandarin of Mandarin monolingually-exposed children, in production.

Further analyses were conducted to evaluate the noun friendliness of Malay relative to both Mandarin and English from Wordbank. Given that our current sample did not include any English monolingually-exposed children, we supplemented the dataset with vocabulary data extracted from Wordbank, in both Beijing Mandarin and American English. The test for the presence of zero-inflation was significant ($p < .001$); hence a zero-inflation model was used. The model revealed that age ($p < .001$) was a significant predictor of excessive zeros. Older participants were more likely to produce verbs.

The distributional part of the analysis (when vocabularies possess at least one verb) revealed an effect of age (p = .002) and that Malay verb-to-noun ratio was significantly larger than English (p = .002), but significantly lower than Mandarin (p = .002). Additional Tukey-adjusted pairwise comparison revealed that the Mandarin verb-to-noun ratios were significantly larger than English, $p < .001$ (see Figure 3-19 and 3-20). Mandarin and Malay are more verb-friendly than English, in production, while Malay was found to be less verb-friendly than Mandarin.

To examine whether Malay and Mandarin retained their relative noun-friendliness (cf. our current findings – Malay was found to be more noun-friendly than Mandarin) in bilingually-exposed children, we compared the Malay verb-to-noun ratio of Malay-English bilinguals to that of the Mandarin of Mandarin-English bilinguals. In comprehension, the model revealed that age (p = .021) but not language group (p = .996) was a significant predictor of excessive zeros. Younger bilingual children were more likely to not possess verbs in their vocabularies yet. In contrast, language group (p = .001) and age ($p < .001$) was a significant

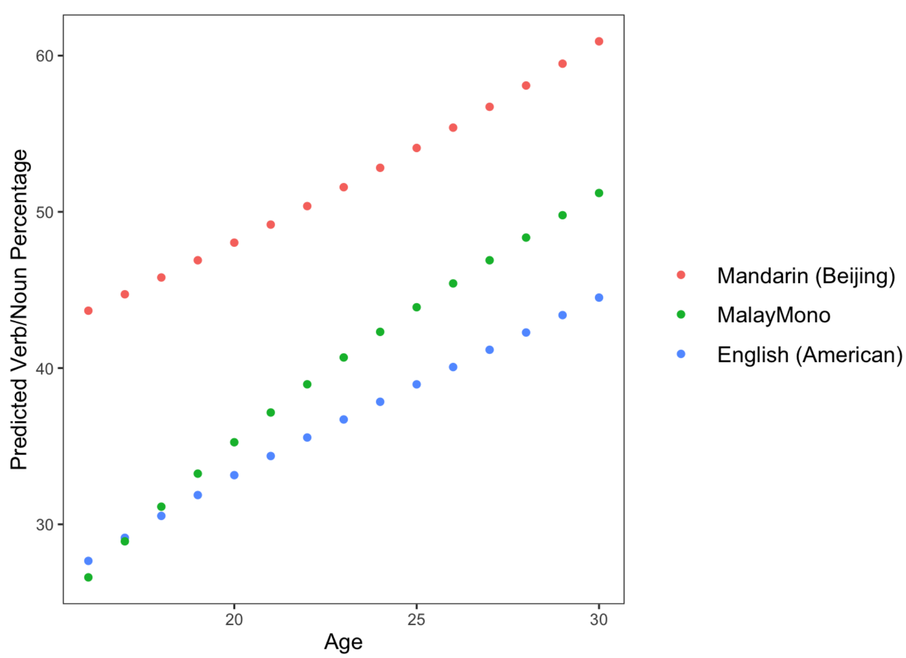


Figure 3-19: *Predicted English verb-to-noun percentage ratios for English (American), Malay for Malay monolingually-exposed children and Mandarin for Mandarin (Beijing) children using zero-inflation negative binomial model, by age.*

predictor of the verb-to-noun ratio. Mandarin of Mandarin bilinguals was richer in verbs than the Malay of Malay bilinguals, in comprehension – consistent with findings of their monolingual pair.

In production, the model revealed that age ($p < .001$) but not language group ($p = .256$) was a significant predictor of excessive zeros. Younger bilingually-exposed children were more likely to not possess verbs in their vocabularies yet. In contrast, language group ($p = .006$) but not age ($p = .645$) was a significant predictor of the verb-to-noun ratio. The Mandarin of Mandarin bilinguals was richer in verbs than the Malay of Malay bilinguals, in production – in contrast with findings of their monolingual peers, where Malay and Mandarin were similar in terms of noun-friendliness, but consistent with the findings of the comparisons made with Wordbank samples (see above).

3.5.5.4 Summary

Our results suggest that in comprehension, but not in production, the English lexicon of Mandarin-English bilingually-exposed children is more verb-friendly than that of Malay-

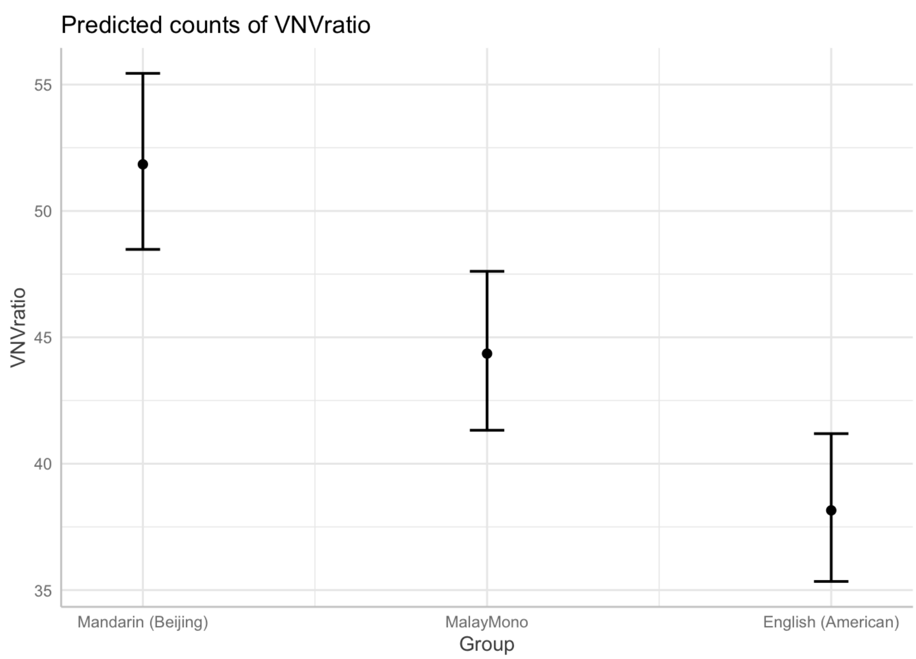


Figure 3-20: *Predicted English verb-to-noun percentage ratios for English (American), Malay for Malay monolingually-exposed children and Mandarin for Mandarin (Beijing) children using zero-inflation negative binomial model, averaged.*

English bilingually-exposed children. This addressed the first question and suggests that extra-linguistic factors can modulate the noun bias. With the exception of Malay productive vocabularies, the noun bias does not appear to differ between monolingually-exposed and bilingually-exposed children, overall (addressing the second question). Finally, we found Malay to be more verb-friendly than English yet more noun-friendly than Mandarin for both bilingually-exposed participants and monolingually-exposed participants (with the exception of productive vocabularies among monolingually-exposed children in the MCDI-M sample), addressing the third question. Overall, the pattern of results suggests that extra-linguistic factors modulate the noun bias in a unidirectional manner – modulating English rather than the children's ethnic language.

3.5.6 Discussion

The current study examined verb/noun distributions in the early lexicon of bilingually exposed infants and toddlers. Our first aim was to investigate whether what is known in the literature as a language noun bias (in our case, English) can be modulated by language-extrinsic factors. Our analysis with young children revealed that their English was less noun-friendly in comprehension (that is, the ratio of verbs to nouns was higher) among Mandarin-English bilinguals than it was among Malay-English bilinguals. This modulation of the noun bias cannot be attributed to language-intrinsic factors, as the focus is on the same language — English.

While the results suggest a modulation of the noun-friendliness of a language due to factors external to that language, we can only speculate about the mechanisms underlying such changes. A first interpretation is that the attentional patterns in parent-infant interactions modulate the structure of early lexicons. Chinese speakers tend to analyse visual scenes in a more holistic manner than English speakers, who tend to have focal attention towards individual objects (Nisbett, 2004; Nisbett & Miyamoto, 2005; Tardif et al., 1999; Waxman et al., 2016). Holistic processing tends to reveal relations between objects, typically referred to with verbs, whereas focal attention emphasises individual objects, typically referred to with nouns. Following this stream of reasoning, infants raised in Chinese-ethnic families are exposed to attentional patterns favouring the acquisition of verbs in all languages that they are learning, including English.

Another potential explanation is that some features of a language (in this case, the verb-friendliness) may bleed into the other language of a bilingually-exposed infant or toddler. One could imagine that, in an immature system, languages are not properly differentiated yet, and that the structure of a language is heavily influenced by the structure of the other language being learnt by the bilingually-exposed infant (e.g., Volterra & Taeschner, 1978; Redlinger & Park, 1980; Meisel, 1989). As the ethnicity of families is confounded with the language environment (Mandarin-English children are raised in Chinese families while Malay-English children are raised in Malay families), one cannot firmly advocate between both explanations from the current findings. Yet, much of the evidence in favour of a unitary language system hypothesis comes from observations of code-switching behaviour during childhood. This perspective is based on the production of speech, whereas the current results suggest that the structure of the English lexicon changes in light of the other language infants and toddlers are exposed to, in comprehension too. Thus, the explanation that infants confuse both of their languages, bringing the verb/noun distributions of each language towards the other, is at odds with strong evidence that infants and toddlers discriminate languages from a very early age (e.g., Werker & Byers-Heinlein, 2008; Genesee, 1989; Meisel, 2001; Bosch & Sebastián-Gallés, 2003). Furthermore, and addressing the second question in the study, the Malay and the Mandarin verb-to-noun ratios did not differ between monolingually-exposed and bilingually-exposed children (with the exception of productive vocabularies in Malay), suggesting an asymmetry in the noun-bias modulation: while the English lexicon changes, the other language being learnt by children remains similar. In other words, it seems to suggest that transfer occurs more in the direction from the stronger/more established L1 (here, Mandarin or Malay) to the L2 (here, English) which functions as a lingua franca. As such, English, as the second language, is more susceptible of socio-cultural modulations than the other languages.

A third explanation is that the characteristics of parental input, in English, differ across groups. One could expect that the English of parents of Mandarin-English bilingual children contains more verbs that the English of parents of English-Malay bilingual children.

Future work will aim at identifying the source of this modulation of the noun bias by comparing the word type and token produced by the parents during parent-child interaction with the verbs and nouns compositions of their children. A recent study has evaluated the verb to noun ratio of speech from parent-child interactions in a Singaporean bilingual population. Setoh, Cheng, Bornstein, and Esposito (2021) found that most Singaporean Chinese mothers either used more verbs than nouns, or used similar number of verbs and nouns, in English and in Mandarin, thus suggesting that differences in the composition of parental input could be driving differences in the composition of their child's vocabulary. However, this finding falls short of explaining how Mandarin-English bilingual adults used more verbs in English than, say, Malay-English bilinguals adults in the first place. Thus, any of the above-mentioned explanations (or a combination thereof) may account for the emergence of differences in the English lexicons of young and adult bilinguals.

In sum, our results suggest that the degree of noun-friendliness of a language can be influenced by factors external to that language. We argue that a likely explanation has the culture in which an infant is raised influencing the pattern of acquisition of verbs and nouns, possibly through differential attentional structures in her learning environment. Future research will investigate the link between attentional patterns in adults of different cultural communities and the verb/noun distributions in the lexicons of their children.

3.6 Conclusions

In this chapter, we have quantitatively examined early vocabulary acquisition patterns in Malaysian infants and toddlers, across three studies. Specifically, we have quantified the effects of internal factors (age and gender) and external factors (language exposure, SES and sociocultural influences) on lexical growth. As early vocabulary acquisition is known to be modulated by all of the above-mentioned factors, it is central that the MCDI-M is able to capture these effects on the Malaysian sample.

In Study 1, overall, we found a positive relationship between age and vocabulary sizes which confirms that the MCDI-M captures vocabulary growth. This relationship is modulated by relative language exposure (Oller & Eilers, 2002; Parra et al., 2011; Pearson et al., 1997; Place & Hoff, 2011) – since language exposure represents the experience the child had with a particular language thus affects the acquisition of the language. We found that girls outperformed boys in vocabulary size in production – a gender effect that was widely observed in CDI studies (see Frank et al., 2021, for cross-linguistic findings).

The findings from Study 2 suggested that external factors such as SES played a significant role in early vocabulary acquisition in Malaysia, as reported in other countries: US - Arriaga et al. (1998) and Layzer and Price (2008); UK - Blanden and Machin (2010); Australia - Taylor et al. (2013). The study provides a glimpse into what is driving language use in households (income more than education) and, in turn, what the result is in terms of their children's language skills (language input facilitates learning of English more than EL).

Study 3 evaluated the verb/noun distributions of children from the Malaysian sample collected for this thesis. The present study compares the relative distribution of verbs and nouns within the same language – English – between Malay-English and Mandarin-English bilingually-exposed infants and toddlers. The English receptive lexicons of Mandarin-English bilingual children contained more verbs than those of Malay-English bilinguals, suggesting that the noun-bias is modulated by factors external to English. The particular set of languages spoken in Malaysia (and Singapore) analysed in the present study, with Mandarin being verb-friendly, English being noun-friendly, and Malay in-between (as addressed by the third research question in the present manuscript), offers a unique opportunity to test competing theories about the origin of the noun bias.

In the next chapter, we introduce two novel computerised tools with the aim of improving the administrations and efficiency of early vocabulary acquisition assessments.

Chapter 4

Developing CDI-based Novel Vocabulary Assessments Tools

4.1 Overview

As mentioned in Chapter 2, CDIs are a cost-effective, reliable and valid set of parent-report instruments for assessing children's early language development from 8 up to 37 months of age (Fenson et al., 2007; Frank et al., 2017). Despite the advantages of using CDIs to collect vocabulary knowledge, the application of the forms is limited in many research and clinical settings due to its size. Although shorter forms of CDIs exist, the development of such forms is language-specific, time-consuming and labour-intensive, that is, to reduce the length of the CDIs while maintaining a balance of items from different semantic categories and with different difficulty levels.

To address these limitations, a language-general, computational approach is introduced and evaluated in this chapter. Study 4 focused on introducing a computational model merging the approach in Mayor and Mani (2018, see Chapter 2 for a review of the method) with item response theory (IRT, see Chapter 2 for a review of the method), hereafter referred to as the MM-IRT model. The performance of the MM-IRT model was evaluated with a range of already-collected CDI dataset extracted from Wordbank (i.e., American English, Danish, Beijing Mandarin and Italian). Building on Mayor and Mani (2018) work, we applied a principled selection of items in place of random selection. More specifically, the

IRT-based CAT was applied in real data simulations, as in Makransky et al. (2016). By applying IRT, maximally informative items are sampled based on each child's estimated ability (Fraley et al., 2000). Study 5 then evaluated the performance of the model on the much smaller and multilingual, Malaysian sample (as presented in Chapter 3).

Furthermore, as mentioned in the literature review, alternative measures to parental reports are needed due to the limitations of parental reports in representing the item-level word knowledge of the children (see literature review in Chapter 2). A direct language measure such as the computerised comprehension task (CCT) can serve both as a convergent and a supplemental measure of parental reports. In light of that, in Study 6, we examined the viability of tablets in assessing toddlers' word comprehension using a word recognition task similar to CCT, with the following objectives.

First, compared to traditional computerised task such as CCT (mounted on a desk in a lab setting, see Chapter 2), tablet-based assessments provide a more engaging and motivating testing environment, since they involve only minor motor movements and are far more portable due to the small form factor of tablets. Given that tablets are simple to use even for young children, and the increasing competency of children with tablets (Azah et al., 2014; Marsh, 2015), there is a need to investigate how such devices may be utilised to collect child language data most efficiently. Second, further research is needed to determine whether parental reports and children's word comprehension are aligned, and, in particular, whether parental evaluations fit best with their toddlers' word recognition in coarse (the semantically unrelated condition) or finer-grained contexts (the semantically related condition). At the item level, children vary in the strength of their word knowledge, and capturing this variability is critical for a robust understanding of a child's lexical development.

4.2 Study 4: Benchmarking the MM-IRT Model using the Stanford Wordbank database

4.2.1 Purpose

To validate our approach, four CDI-WS versions that varied in sample sizes on Wordbank were selected: American English (a very large dataset; Fenson et al., 2000), Danish (a large dataset; Bleses, Vach, Jørgensen, & Worm, 2010), Beijing Mandarin (a medium-sized dataset; Tardif & Fletcher, 2008), and Italian (a small dataset; Rinaldi, Pasqualetti, Stefanini, Bello, & Caselli, 2019)[1]. This, in turn, helped to evaluate the possibility of applying the MM-IRT model to languages for which only small samples are available (e.g., the Malaysian sample in Chapter 3). An evaluation of performance (in terms of correlation, standard error and reliability) was conducted across different age groups and genders.

The following subsections detail the two main components of the MM-IRT: i) the IRT-based selection of test items, coupled with ii) full-CDI score estimation method based on Mayor and Mani (2018)'s model.

4.2.2 Methods

4.2.2.1 IRT-based item selection

Prior to an assessment, items on the CDI are fitted to a two-parameter IRT model (2PL)[2]. Each item is assigned two parameters, i.e., a *discrimination* parameter and a *difficulty* parameter. These item parameters are computed with marginal maximum likelihood estimation using the Expectation-Maximization (EM) algorithm (Bock & Aitkin, 1981). During CAT assessments, items with maximum information are prioritised and tested. The estimation of the ability parameter of each child is conducted using the Weighted Likelihood Estimation (WLE) method (Warm, 1989) and is subsequently used to select items that can maximally inform about the knowledge level of the child. The test items are selected using

[1]These versions of the CDI-WS were selected for having relatively homogeneous sample sizes across all ages.

[2]Item parameters are estimated using the `mirt` function from Chalmers (2012)'s `mirt` package in `R`.

Chalmers (2016)'s `mirtCAT` package in `R`. This principled selection of test items on the basis of both the child's estimated ability and the properties of the test items is expected to improve the relevance of the items sampled in the algorithm. Based on children's responses on the items administered in CATs, the next step computes estimates of their full CDI scores.

4.2.2.2 CDI score estimation

The method of fitting each of the item-based histograms of full-CDI scores to a normal distribution and the estimation of CDI scores closely resembled those described in Mayor and Mani (2018). First, the language, gender and age in months of the child being assessed are matched with other children from Wordbank[3] (Frank et al., 2017). Then, for each test item *i* that is reported by the parent as either known or not known by the child, a histogram of full-CDI scores of all other children with the same response is extracted from Wordbank. The resulting histogram is fitted with a normal distribution using maximum likelihood estimation. A polynomial is fitted with the parameters (mean and standard deviation) extracted from the fitted histogram to smoothen out random fluctuations. Unlike Mayor and Mani (2018) whose polynomial fitting was conducted with a degree of three, the present model, took a more flexible approach, in which the degree of the polynomial adjusts to the breadth of the distribution of the vocabulary counts[4]. Once normalised, this histogram can be seen as the probability distribution of full-CDI scores given the child's response for item *i*. The histograms of all test items are subsequently log-summed and we retrieve the mode of the resulting histogram. Finally, a linear transformation of the mode, ensuring that the full range of CDI values associated with language-, age- and gender-matching children can be reached, produces the estimate of the child's full CDI score.

[3]The Wordbank database is accessed via Braginsky (2018)'s `wordbankr` package in `R`.

[4]To this end, the median absolute deviation (MAD) of productive vocabulary is computed for each age, in months respectively. When MAD < 100, a linear function is fitted to improve generalisation. When MAD > 100, a cubic polynomial is fitted to obtain a finer model.

4.2.2.3 Real data simulations

The data used in this study were from the American English (Fenson et al., 2000), Danish (Bleses et al., 2010), Beijing Mandarin (Tardif & Fletcher, 2008), and Italian (Rinaldi et al., 2019) CDI-WS, retrieved from Wordbank (Frank et al., 2017). The American English dataset was categorised as *very large-sized*, with more than 200 samples for each age (in months); the Danish dataset was categorised as *large-sized*, with between 100 and 200 samples for each age; the Beijing Mandarin dataset was categorised as *medium-sized*, with between 50 and 100 samples for each age; whereas the Italian dataset was categorised as *small-sized*, with fewer than 50 samples for each age[5]. These versions of the CDI-WS were selected for having relatively homogeneous sample sizes across all ages.

Using real data simulations, we compared the performance of the present model (the *IRT version*) with the original model by Mayor and Mani (2018) in estimating full-CDI scores. In addition, the raw vocabulary score on randomly selected items, scaled to the full CDI size, were used as a baseline measure and compared with the *IRT version*. Results for the original model and the baseline measure were averaged over ten simulations, whereas those for the *IRT version* were based on single simulations, since the item selection process relied on the ability estimation of each child, i.e., the same items will always be selected for the same children. In line with previous work using real-data simulations (i.e. Mayor & Mani, 2018; Makransky et al., 2016), correlations between the estimates (based on 5, 10, 25, 50, 100, 200, 400 and all items on each CDI) and the full-CDI scores, average *SE*s, and reliability (1 - SE^2), were reported across different age groups and genders.

Further evaluation of the performance of the *IRT version* was conducted with existing CDI-SFs. CDI-SF scores were obtained by summing raw vocabulary count based on responses on CDI-SF items and scaling these scores up to the instrument size to fit the range of full-CDI scores. Likewise, correlation coefficients of these scores with the full-CDI scores, reliability, and average *SE*s were computed and compared to those obtained from the *IRT version*.

[5]These classifications were made with regard to all normative data that is currently available in the Stanford Wordbank (Frank et al., 2017)

4.2.3 Results

4.2.3.1 Model selection

Two changes were made to the original model (Mayor & Mani, 2018): the application of IRT in item selection and the flexible approach to polynomial fitting. Preliminary comparisons between correlation coefficients of the original model, the original model equipped with flexible polynomial fitting, the original model with IRT (but without flexible polynomials), and the *IRT version* (with both flexible polynomial fitting and IRT) are shown in Figure 4-1. Correlations were compared using two CDIs having different sample sizes: the *very large-sized* American English CDI dataset and the *medium-sized* Beijing Mandarin CDI dataset, prior to selecting the final model. When applied to the *medium-sized* dataset, the combination of flexible polynomial fitting and IRT led to the largest improvements. For the *very large-sized* dataset, the mere application of IRT improved the model the most, although performance were comparable to the level of performance attained with the maximal model (IRT and flexible polynomials). With the merit that improvements were observed with the smaller dataset when both flexible polynomial fitting and IRT were

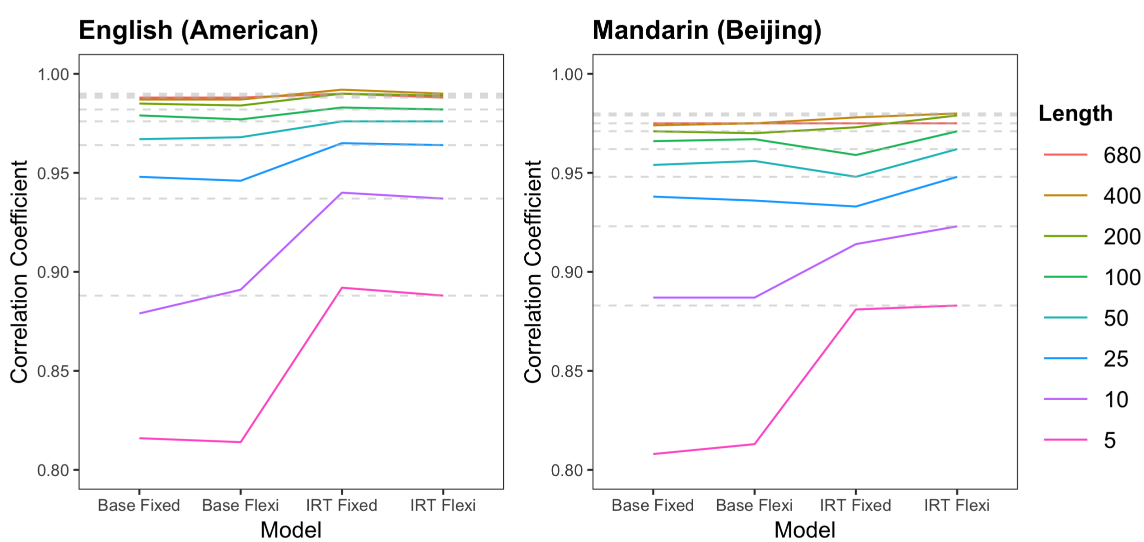


Figure 4-1: *A comparison between correlation coefficients of the original model (Base Fixed), the original model with flexible polynomial fitting (Base Flexi), the original model with IRT (IRT Fixed) and the IRT version with flexible polynomial fitting (IRT Flexi) for American English CDI and Beijing Mandarin CDI (the grey dashed reference lines represented the values for IRT Flexi).*

applied, we selected the *IRT version* as the final model.

4.2.3.2 American English CDI-WS

The original model and the *IRT version* were used in real data simulations on the *very large-sized* American English CDI-WS dataset for children between 16 and 30 months of age, for each gender, using a different number of test items (5, 10, 25, 50, 100, 200, 400, and 680, the full CDI size). Figure 4-2 shows the correlations between the estimated scores and the full-CDI scores using 5, 10, 25, 50, and 100 items, along with the average *SE*s, and the reliability of both the *IRT version* and the original model, as well as Makransky et al. (2016)'s values for comparison[6].

In terms of correlations, the *IRT version* performed better than the original model, with most correlations higher than .9, except at the 5-item test. In terms of both average *SE*s and reliability, the *IRT version* had values similar to the original model at 100 items, but outperformed the latter at 50 items and below. At 25 items, correlations greater than the .95 cut-off point, as suggested by Makransky et al. (2016), were already achieved. Additional real data simulations revealed that a correlation of .95 was already achieved at 14 items, with an average SE of .07 and reliability of .995.

[6]For the full list of values at all test lengths, across gender, see Table A.1.

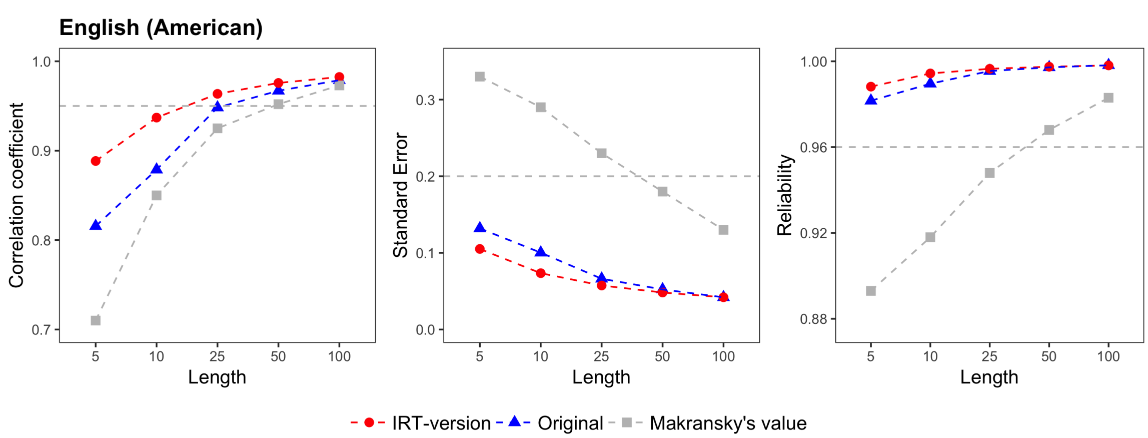


Figure 4-2: *Comparison of the IRT version and the original model with different test lengths (100, 50, 25, 10, 5) on the American English CDI-WS, with Makransky et al. (2016)'s values for reference. The grey dashed lines at .95 on correlation, .20 on SE and .96 on reliability represent the cut-off points suggested by Makransky et al. (2016). Note the x-axes are non-linear — the axes increase by a factor of two, except from 10 to 25.*

Table 4.1. Correlations from the *IRT version* and the original model (in parentheses) on the American English CDI-WS, by age group.

	16 - 18	19 - 21	22 - 24	25 - 27	28 - 30
680	.97 (.97)	.99 (.99)	1.00 (1.00)	1.00 (1.00)	.98 (.98)
400	.98 (.96)	.99 (.99)	1.00 (1.00)	.99 (1.00)	.98 (.98)
200	.99 (.96)	.99 (.99)	.99 (.99)	.99 (.99)	.98 (.98)
100	.98 (.95)	.99 (.98)	.99 (.99)	.98 (.99)	.98 (.98)
50	.98 (.94)	.99 (.97)	.98 (.98)	.97 (.98)	.97 (.96)
25	.96 (.92)	.98 (.95)	.97 (.96)	.96 (.96)	.95 (.94)
10	.92 (.81)	.95 (.87)	.95 (.90)	.94 (.90)	.92 (.89)
5	.87 (.74)	.92 (.82)	.92 (.84)	.89 (.85)	.84 (.82)

To further examine the effectiveness of the *IRT version* across ages, another analysis was conducted per age group (see Table 4.1). Improvements in correlations were observed at all age groups when compared to the original model. It is noteworthy that at 25 items, correlations across all age groups were already higher than .95, while in the original model, the youngest age group (16–18 months) and the oldest age group (28–30 months) required at least 100 and 50 items respectively, to achieve correlations of .95. In line with Makransky et al. (2016) and Mayor and Mani (2018), a marked reduction in performance was observed for both the youngest and the oldest age groups, when less than 10 items were used.

4.2.3.3 Danish CDI-WS

Real data simulations were conducted using the *large-sized* Danish CDI-WS (Bleses et al., 2008b) dataset, Figure 4-3 depicts performance of the *IRT version* and the original model for children between 16 and 30 months of age when having 5 to 100 items, with random lists as the baseline measure for comparisons[7]. Similar to the American English data, consistent improvements in correlations, average *SE*s and reliability were observed for the *IRT version*. With the Danish CDI-WS dataset, the *IRT version* was again able to achieve

[7]For the full list of values at all test lengths, across gender, see Table A.2.

correlations of above .95 at 25 items, as it did with the American English CDI-WS dataset, whereas the original model required at least 50 items to achieve a similar performance. Furthermore, the *IRT version* had better correlations, average *SE*s and reliability than the baseline measure at 50 items and below. Additional real data simulations revealed that a correlation of .95 was already achieved at 17 items, with an average SE of .06 and reliability of .997.

4.2.3.4 Beijing Mandarin CDI-WS

Real data simulations were run on the *medium-sized* Beijing Mandarin CDI-WS (Tardif et al., 2009) dataset for children between 16 and 30 months of age and the results[8] are illustrated in Figure 4-4 for tests with 100 items and below. In terms of correlations, the *IRT version* performed better than the original model, with similar, or better average *SE*s and reliability. With a relatively smaller sample size, the *IRT version* achieved correlations of .95 for male at 25 items and at 50 items for female. When compared to the baseline measure, the *IRT version* had better correlations at 25 items and below, while the average *SE*s and reliability were similar at 25 items and only better at 10 items and below. Additional real data simulations revealed that a correlation of .95 was achieved at 23 items for male, with

[8]For the full list of values at all test lengths, across gender, see Table A.3.

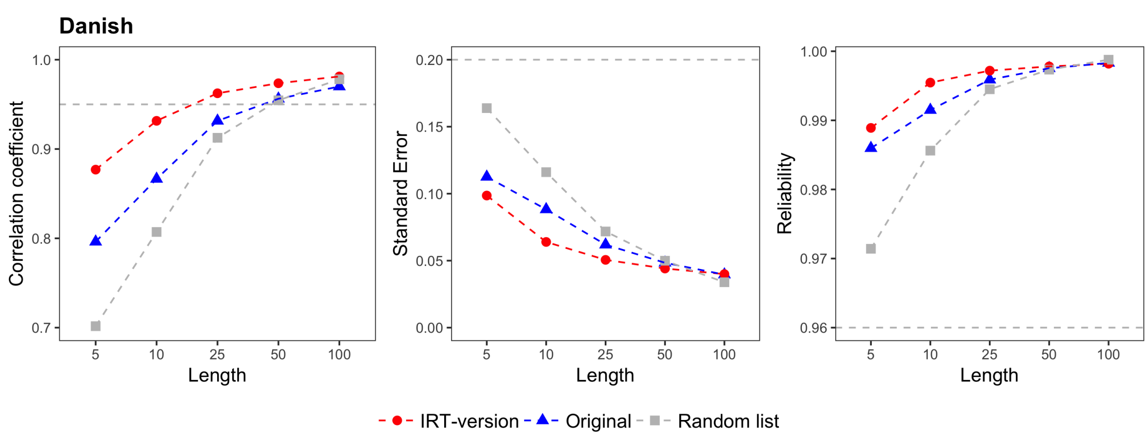


Figure 4-3: *Comparison of the IRT version and the original model with different test lengths (100, 50, 25, 10, 5) on the Danish CDI-WS, with random list as the baseline measure. The grey dashed lines at .95 on correlation, .20 on SE and .96 on reliability represent the cut-off points suggested by Makransky et al. (2016). Note the x-axes are non-linear — the axes increase by a factor of two, except from 10 to 25.*

an average SE of .09 and reliability of .992; and at 36 items for female, with an average SE of .08 and reliability of .993.

4.2.3.5 Italian CDI-WS

Real data simulations were conducted on the *small-sized* Italian CDI-WS (Caselli et al., 1995) dataset for children between 18 and 30 months of age. For this dataset, the original

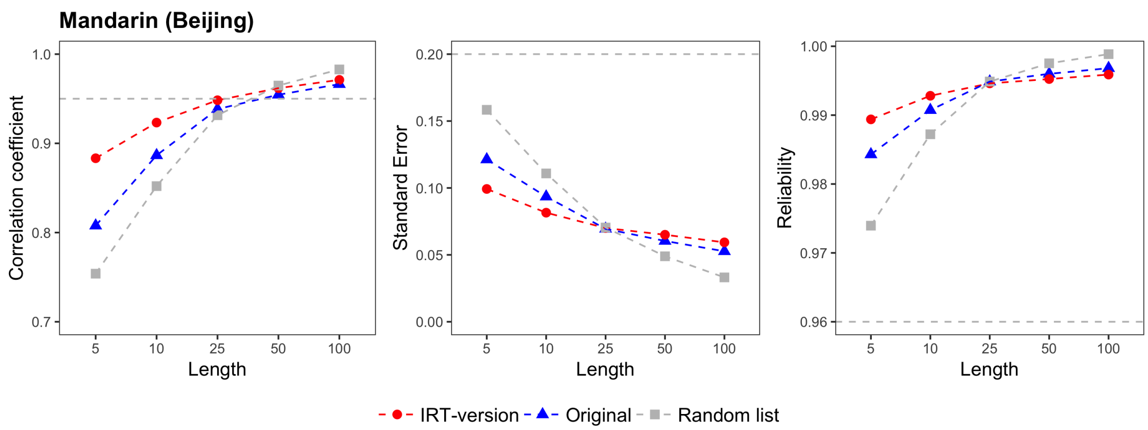


Figure 4-4: *Comparison of the IRT version and the original model with different test lengths (100, 50, 25, 10, 5) on the Beijing Mandarin CDI-WS, with random list as the baseline measure. The grey dashed lines at .95 on correlation, .20 on SE and .96 on reliability represent the cut-off points suggested by Makransky et al. (2016). Note the x-axes are non-linear — the axes increase by a factor of two, except from 10 to 25.*

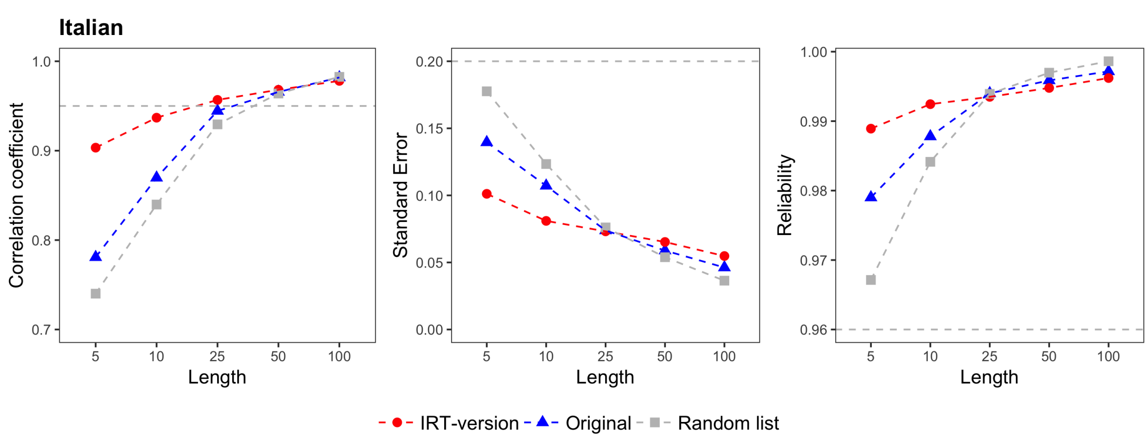


Figure 4-5: *Comparison of the IRT version and the original model with flexible polynomial fitting with different test lengths (100, 50, 25, 10, 5) on the Italian CDI-WS, with random list as the baseline measure. The grey dashed lines at .95 on correlation, .20 on SE and .96 on reliability represent the cut-off points suggested by Makransky et al. (2016). Note the x-axes are non-linear — the axes increase by a factor of two, except from 10 to 25.*

Table 4.2. Comparisons between the *IRT version* of the model, Fenson et al. (2000)'s American English CDI-SF and the baseline measure using 100 test items, by age group.

	IRT version			CDI-SF			Baseline		
Age	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
16 - 18	.982	.03	.999	.954	.04	.998	.975	.03	.999
19 - 21	.990	.03	.994	.973	.05	.997	.985	.04	.999
22 - 24	.985	.04	.986	.984	.05	.997	.988	.04	.998
25 - 27	.978	.05	.985	.986	.06	.997	.988	.04	.999
28 - 30	.978	.05	.994	.985	.04	.998	.987	.04	.999

model (with fixed degree of polynomial fit of 3) was unable to reliably estimate scores. Thus, we could only compare the performance of the *IRT version* with the original model with flexible polynomial fitting. As shown in Figure 4-5, the *IRT version* outperformed the original version in terms of correlations at 25 items and below, with better or similar average *SE*s and reliability[9]. Correlations of above .95 were already achieved with the *IRT version* starting at 25 items, whereas the original model achieved the same for females, but not for males (starting at 50 items). Correlations of the *IRT version* were better than the baseline measure at 50 items and below, while average *SE*s were better at 25 items and below, and reliability, at 10 items and below. Additional real data simulations revealed that a correlation of .95 was already achieved at 15 items, with an average SE of .08 and reliability of .993.

4.2.3.6 Comparisons with existing short-form versions of CDIs

Using 100-item tests (110-item tests for Beijing Mandarin), the performance of the *IRT version* was compared with the American English (Fenson et al., 2000), Danish (Bleses et al., 2010), Beijing Mandarin (Tardif & Fletcher, 2008), and Italian (Rinaldi et al., 2019) CDI-SFs. For a more detailed evaluation, comparisons were made across different age groups. Overall, all three approaches met the criterion for test acceptability suggested in Makransky et al. (2016) across all age groups and CDIs.

[9] For the full list of values at all test lengths, across gender, see Table A.4.

Table 4.3. Comparisons between the *IRT version* of the model, Bleses et al. (2010)'s Danish CDI-SF and the baseline measure using 100 test items, by age group.

	IRT version			CDI-SF			Baseline		
Age	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
16 - 18	.986	.02	.999	.968	.02	1.000	.972	.02	1.000
19 - 21	.978	.05	.997	.969	.03	.999	.973	.03	.999
22 - 24	.990	.04	.999	.983	.04	.998	.982	.04	.998
25 - 27	.981	.05	.997	.984	.05	.997	.983	.04	.998
28 - 30	.971	.06	.997	.985	.05	.997	.980	.04	.998

For the very large-sized American English CDI dataset, comparisons were made using Form A version of the American English CDI-SF (Fenson et al., 2000). Table 4.2 reported the correlations, the average *SE*s, and the reliability scores for the *IRT version*, CDI-SF, and the baseline measure, across five age groups. The *IRT version* performed better than CDI-SF in terms of correlations between 16 and 24 months, while CDI-SF performed better between 25 and 30 months. Overall, the IRT version had similar or better average *SE*s than CDI-SF, but reliability was consistently lower (except for the youngest age group, i.e., 16–18 months). The baseline measure outperformed the *IRT version* between 22 and 30 months as well as CDI-SF across all age groups, with better correlations, average *SE*s, and reliability.

Comparisons made using the *large-sized* Danish CDI dataset, across five age groups, among the *IRT version*, the Danish CDI-SF, and the baseline measure are reported in Table 4.3. The *IRT version* had better correlations than CDI-SF and the baseline measure between 16 and 24 months. After 24 months, CDI-SF performed best in terms of correlations. Overall the average *SE*s and reliability of the *IRT version* were similar, if not slightly poorer, when compared to both CDI-SF and the baseline measure.

For the *medium-sized* Beijing Mandarin CDI dataset, 110 items were administered in accordance with the number of items in the Beijing Mandarin CDI-SF. The results reported across five age groups in Table 4.4 indicated poorer performance of the *IRT version* in terms of the correlations, except for the youngest age group, i.e., 16–18 months. Average *SE*s and

Table 4.4. Comparisons between the *IRT version* of the model, Tardif and Fletcher (2008)'s Beijing Mandarin CDI-SF and the baseline measure using 110 test items, by age group.

	IRT version			CDI-SF			Baseline		
Age	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
16 - 18	.986	.06	.995	.980	.04	.999	.979	.04	.999
19 - 21	.984	.05	.998	.990	.05	.998	.990	.05	.998
22 - 24	.963	.07	.995	.981	.04	.998	.986	.04	.998
25 - 27	.961	.06	.997	.979	.04	.998	.983	.04	.999
28 - 30	.970	.06	.996	.981	.03	.999	.976	.03	.999

Table 4.5. Comparisons between the *IRT version* of the model, Rinaldi et al. (2019)'s Italian CDI-SF and the baseline measure using 100 test items, by age group.

	IRT version			CDI-SF			Baseline		
Age	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
18 - 21	.981	.04	.998	.972	.03	.999	.975	.03	.999
22 - 24	.990	.03	.999	.983	.04	.998	.982	.04	.998
25 - 27	.981	.04	.998	.984	.05	.997	.983	.05	.998
28 - 30	.971	.05	.998	.985	.05	.997	.980	.05	.998

reliability scores were also poorer than both CDI-SF and the baseline measure.

The final comparisons were made using the *small-sized* Italian CDI dataset, among the *IRT version*, the Italian CDI-SF, and the baseline measure, across four age groups. As reported in Table 4.5, the *IRT version* had better correlations than CDI-SF and the baseline measure between 18 and 24 months. Between 25 and 30 months, CDI-SF had the highest correlations. Average *SE*s and reliability were similar across all three approaches.

4.2.4 Discussion

CDIs are a cost-effective, reliable, and valid set of parent-report instruments for assessing children's early language development from 8 up to 37 months of age. However, due to their size, the administration of CDIs is time-consuming, thus restricting the applicability of CDIs when rapid assessments are desirable. The present approach, i.e., the MM-IRT (*IRT version* of the original model) combined Mayor and Mani (2018) Bayesian-inspired approach with an IRT-based CAT that dynamically selects test items that are maximally informative based on both the child's ability and the properties of the test items (as in Makransky et al., 2016). To evaluate the IRT version, real data simulations were conducted using four CDI-WS versions with varying sample sizes on Wordbank (Braginsky, 2018). Results obtained were subsequently compared with three other approaches: Mayor and Mani (2018)'s model (in a novel implementation, in `R`), a baseline measure (i.e., the sum of responses obtained directly from a set of items sampled randomly from the full forms), and CDI-SF. For the American English CDI-WS, Makransky et al. (2016)'s results were also included in the comparisons.

Overall, the MM-IRT model achieved correlations (with the full CDI) above .95, average *SE*s below .20, and reliability above .96 (a criterion for test acceptability suggested in Makransky et al. (2016)) with as few as 17 items for American English, Danish and Italian. For the Mandarin dataset, this criterion was only met from 23 items for males and 36 items for females.

To explain the uneven performance between both genders in the Mandarin dataset, we further inspected the dataset and found a much lower variation (quantified by MAD) in the female data than in the male data. More specifically, starting from 23 months, the female data was more left-skewed than the male data, i.e., a majority of females had high CDI scores, while males' scores continued to vary until about 27 months, when a majority, like females, began to have high CDI scores. The implication is twofold: first, it may be that, for girls, a larger sample size is needed for a better representation of the population; second, many items in the Mandarin CDI appear to be too easy, in particular for girls older than 23 months, hence reaching a ceiling earlier than boys. Despite this exception, our results suggest that a 25-item test can reliably estimate a child's CDI scores in most cases, regardless of gender and language. Analyses conducted per age group on the American English dataset extends this finding, further suggesting that a 25-item checklist is suitable for use with children across all age groups (16 – 30 months).

Comparisons with Mayor and Mani (2018) revealed that the MM-IRT model had similar or better performance in terms of correlations, average *SE*s, and reliability, across all four CDIs and both genders, regardless of the number of test items. In other words, the scores established by the MM-IRT matches more closely the full CDI scores. When compared against the baseline measure, the MM-IRT performed better in terms of correlations, average *SE*s, and reliability at shorter tests, i.e., having 50 items and below. It is noteworthy, however, that the baseline measure — summing a random selection of words — performed well across all four CDIs, already achieving correlations of above .95 with good average SEs and reliability, starting at just 50 items. At 100 items, the baseline measure's performance was also comparable to CDI-SFs. Such impressive performance should be attributed to the high internal consistency of CDI (e.g., Fenson et al., 1994; Bleses et al., 2008a; Tardif et al., 2009).

The final comparisons were made with CDI-SFs and the baseline measure across age groups, with 100 test items (or 110 for the Beijing Mandarin CDI-SF). While the MM-IRT model typically outperformed CDI-SFs in the younger age groups, i.e., between 16 and 24 months (with the exception of the Mandarin CDI-SF), both the baseline measure and CDI-SFs performed better in the older age groups, i.e., between 25 and 30 months. Nevertheless, the performance of the MM-IRT model was still comparable to the baseline measure and CDI-SFs, with all three approaches meeting Makransky et al. (2016)'s suggested criterion for test acceptability across all age groups. An important point to note here is that the development of CDI-SF for even just one language is often time-consuming and labour-intensive, whereas the MM-IRT model has the advantage of being cost-effective in that it is generalisable to a different language, i.e., it can be directly applied to CDIs of any languages, as long as sufficient CDI data is available online. Crucially, our objective is to develop a brief test that allows for rapid assessments – a 100-item checklist may still be considered too long in cases requiring multiple forms to be completed (e.g., in a linguistically diverse environment, a clinical setting) or intimidating when parents have low literacy. The MM-IRT model, In parallel, is able to provide reliable estimates with just 14 to 25 items, gaining a factor of four to seven compared to CDI-SFs.

The results reported here are based on real-data simulations. A full assessment of the psychometric properties of the IRT version should be conducted with new participants, in particular, to establish its test–retest reliability and its validity using an array of validity tests. With new participants, we also expect reduced level of performance as a result of parents responding differently to the same item in the full and short forms. This was demonstrated in Mayor and Mani (2018), with parents responding more positively in both the 25- and 50-item checklists than in the full CDI. In addition, as opposed to the more structured full forms that organize items into different semantic categories, our approach presents items in a semantically unstructured order, which may in turn affect parents' response behaviour. Therefore, the future research should include investigating the differences in parents' response behaviour and validating the model on new participants.

Finally, the reliability of our generic approach relies upon the availability of CDI data from children with matching key demographics (e.g., language, age, and gender). Based on our findings, even with a small data set having less than 50 samples available online for each age group (in months), our approach is able to reliably estimate children's full CDI scores with just 25 items, effectively reducing administration time to a mere couple of minutes. Thus, it is vital that data collected on children's vocabulary be shared publicly to enable access to and reuse of these data that will allow for the establishment of computerized adaptive tests that are tailored to each child.

The next study will explore the performance of the MM-IRT approach when applied to the current Malaysian sample. It will first be fitted to the paper version of the sample, and then be validated to assess how well it generalises to novel data (the online sample).

4.3 Study 5: Application of the MM-IRT Model on the MCDI-M data

4.3.1 Purpose

We have shown in previous chapters that the MCDI-Ms can be effective in capturing developmental trends and in providing insights inon the effects of environmental factors (i.e., SES and culture) on the language development of children in Chapter 2. Yet completion of the forms is time-consuming (it contains as many as $1,800$ items for trilingual), even for the online version[10], with completion times that ranged from 25 minutes to 2 hours and 30 minutes for the majority (80%) of these checklists. Given that a majority of children (73.6%) sampled using the online CDI were exposed to more than one language, the parents of these children still needed to assess at least $1,200$ words (or $1,800$ for trilingual), which explained the long completion time. Hence, the shorter version of CDI is even more beneficial in the Malaysian context, in which an overwhelming majority of its people are non-monolingual speakers. As an extension of the study from the previous section – which focused on the monolingual population, the current chapter sought to investigate the effectiveness of the MM-IRT model on the current Malaysian data (reported in Chapter 3).

The model was tested for its ability to generalise to a new sample. The term 'generalisation', in the context of a machine learning model, represents the model's ability to perform similarly well when deployed on a new dataset, rather than just the data it was trained on (Grilli & Remondino, 2020; Mitchell et al., 1997). In order to examine the generalisation performance of the MM-IRT model, we trained and deployed the model on the paper sample (the train set) and compared its ability to generalise when deployed to a different set of data, i.e., the online sample (the test set). The ability of the model to generalise is defined as follows: a small difference in performance (in terms of correlation, standard error and reliability) between the train and test sample would suggest that the model generalises well; whereas a poorer performance with the test sample would suggest that the model over-fitted the data and does not generalise well to a novel sample. Over-fitting occurs when a machine-learning algorithm fits an overly complex model to explain noises in the data due to a lack of an underlying pattern (see Mitchell et al., 1997). Any over-fitting we observe in the current study using the MM-IRT model would likely suggest poor sample sizes rather than a poor model, as the performance of the MM-IRT model was validated in Study 4, as well as the model's constituents were validated in previous studies, in simulation and empirically (Mayor & Mani, 2018; Makransky et al., 2016). Additional comparisons were also made with a manually assembled short-form of the MCDI-M, developed using the guidelines provided by Fenson et al. (2000).

[10]The online CDI forms were tailored to the language of the participants. That is, only vocabulary in languages that parents reported their children were exposed to was shown to the participants.

The current study used conceptual productive vocabulary of the children for two reasons: first, the conceptual vocabulary focuses on language knowledge of the children at a conceptual level and not on differences in vocabulary acquisition between languages; second, focusing on conceptual vocabulary increases the efficiency of the MM-IRT and the hand-picked list by reducing the time needed to assess each individual word, especially for non-monolingually exposed children.

4.3.2 Methods

4.3.2.1 Participants

The current study included 399 toddlers aged between 16 and 30 months of age (*M* = 23.23, *SD* = 4.22) from the Malaysian CDI sample. There were 269 toddlers from the sample of the paper version (131 girls and 138 boys) and 130 toddlers from the sample of the online version (57 girls and 73 boys).

4.3.2.2 The MM-IRT Model

The Bayesian-inspired IRT-based model approach introduced in the previous section was adopted in the current study to investigate its performance on the Malaysian CDI. Unlike Study 4, in which the model was implemented at each month of age, the toddler sample in the current study was aggregated into intervals of three months – 16 – 18, 19 – 21, 22 – 24, 25 – 27 and 28 – 30-month-old – to obtain a sufficient sample size[11] within each month-age group. The capability of the MM-IRT model in generalising was evaluated by using different datasets during fitting and testing – first, the model was fitted with the sample from the paper version and then performances on the paper and the online sample were compared.

Figure 4-6 displays the descriptive statistics in terms of means and .95 bootstrapped confidence intervals of the productive scores between the paper and online samples across age and gender (refer Table A.5 for the full set of descriptive statistics). It shows that the

[11] At least 20 samples per group were needed in order to fit both the IRT (for computation of item information) and the Mayor and Mani (2018)'s model (for full-CDI score estimation).

mean productive vocabulary sizes tend to increase over months of age, with consistently high variations in vocabulary size. It is noteworthy that most vocabulary sizes between the paper and online sample were similar, except for the 16 – 18-month-old group.

4.3.2.3 Establishing a Short-Form Baseline by Creating a Hand-Picked Selection of Words

The hand-picked list involved manual selection of 100 words with reference to the guidelines outlined by Fenson et al. (2000). First, for each age group (in months: 16 – 18, 19 – 21, 22 – 24, 25 – 27, 28 – 30), only words that were produced by more than 50% of children were shortlisted. Then, lower frequency words were excluded to sample consistently from each semantic category. Third, words related to religion (i.e, *church*, *mosque* and *temple*) or class discrimination words[12] (i.e., *maid* and *poor*) were excluded.

The next step involved ranking children from each month-old group into 3 percentile groups based on their production sizes: low (0 – 33), medium (34 – 67) and high (68 – 99) percentile groups as it is essential that the short forms be valid at different stages of

[12] This is to avoid the effect of SES – children from high-income families might be more likely to produce these words, compared to children from lower income families.

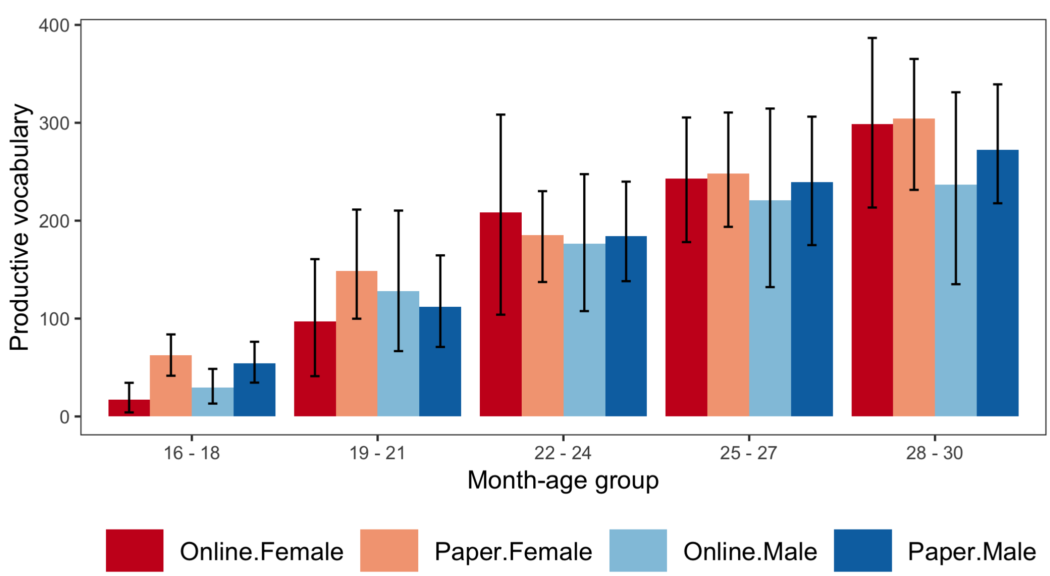


Figure 4-6: *Mean conceptual productive vocabulary and bootstrapped .95 confidence interval across age and gender.*

development across all ages (Fenson et al., 2000). The frequency of each word was then measured for each subgroup of children (5 levels for the age group × 3 levels for the percentile range group).

Then, deviating from Fenson et al. (2000), words with progression in rank across month-old groups were given priority in the selection process. For instance, words such as *No*, *Water*, *Hand* used by a majority (more than 60%) of the children aged between 16 – 18 months in the high percentile group were selected, given that older children from lower the percentile group were able to use these words, thus signalling the informativeness of these words about the language development of each child. In addition, similar to Fenson et al. (2000), difficult words (produced by 50 – 55% of all children) and easy words (produced by more than 80% of all children) were given priority to reduce ceiling and floor effects.

Finally, the distribution of items across semantic categories was compared with Fenson et al. (2000)'s American English CDI-SF and words were replaced to maintain similar composition. Specifically, for semantic categories with a number of items that exceeded (by more than 2) their counterpart in the American English CDI-SF, excess words with the lowest frequency were removed. Vacancies in the list were then filled up with initially excluded high-frequency items from other semantic categories, which had fewer items than their counterparts in the American English CDI-SF.

When conducting real-data simulation, each child's set of responses on the hand-picked list were summed up to generate a score, and this score was then scaled up to the instrument size to fit the range of full-CDI scores. Similar to the MM-IRT model, the manual word selection process was based on the data from the paper sample and then applied to both the paper and online data for evaluation.

4.3.3 Results

4.3.3.1 Items Selected from Both Approaches

Table 4.6 reports the categorical composition of the 100-item hand-picked list and the average percentage of conceptual items selected using the MM-IRT model[13], with the American

[13]The items were reported in percentages rather than in counts due to the dynamic nature of the word-lists generated using the MM-IRT model method (refer Figure A-1 for the histograms of the number of items

Table 4.6. Compositions of 100-word list from the MM-IRT model (mean number of items, with standard deviation in parentheses) when deployed on the paper and online sample; and the manual selection of items in the CDI-SF from the Malaysian list and from the American English list for comparison.

	MM-IRT model, MeanN_{items}(SD)		CDI-SF, N_{items}	
	Paper sample	Online sample	Malaysian	English
Nominals				
Animals	6.35 (5.23)	8.41 (7.34)	8	6
Vehicles	4.35 (1.61)	4.90 (1.55)	4	3
Toys	3.40 (1.51)	4.41 (1.69)	2	3
Food & drinks	6.50 (4.44)	7.14 (4.75)	9	8
Clothings	2.53 (1.46)	3.46 (2.01)	5	4
Body parts	9.26 (4.95)	7.48 (4.77)	5	4
Household items	12.40 (6.25)	10.26 (6.53)	6	7
Furniture & rooms	5.43 (2.70)	4.13 (2.81)	4	5
People	4.47 (4.34)	5.94 (5.69)	6	3
Places to go	-	-	2	2
Outside things	8.73 (4.56)*	8.63 (5.83)*	4	5
Sound effects & animal sounds	-	-	-	5
Total	54.17 (11.76)	53.09 (11.67)	55	55
Non-nominals				
Games & routines	6.69 (3.97)	7.38 (4.76)	5	5
Action words	21.08 (9.16)	18.62 (10.31)	13	14
Descriptive words	10.69 (8.64)	13.74 (9.78)	10	9
Words about time	2.61 (1.32)	2.41 (1.39)	3	3
Pronouns	4.17 (2.84)	3.53 (2.13)	3	4
Question words	1.72 (1.13)	1.43 (0.96)	1	1
Prepositions & locations	6.67 (2.86)	6.02 (3.27)	3	3
Quantifiers & articles	2.47 (2.11)	2.28 (1.58)	3	2
Helping verbs	2.01 (1.50)	2.04 (1.21)	3	3
Connecting words	4.12 (2.00)	3.27 (2.34)	1	1
Final particles	-	-	-	-
Total	45.83 (11.76)	46.91 (11.67)	45	45

*Some items in this category are translation equivalents of items from the *Places to go* category in other CDIs.

selected.).

English for comparison. Both approaches were applied to the conceptual vocabulary of the children but not to the total vocabulary nor the vocabulary in language(s) in order to focus on the language skills of the children on the conceptual level.

As expected, the item composition of the hand-picked word list (see Table A.6 for the 100-item list) is very similar with Fenson et al. (2000)'s American English CDI-SF given that the development of the forms were based on Fenson et al. (2000)'s guideline. This demonstrated the development effort of the Malaysian CDI-SF to maintain the number of words across semantic categories. In addition, despite a deviation of items due to dynamic selection of words (in terms of standard deviation in the number of items selected in parentheses of Table A.6), the composition of items selected using the MM-IRT model was similar with both the American English and the Malaysian CDI-SFs and this indicated that the MM-IRT model was able to preserve proportion of items as seen in the CDIs. One-sample t-tests revealed that the number of nominal and non-nominal items selected using the model did not differ from the number of nominal and non-nominal items in the CDI-SFs (i.e., 55 nominal items and 45 non-nominal items), when deployed on paper (ps = .249) and online sample (ps = .065).

For the MM-IRT model, the item compositions were similar between the paper and online sample. Yet it is noteworthy that the variations (measured as standard deviation) in the number of items selected from some semantic categories (e.g., animals, people, descriptive words) were as large as the mean number of items selected, in both samples. Given that the item selection of the MM-IRT model adjusts to the knowledge level of the child and that knowledge level increase over age, a close inspection was carried out to examine the item composition for semantic categories involved.

Figure 4-7 shows that average number of items selected tends to decrease over age for categories such as *Animals*, *Games & Routines* and *Food & Drinks*, in a factor of 2 to 3 when comparing between 16 – 18- and 28 – 30-month-old – an indication that these categories tend to lose its ability to inform about the relative knowledge level of older children. The words from these categories were often introduced in daily activities such as mealtime, storytelling and other routines thus are learnt earlier in life. In contrast, categories such as *Outside things*, *Descriptive words*, and *Question words* tend to increase over age, in a factor

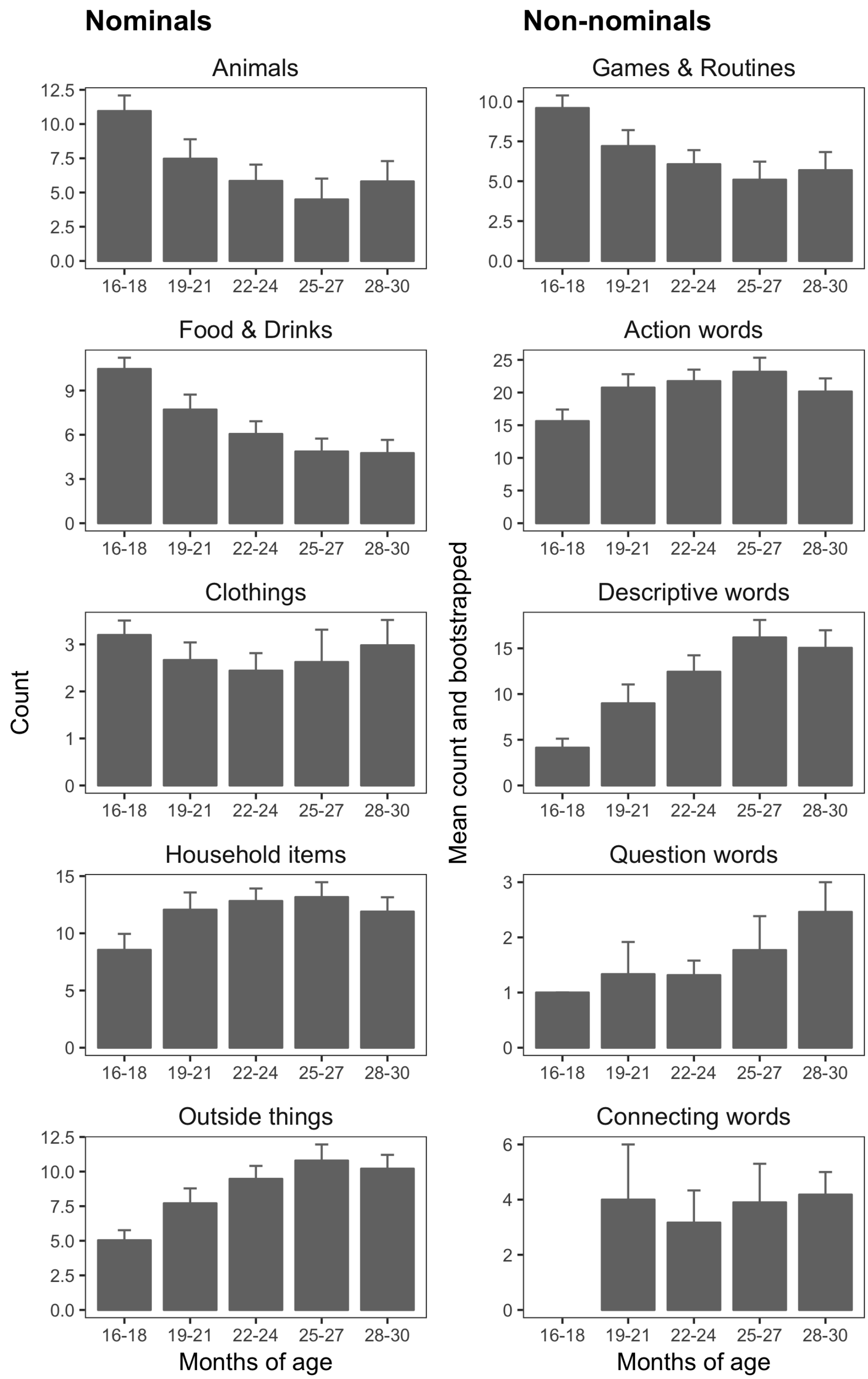


Figure 4-7: *Mean word count and bootstrapped .95 confidence interval of a selection of item categories from the MM-IRT model across month-old groups.*

of 2 to 3 between 16 – 18- and 28 – 30-month-old. The increase of *Outside things*-related items reflects the richer outdoor experiences that older children had relative to younger children. Older children were also more developed in their language skills thus capable of using more abstract words, e.g., *Descriptive words*- and *Question words*-related items. Nevertheless, some categories were relatively stable across month age, e.g., *Action words*, *Clothings*, and *Household items*. Finally, items from categories that required advanced level of language development such as *Connecting words* were not selected at all for the 16 – 18-month-old group.

4.3.3.2 Real-data Simulations

Real-data simulations were conducted to compare the performance of the MM-IRT model and of the hand-picked word list. Both approaches were tested on paper and online samples. Correlation coefficients between estimated and full-CDI scores, average standard errors and reliability scores were computed for both the paper (train sample) and the online sample (test sample) and were compared to evaluate the generalisability of the approaches. The analysis of the generalisation of the model examined whether the model is able to perform similarly when deployed on a new dataset.

In terms of the ability to generalise, Figure 4-8 contrasts the results of the real-data simulation of both approaches on the paper and the online sample using a series of test items (5, 10, 25, 50 and 100). The results of the MM-IRT model were based on single simulations, given that items selected from the CAT are fixed and specific to each child, whereas the results of the hand-picked word list were averaged over ten simulations due to random sampling of items for each run, except with the 100-item list, in which the full list of the hand-picked list was used. For correlation coefficients, smaller differences between performance on the paper and the online sample were observed for the hand-picked list when compared to the MM-IRT model, suggesting slightly better generalisation capacity of the hand-picked list. In contrast, the standard error and reliability measures showed that the MM-IRT model performed better than the hand-picked list, since relatively smaller differences between the train and test sample were observed for the MM-IRT model (with the exception at 5-item list). Importantly, starting at 100 items, all values for both samples from the MM-IRT model have already reached or exceeded the cut-off points for correlation, standard errors and reliability as suggested by Makransky et al. (2016).

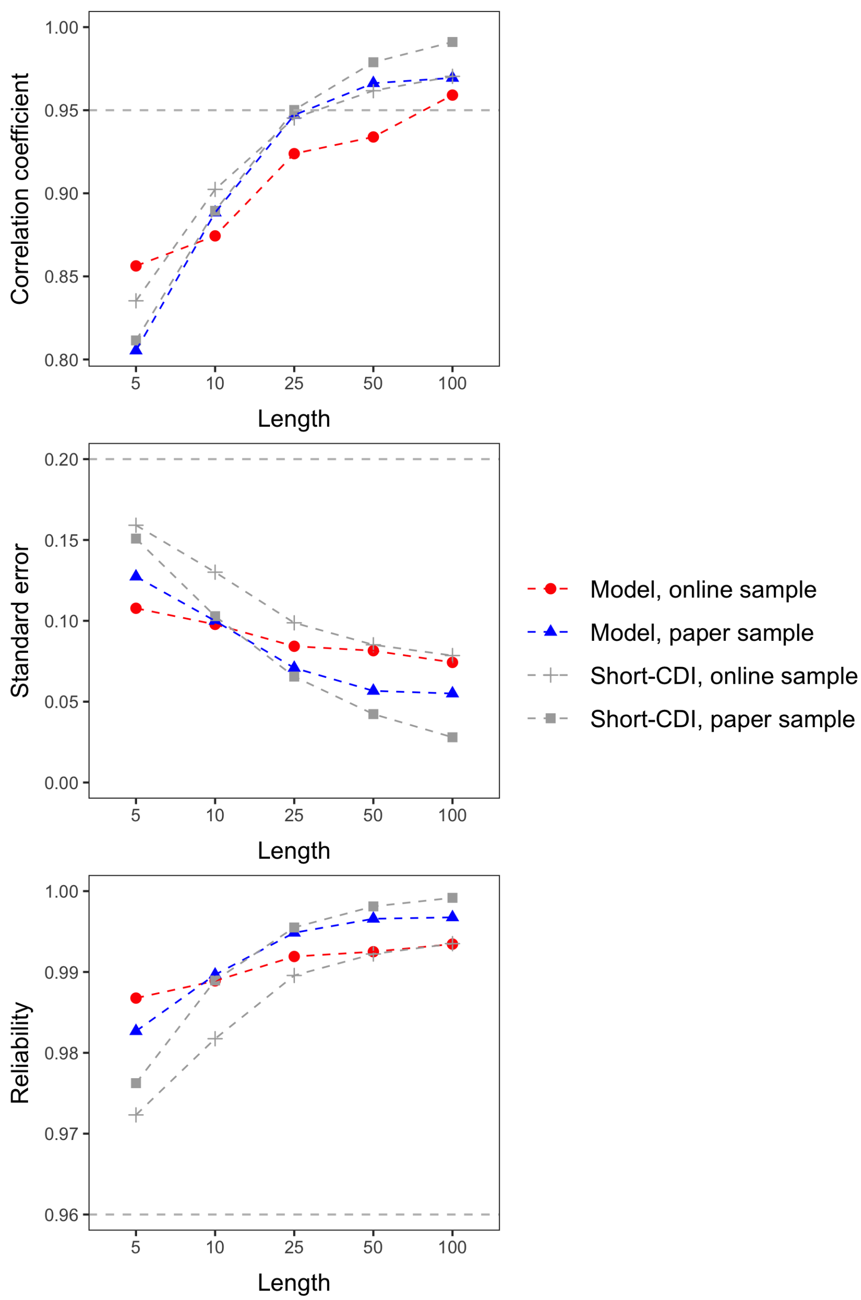


Figure 4-8: *Comparison of the performance of the MM-IRT model and the hand-picked 100-item short-CDI for the paper sample (train set) and online sample (test set).*
Note. *The grey dashed lines at .95 on correlation, .20 on SE and .96 on reliability represent the cut-off points suggested by Makransky et al. (2016). The x-axes are not linear. The values for the short-CDI are averages of 10 simulations.*

Upon closer inspection of the 100-item tests, Table 4.7 shows that the correlation scores of the MM-IRT model exceeded the .95 cut-off point for children aged between 22 and 30 months of age for both sets, with comparable standard error and reliability scores. In contrast, low correlation scores were reported for the 16 – 18-month-old and 19 – 21-month-old groups for the train and test sets respectively, and further simulations revealed that additional items were needed to achieve a correlation of .95 for both sets (125 items for both groups). For the other age groups, additional simulations revealed that a correlation of .95 was already achieved, for both sets of data, at 50 items in the 22 – 24-month-old, and at 25 items in the 28 – 30-month-old groups.

Table 4.7. Performance of the MM-IRT model on the Malaysian CDI (training/paper and test/online set), by month-old group, using 100 items.

100-item list	16-18	19-21	22-24	25-27	28-30
Training set					
r with full CDI	.970	.940	.981	.980	.976
Avg. SE	.05	.07	.05	.05	.06
Rel.	.998	.995	.997	.998	.996
Test set					
r with full CDI	.938	.984	.966	.951	.958
Avg. SE	.02	.09	.08	.08	.11
Rel.	1.000	.992	.994	.994	.989

4.3.4 Discussion

In contrast to a static word list (i.e., the hand-picked list), the MM-IRT model dynamically selects test items that are best at capturing the child's ability. The fact that more difficult words (e.g., *Descriptive words*, *Question words*) were selected in place of easier ones for older children by the MM-IRT model demonstrates that these children were more advanced in language skills. In contrast, easy words were prioritised for younger children as they are still in the process of grasping basic concepts such as names for food, drinks and animals. As children become more active as they grow, their outdoor experiences increase too. This is indicated by more *Outside things*-related items being selected for older children. On a side note, it would be interesting to compare the onset of *Outside things*-related items with other cultures, such as the Scandinavian culture, where *Forest School*s are very common due to the Scandinavian tradition of being close to nature (Knight, 2013).

To evaluate their effectiveness and the ability to generalise, both approaches were first applied to the dataset from the paper version of MCDI-M (i.e., the training set) and then tested with a test set (the dataset from the online version of the forms). Overall, the evaluation of correlation coefficients, standard error and reliability measures suggested that the MM-IRT model and the hand-picked list performed comparably in the ability to generalise – evidence that both approaches are able to perform similarly with new data.

Unlike in Study 4, the results are mixed when concerning the minimum number of test items required to reach the threshold of .95 across age groups, for both training and test set. Additional simulations suggested 125 items are needed for children between 16 and 21-month-old, whereas less than 100 items are needed for children between 21 and 30-month-old groups. It is noteworthy that the key difference between Study 4 and Study 5 lies in the treatment of age groups – while children were grouped according to their months of age in Study 4, the children in Study 5 were grouped into 3-month intervals. Hence, future work could improve the performance by building a larger dataset than in the current work, and grouping the children by their months of age, similar to Study 4.

In summary, for the current MCDI-M sample, test items ranged between 25 and 125 are needed when using the MM-IRT model, with adequate performance in terms of correlations with full-CDI scores and its ability to generalise across a range of month-old groups (except 16 – 18-month-old group). The item selection of the MM-IRT model adjusts to the knowledge level of each child, resulting in consistent performance across different samples (online vs. paper sample). Yet, the results reported here are still based on real-data simulations using existing data collected via full-length CDI. Hence, future work would be to conduct empirical validation – to validate the MM-IRT model with new participants, by comparing the performance of the model with the vocabulary data from the full-CDI forms. In the future, the assessments run by the MM-IRT algorithm and the full-CDI will be administered to a new group of parents on a separate occasion. The overall correlation between the estimated scores using the MM-IRT model and the scores obtained from the full-CDI forms will be computed to evaluate the convergence of the model. Item-level analysis will also be conducted to examine whether parents responded differently to the same items in the short list of words and in the full-CDI. Correspondingly, the next section examines a tablet-based direct-measure word recognition task as a supplemental index for vocabulary comprehension.

4.4 Study 6: CDI-based Tablet Assessment of Early Word Comprehension

4.4.1 Purpose

The current study utilised a Two-Alternative Forced Choice (2AFC) word recognition paradigm (similar to the CCT) to investigate the viability of using a tablet-based measure in assessing early word comprehension and to investigate the role of semantic relatedness in early word recognition, starting with monolingually-exposed Norwegian toddlers between 18- and 20-month-old. As the CCT is only available in three languages (English, Spanish, and French), lexical items with varying levels of difficulty (defined based on the normative sample) were selected from the Norwegian adaption of the CDI–Words and Gestures (CDI-WG; Simonsen et al., 2014)[14]. Within each trial, two images are displayed on the screen: one representing the lexical target and the other representing the distractor.

The current study examined the role of semantic relatedness in toddlers' word recognition performance, by pairing the lexical target with a distractor from a different semantic category (e.g., a car and a cat) and with another distractor from the same semantic category (e.g., a car and an airplane), respectively. It was expected that toddlers would be more accurate in semantically unrelated trials than in related trials. Accuracy was expected to mirror a priori difficulty levels (based on data from parental reports), with accuracy decreasing with increasing difficulty, based on previous work using the CCT (Friend & Keplinger, 2003, 2008).

[14] The CDI-WG assesses word comprehension.

Then, potential differences between data collected online and in the laboratory were examined to verify whether toddlers' motivation to produce a response (attempted trials) and their performance (correct trials) differed between settings.

Finally, we evaluated whether parental reports converge with toddler's word recognition – a positive relationship between parent-reported comprehension and the toddler's accuracy in word recognition would suggest convergence of the measures.

4.4.2 Method

4.4.2.1 Participants

Parents of 49 monolingually-exposed Norwegian toddlers (aged between 18- and 20-month-old) from the Greater Oslo Region, Norway, were contacted to participate in the current study through social media, leaflets distributed in a kindergarten, postal mailing lists, and email lists. After consenting to participate in the study, parents completed the Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) online within one week prior to the study so that the current estimates of their child's vocabulary size could be obtained. Parents' socioeconomic status (SES), indicated by the mother's highest education level, ranged from 0 (primary school) to 5 (doctoral degree), with the mean score 3.57 (SD = .82).

All recruited toddlers were full-term at birth, had no hearing or visual impairments, and had Norwegian as their native language. Toddlers participated in the study in one of three settings: the BabyLing laboratory, a municipal kindergarten, and online (i.e., at toddlers' own homes). Data was initially collected in the laboratory and kindergarten. Due to the COVID-19 pandemic-related lockdown in Norway (Kalajdzic, Krüger, & Venli, 2020), data collection proceeded online. In both the laboratory and the kindergarten settings, toddlers were tested by an experimenter, whereas online, toddlers were tested by their parents. Parents consented to not interfere with the task or influence their children's responses. Thus, for simplicity, both the laboratory and kindergarten samples (n = 21; 16 females, 5 males) were categorized under the laboratory setting, and the online sample (n = 28; 15 females, 13 males), the online setting. An additional 11 participants had to be excluded for failing to complete the task (n = 7; 2 laboratory and 5 online) and for attempting the task more than once (n = 4; all online). The age range for the laboratory setting was between 18- and 20-month-old (*M* = 19.29, *SD* = .60) and for online setting was between 19- and 21-month-old (*M* = 19.63, *SD* = .63).

The study was approved by the ethics committee at the Department of Psychology, University of Oslo and by the Norwegian Centre for Research Data (NSD, ref. 807456).

4.4.2.2 Design

The present study used a within-subjects design. Toddlers' comprehension of 24 lexical items of three levels of difficulty (easy, moderately difficult, and difficult; see Lexical Items section, below) was assessed using a tablet-based 2AFC word recognition task. Lexical targets were assessed under two conditions: semantically related (i.e., the lexical target was presented with a distractor from the same semantic category) and semantically unrelated (i.e., the lexical target was presented with a distractor from a different semantic category).

4.4.2.3 Lexical Items

Four highly familiar[15] lexical items were selected for the familiarization phase: "ball" [ball], "hus" [house], "sko" [shoe], and "tre" [tree]. For the test phase, a total of 24 lexical items that belong to the nominal categories (i.e., animals, vehicles, household items, clothings, outside things, food) were selected. Each lexical target was assessed twice, by pairing its referent with semantically related and unrelated referents as distractors. Word pairs varied in difficulty (defined a priori on the basis of the Norwegian CDI–WG normative data for 20-month-old; Frank et al., 2017; Simonsen et al., 2014) and were comprised of an equal number of easy (comprehended by more than 80% of the normative sample), moderately difficult (comprehended by 40%–80% of the normative sample), and difficult (comprehended by less than 40% of the normative sample) items. Within each level of difficulty, there was also an equal representation of animate and inanimate referents. The list of word pairs is provided in Table 4.8.

[15]On the basis of the Norwegian CDI-WG normative data (Frank et al., 2017; Simonsen et al., 2014).

4.4.2.4 Two-alternative forced choice (2AFC) Recognition Task

The study was conducted via a custom-based online experimental platform developed by Lo et al. (2021). In the laboratory setting, a Samsung Galaxy Tab S4 was used to run the study, whereas in the online setting, parents' own touchscreen devices were used. The Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) was used as a measure of comprehensive vocabulary size.

Audio stimuli were used as a prompt for the toddlers to touch the target. The audio stimuli consisted of a series of audio prompts recorded by a female, native Norwegian speaker. A total of 24 unique sentences were recorded, with the target object at the end of each sentence: 'Can you touch the ____?'. The speaker was asked to speak in child-directed speech, i.e., with a slower tempo and an elevated pitch. The recordings were processed in Praat (Boersma, 2001) to reduce noise and to equalise the intensity across the 24 audio prompts. The same speaker also recorded encouraging sentences such as 'Here you go!' and 'You are almost done!' to maintain child's attention and to keep the child motivated.

Table 4.8. Word pairs used in the 2AFC task, that are either semantically related or semantically unrelated, across difficulty levels.

Difficulty level	Semantically related	Semantically unrelated
Easy	bil [car]—fly [airplane]	hest [horse]—banan [banana]
	eple [apple]—banan [banana]	hund [dog]—fly [airplane]
	hest [horse]—ku [cow]	katt [cat]—bil [car]
	hund [dog]—katt [cat]	ku [cow]—eple [apple]
Moderate	elefant [elephant]—tiger [tiger]	elefant [elephant]—saks [scissors]
	lastebil [truck]—tog [train]	løve [lion]—tog [train]
	saks [scissors]—blyant [pencil]	sjiraff [giraffe]—lastebil [truck]
	sjiraff [giraffe]—løve [lion]	tiger [tiger]—blyant [pencil]
Difficult	elg [moose]—pingvin [penguin]	elg [moose]—pasta [pasta]
	gås [goose]—ugle [owl]	gås [goose]—shorts [shorts]
	pasta [pasta]—sukkertøy [candy]	pingvin [penguin]—sukkertøy [candy]
	shorts [shorts]—glidelås [zipper]	ugle [owl]—glidelås [zipper]

The visual stimuli consisted of 48 unique images with 2 images representing each word to reduce familiarity bias due to repeated testing. Identical items could differed in terms of orientation (e.g., airplanes, elephants), colour (e.g., cars, horses), posture (e.g., cats, giraffes) and/or variants (e.g., candies, houses) – see Appendix B-3. Images were rated for typicality by 21 native Norwegian speakers using Likert scale that ranged from 1 to 5 ('Not at all typical' to 'Very typical'). Among them, 90% of the raters agreed that all pictures were very typical (*M* = 4.69, *SD* = .25). This is important to ensure children's responses do not systematically differ in terms of accuracy, given that children are more accurate and quicker in identifying typical items than atypical items (Jerger & Damian, 2005; Rosch, 1973). The images were presented in pairs on a white background (960 x 960 px). The image pairs were edited to be approximately matched in brightness and size. Within each word pair, the side (left or right) on which a target appeared was counterbalanced. Figure 4-9 and Figure 4-10 depict image pairs with either semantically related or semantically unrelated (for the complete set of stimuli in the main trials, see B-3; for stimuli in the

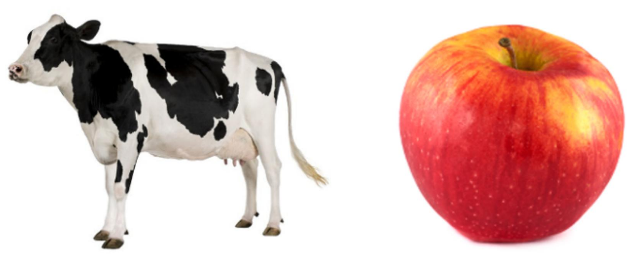


Figure 4-9: *Example of semantically-unrelated (cow-apple) pairs.*

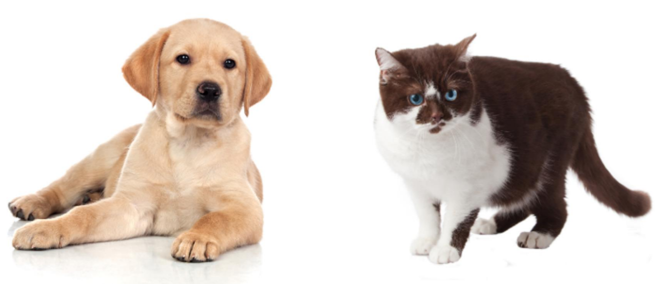


Figure 4-10: *Example of semantically related (dog and cat) pairs.*

practice trials, see B-4).

There are two phases in the recognition task: the training phase and the test phase. During the training phase, four control trials consisting of highly common objects, i.e., house, shoes, ball and trees, are presented in pairs to familiarise toddlers with the task.

During the test phase, each participants are presented with 48 test trials in a random order, with 24 words assessed twice, to appear once as a target and once as a distractor, and counterbalanced for image position. Encouraging messages were played between the trials to signal breaks and the 'next' button must be pressed to continue.

4.4.2.5 Procedure

Parents were given a consent form and subsequently an online questionnaire that asked about basic demographic information and their child's linguistic environment, as well as the Norwegian adaptation of the CDI:WG digitalised form (Simonsen et al., 2014). All parents completed the online questionnaires within a week before the word recognition task.

For the experimenter-ran sessions, prior to the task, the experimenter performs a warm-up session to ensure that each toddler is comfortable proceeding to the task. To initiate the task, the experimenter told the toddler that they were going to play a game. Parents were told not to interfere with their children during the test. Both the experimenter and the parents stayed with their children throughout the test. All instructions during the task were pre-recorded and delivered by the online experimental platform. After the test, the toddlers were compensated with a small toy that they picked. The data collection process in the kindergarten was similar with the lab. Each toddlers was accompanied by the experimenter and a kindergarten teacher instead of their parents. The test was conducted in a quiet room that is familiar to the toddler. The kindergartens were compensated with a donation of picture books to their library.

For the parent-ran sessions, in addition to the consent form and the online questionnaire, parents were also sent an additional link to the online experimental platform for the word recognition task. Again, parents were told not to interfere with their children during the test. Parents were allowed to exit the task if their children were not interested in playing with the task. When touch-screen device is not available, parents were told to substitute with a computer. The children make their responses by pointing on the monitor and their parents records those responses using a mouse.

4.4.3 Results

The results attempt to answer three central questions. First, potential differences between settings were examined in terms of toddlers' motivation to produce a response (attempted trials) and their performance in producing a correct response (correct trials) in the word recognition task. Second, we examined the influence of semantic relatedness and word difficulty on attempted trials as well as on correct trials. Finally, the convergent relationship between toddlers' performance and parental report (CDI–WG) was assessed in terms of parent-child agreement. In accordance with previous work using the CCT (Friend & Keplinger, 2003; Friend et al., 2012), missing responses (i.e., trials in which the child did not produce a response) were treated as errors of comprehension[16].

4.4.3.1 Attempted trials

The toddlers' motivation to produce a response during the word recognition task was measured in terms of the number of trials in which a tap response was produced, regardless of whether the response was correct (i.e., tap on target) or incorrect (i.e., tap on distractor).

Potential differences between settings were compared in a Welch's t-test, which indicated that toddlers who were tested online (M = 44.27, SD = 6.36) and those who were tested in the laboratory (M = 4.81, SD = 7.06) did not differ significantly in the number of attempted trials ; $t(4.60) = -1.78$, $p = .08$ (see Figure 4-11).

To assess whether toddlers' motivation differed across semantic relatedness and difficulty of the trials, a binomial Generalised Linear Mixed Model (GLMM) with a logit link function was fitted and analysed using the *mixed* function from the *afex* package (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2015), which relies on the *lme4* pack-

[16] Toddlers produced fewer tap responses for difficult trials than for easy trials, suggesting that unanswered trials are not random, but predominantly reveal errors of comprehension (Friend & Keplinger, 2003; Friend et al., 2012).

age (D. Bates, Mächler, Bolker, & Walker, 2015) for model fitting.

The model included semantic relatedness (related, unrelated), difficulty (easy, moderately difficult, and difficult), toddlers' age (in months), and the interaction between semantic relatedness and difficulty as fixed effects, as well as participant and selected object as random intercepts. Both semantic relatedness (-1: unrelated; 1: related) and difficulty (-1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age was centred on the mean. To determine a model with a parsimonious random effect structure (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), the α = .20 was used to test random slopes for inclusion using the Forward "best path" algorithm (Barr, Levy, Scheepers, & Tily, 2013). The random slopes were included in the final model, as none of the random slopes fell below the inclusion criterion:

$$\text{Attempted} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 \mid \text{Participant}) + (1 \mid \text{Object})$$

The model fitted and comparisons are presented in Table 4.9, with chi-square statistics and p-values obtained using likelihood ratio tests. Significant effects are further analysed,

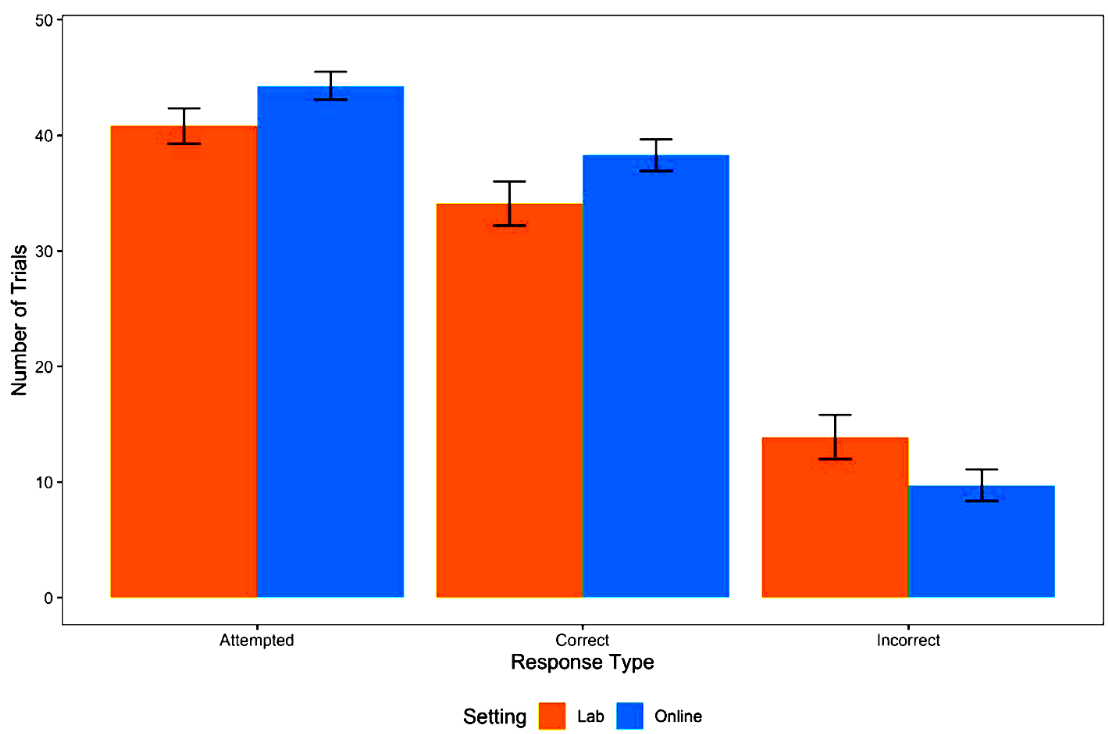


Figure 4-11: *The number of attempted, correct, and incorrect trials across lab and online setting.*

with p-values adjusted using the Tukey method, using the *pairs* function in the *emmeans* package (Lenth, 2020).

As shown in Table 4.9, there were significant main effects of trial difficulty and age, with the number of attempted trials increasing with age. No significant main effect of semantic relatedness was found; neither did semantic relatedness interact with difficulty. Results from the follow-up tests indicated that toddlers attempted significantly more easy than difficult trials (β = .556, *SE* = .186, z = 2.995, p = .008), while no such difference was found between easy and moderately difficult trials (β = .363, *SE* = .189, z = 1.917, p = .134) as well as moderately difficult and difficult trials (β = .193, *SE* = .176, z = 1.096, p = .517; see also Figure 4-12).

Table 4.9. GLMM Model summary and comparisons for attempted trials.

	Model summary			Model comparison		
	β	*SE*	z	χ^2	*df*	p
Intercept	3.080	.281	1.956	103.539	1	<.001
Relatedness	-.087	.075	-1.163	1.355	1	.244
Difficulty				8.516	2	.014
Moderate	-.057	.105	-.542			
Difficult	-.249	.103	-2.432			
Age	.949	.395	2.402	5.686	1	.017
Relatedness:Difficulty				1.618	2	.445
Relatedness:Moderate	-.106	.105	-1.006			
Relatedness:Difficult	.116	.102	1.136			

4.4.3.2 Correct trials

The toddlers' accuracy was measured in terms of the number of trials in which they correctly identified the target referent.

Results from a Welch's t-test indicated that there was no statistically significant difference in toddlers' performance between those who were tested online (*M* = 38.286, *SD* = 7.262) and those who were tested in the laboratory (*M* = 34.095, *SD* = 8.717), $t(38.508) = -1.787$, $p = .082$ (see Figure 4-11).

To assess whether toddlers' accuracy differed across semantic relatedness and difficulty of the trials, a binomial GLMM with a logit link function was again fitted and analysed. The model included the same fixed effects as the previous model (i.e., semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty) and the same random intercepts (i.e., participant and selected object), with by-participant adjustments to the slope of difficulty[17].

[17] The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

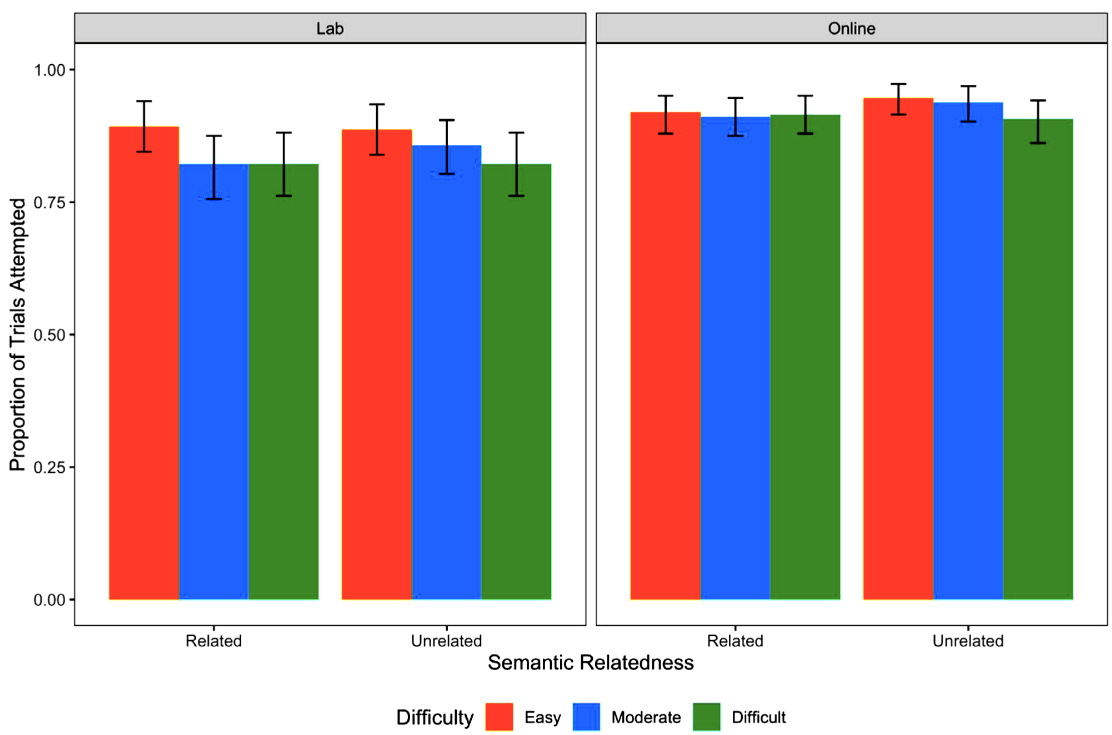


Figure 4-12: *Proportion of attempted trials across settings by semantic relatedness and difficulty*

$$\text{Accuracy} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 + \text{Difficulty}|\text{Participant}) + (1 \mid \text{Object})$$

The results are detailed in Table 4.10, with chi-square statistics and p-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p-values adjusted using the Tukey method.

As shown in Table 4.10, there were significant main effects of semantic relatedness, difficulty, and age. Specifically, toddlers responded with higher accuracy in semantically unrelated than related trials. No interaction was found between semantic relatedness and difficulty. A significant positive relationship was also found between toddlers' accuracy and age. Results from the follow-up tests indicated that toddlers were significantly more accurate in easy trials relative to both moderately difficult (β = .523, SE = .183, z = 2.861, p = .012) and difficult trials (β = 1.113, SE = .164, z = 6.799, p < .001). Toddlers were also significantly more accurate in moderately difficult than difficult trials (β = .590, SE = .150, z = 3.924, p < .001; see also Figure 4-13)[18].

[18] A Spearman correlation between toddlers' overall word recognition accuracy and SES revealed no relationship, rho = .1, p = .46.

Table 4.10. GLMM Model summary and comparisons for accuracy.

	Model summary			Model comparison		
	β	SE	z	χ^2	df	p
Intercept	1.438	.143	1.038	56.979	1	<.001
Relatedness	-.141	.054	-2.624	6.782	1	.009
Difficulty				36.405	2	.001
Moderate	.022	.097	.229			
Difficult	-.568	.085	-6.660			
Age	.537	.193	2.779	7.233	1	.007
Relatedness:Difficulty				3.887	2	.143
Relatedness:Moderate	-.114	.076	-1.511			
Relatedness:Difficult	.127	.071	1.785			

4.4.3.3 Convergent Validity

At the general level, toddlers' receptive vocabulary size, as measured by the CDI–WG, and their overall accuracy in the word recognition task significantly correlated in both unrelated, $r(47) = .631$, $p < .001$ and related trials, $r(47) = .603$, $p < .001$. Partial-ling out the effect of age further revealed that toddlers' receptive vocabulary size accounted for a significant proportion of unique variance in their recognition accuracy, beyond that accounted for by their age in both unrelated, $r(46) = .593$, $p < .001$, $R^2 = .352$ and related trials, $r(46) = .538$, $p < .001$, $R^2 = .289$.

To explore the consistency between toddlers responses and parent-reported comprehension on the test items (i.e., parent–child agreement), item-level agreement was calculated (see Table 4.11) and a binomial GLMM with a logit link function was fitted. The model included semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty as fixed effects. Both semantic relatedness (-1: unrelated; 1: related) and difficulty (-1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age

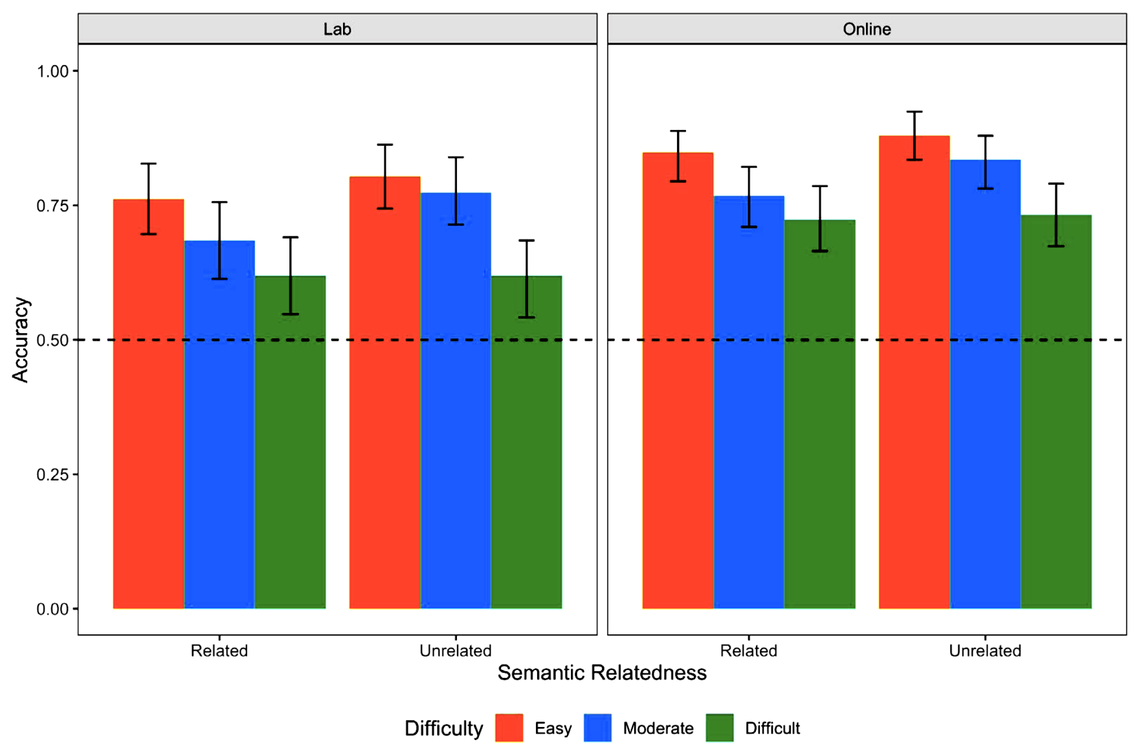


Figure 4-13: *Accuracy by semantic relatedness and difficulty across different settings. Dashed line represents chance level at .5.*

was centred on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slopes of semantic relatedness, difficulty, and their interaction[19]:

$$\text{Accuracy} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 + \text{Relatedness*Difficulty|Participant}) + (1 \mid \text{Object})$$

Table 4.11. Item-level agreement between parental report and toddlers' performance.

Difficulty level	Semantically related	Semantically unrelated	Overall
Easy	.781	.827	.804
Moderate	.564	.538	.551
Difficult	.615	.661	.638
Overall	.653	.675	.664

The GLMM results are detailed in Table 4.12, with chi-square statistics and p-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p-values adjusted using the Tukey method.

Overall, as shown in Table 4.12, there was good item-level agreement between parental reports and toddlers' responses, although this was modulated by item difficulty. The results from the GLMM indicated that semantic relatedness, difficulty, as well as the interaction between semantic relatedness and difficulty (but not age) significantly predicted parent–child agreement (see also Figure 4-14). The follow-up tests revealed that parent–child agreement was significantly higher in semantically unrelated than in related easy trials (β = .795, *SE* = .299, z = 2.662, p =.008), but no significant differences were found across the different semantic conditions in the moderately difficult (β = .253, *SE* = .169, z = 1.495, p =.135) and difficult trials (β = -.166, *SE* = .164, z = -1.014, p = .311).

To further examine whether item-pair comprehension status (i.e., whether the target or the distractor label was known or not known by the toddler as indicated by parental responses on the CDI–WG) was an accurate predictor of toddlers' performance in the word

[19]The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

Table 4.12. GLMM results for parent–child agreement.

	Model summary			Model comparison		
	β	*SE*	*z*	χ^2	*df*	*p*
Intercept	.921	.163	5.663	68.207	1	<.001
Relatedness	-.147	.066	-2.237	5.436	1	.020
Difficulty				21.564	2	.001
Moderate	-.240	.168	-1.423			
Difficult	-.752	.182	-4.134			
Age	.074	.153	.486	.218	1	.641
Relatedness:Difficulty				9.994	2	.007
Relatedness:Moderate	.020	.082	.249			
Relatedness:Difficult	.230	.076	3.030			

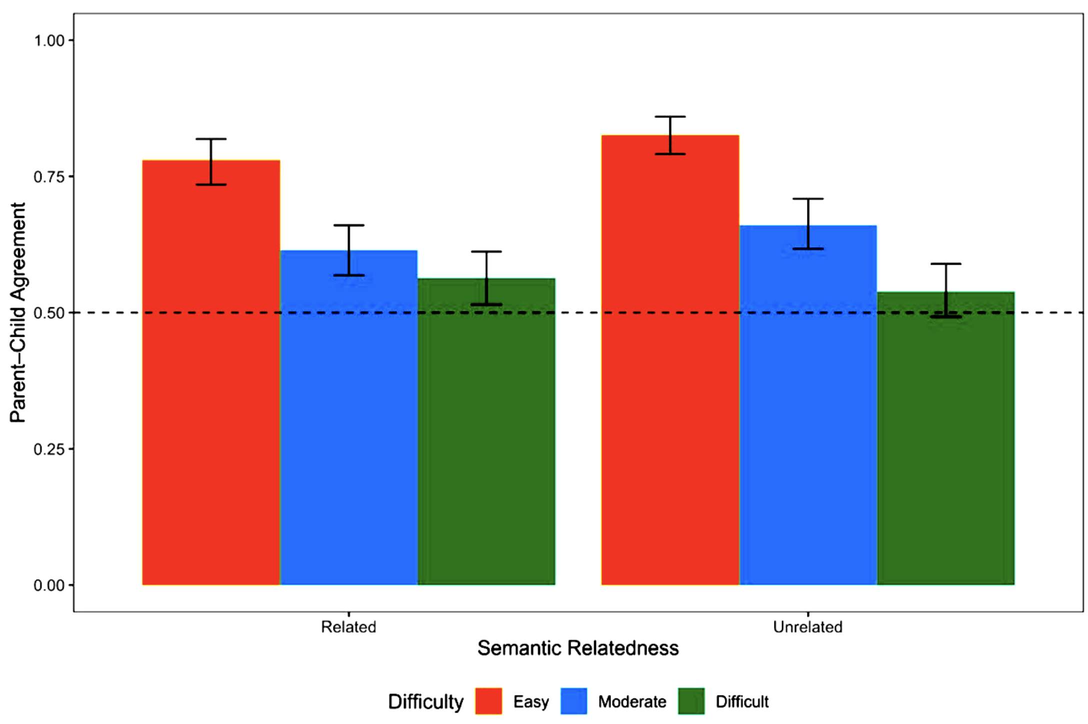


Figure 4-14: *Parent–child agreement by semantic relatedness and difficulty. Dashed line represents chance level at .5.*

recognition task, another binomial GLMM with a logit link function was fitted, with semantic relatedness, difficulty, item-pair comprehension status, age, and the interaction between semantic relatedness and difficulty as fixed effects. Semantic relatedness (-1: unrelated; 1: related), difficulty (-1: easy; 1: moderately difficult, difficult), and item-pair comprehension status (-1: both unknown; 1: both known, target known only, distractor known only) were sum-coded, whereas age was centred on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slope of difficulty[20]:

Accuracy ~ Relatedness ∗ Difficulty + Pair Comprehension + Age + (1 + Difficulty|Participant) + (1 | Object)

The results are detailed in Table 4.13, with chi-square statistics and p-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with p-values adjusted using the Tukey method.

As shown in Table 4.13, parent-reported item-pair comprehension was a significant predictor of toddlers' performance, along with semantic relatedness, difficulty, and age. No significant interaction effect between semantic relatedness and difficulty was found. Results from the follow-up tests indicated that toddlers were significantly less accurate when both target and distractor were reported as unknown compared to when both were known (β = -.628, *SE* = .190, z = -3.300, p = .005) and when only the target was known (β = -.769, *SE* = .196, z = -3.923, $p < .001$). No significant differences were found in other cases: (a) both known and target known only (β = -.141, *SE* = .195, z = -.725, p = .887); (b) both known and distractor known only (β = -.284, *SE* = .184, z = 1.539, p = .414); (c) target known only and distractor known only (β = .425, *SE* = .205, z = 2.070, p = .163); (d) distractor known only and both unknown (β = -.344, *SE* = .186, z = 1.846, p = .252; see also Figure 4-15).

4.4.4 Discussion

The current study examined the viability of using a tablet-based 2AFC word recognition task in assessing early word comprehension, with the aim of serving as a convergent and supplemental measure of parental reports. Toddlers aged between 18- and 20-month-old

[20]The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

Table 4.13. GLMM results for accuracy (with parent-reported item-pair comprehension as predictor)

	Model summary			Model comparison		
	β	*SE*	*z*	χ^2	*df*	*p*
Intercept	1.402	.144	9.749	58.245	1	<.001
Relatedness	-.139	.054	-2.588	6.586	1	.010
Difficulty				14.702	2	.001
Moderate	.007	.098	.068			
Difficult	-.403	.107	-3.776			
Pair comprehension				18.108	1	.001
Both known	.193	.114	1.685			
Target known	.334	.125	2.667			
Distractor known	-.091	.117	-.778			
Age	.511	.181	2.817	7.428	1	.006
Relatedness:Difficulty				4.141	2	.126
Relatedness:Moderate	-.120	.076	-1.581			
Relatedness:Difficult	.132	.072	1.832			

were examined on their comprehension of 24 lexical items selected from the Norwegian CDI–WG (Simonsen et al., 2014), either in the laboratory setting by an experimenter or online (and at home) by their parents. During the task, toddlers were asked to identify the target that was presented alongside a distractor. Each lexical target was paired twice, once with a semantically relevant distractor and once with a semantically unrelated distractor. Word pairs were classified into three difficulty levels (determined using Norwegian CDI–WG normative data for age-matched toddlers).

Toddlers were equally motivated to produce a response in the task and neither setting resulted in better or worse performance, as indicated by a lack of significant differences in terms of the number of attempted trials (regardless of whether the response was correct or incorrect) and the number of trials in which toddlers provided a correct response. This illustrates that remote infant data collection with fully automatized tasks can be as efficient and reliable as in situ laboratory assessments. Remote administration of high-quality data is not only a critical enabler during the global COVID-19 pandemic, but it also offers a promising avenue for data collection in developmental research, with higher efficiency, lower cost, and the potential for improved sample diversity by reaching a wider socio-demographic background than traditional laboratory-based methods (Sheskin et al., 2020).

With regards to the role of semantic relatedness, the findings indicated that toddlers displayed more robust recognition in semantically unrelated than related conditions, indicating that semantic relatedness between the target and the distractor triggered competition effects in referent selection, similar to research in younger infants (Bergelson & Aslin, 2017a). In the current study, lower recognition on some related trials could be due to an increased burden of visual discrimination and feature overlap (e.g., both goose and owl are birds with wings, feather, and a beak), as observed in Arias-Trejo and Plunkett (2010)'s

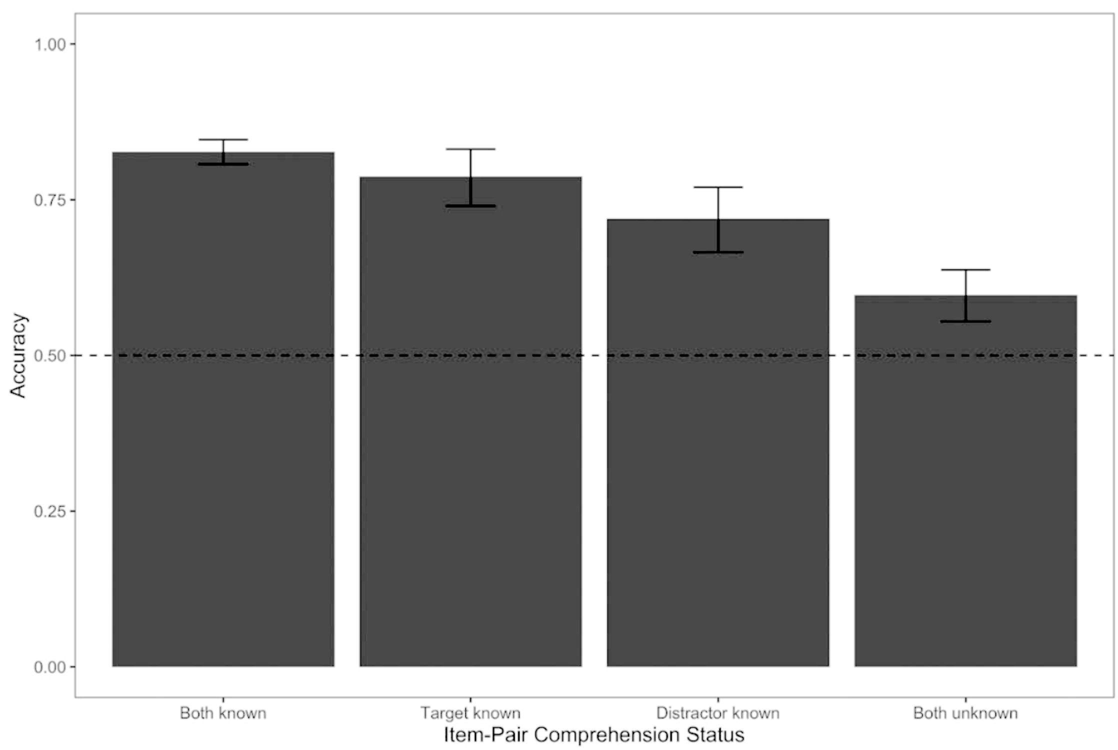


Figure 4-15: *Accuracy by parent-reported item-pair comprehension status. Dashed line represents chance level at .5.*

study with 18–24-month-olds. It is likely that, when toddlers heard the lexical target, they co-activated related (and thus, competing) word referents, which subsequently interfered with their lexical decision about the target. Even older children, between the ages of 3 and 9 years, have been observed to take longer time to provide a correct response in a visual search task when a related distractor was present compared to when an unrelated distractor was present (Vales & Fisher, 2019).

The overall viability of the tool is in line with Friend and Keplinger (2008), in which toddlers attempted significantly more easy than difficult trials. The findings suggest that toddlers were responding non-randomly and that non-responses represent toddlers' true inability to map the lexical target to its referent, rather than non-compliance or lack of motivation, while incorrect responses could be interpreted as evidence of partial word knowledge, and correct responses as evidence of robust word knowledge (Hendrickson et al., 2015). In terms of accuracy measure, again, congruent with previous work (Friend & Keplinger, 2003, 2008), toddlers' performance was consistent with the a priori, age-matched, difficulty categorization, as their best performance was obtained for easy trials and their worst performance for difficult trials. As would be expected, older toddlers also performed with greater accuracy relative to younger toddlers.

The acceptable convergent validity of the word recognition task was demonstrated by the findings which showed a significant and moderate correlation between toddlers' receptive vocabulary sizes (as measured by the CDI–WG) and their recognition accuracy (comparable to that achieved with the CCT; Friend & Keplinger, 2008), and also supported the feasibility of the CDI–WG as a general proxy for receptive vocabulary.

The current findings found moderately good item-level agreement between toddlers' responses and parental reports across both semantic conditions, with easy items having the highest agreement and difficult items having the lowest agreement, consistent with the CCT (Friend et al., 2012; Friend & Zesiger, 2011). In addition, parent-reported item-pair comprehension (i.e., whether the target or distractor label was known or not known by the child) was found to be a significant predictor of toddler recognition accuracy. Specifically, compared to trials where both the target and distractor were reported by parents as "not understood" on the CDI–WG, toddlers were more likely to respond correctly in trials where either the target or both the target and distractor were reported as "understood", indicating that parents are adequate informants of their child's language abilities.

In summary, the consistency between toddler's performance and the a priori word difficulty categorization, as well as the good item-level agreement between parental reports and their children's performance, provide encouraging results, even when the present study focused solely on nouns, unlike CCT which consists of an equal representation of nouns, verbs, and adjectives.

4.5 Conclusions

In this chapter we have introduced and discussed two novel computerised tools with the aim of improving early vocabulary acquisition assessment using CDI.

Study 4 described the MM-IRT model, a computational approach for creating short versions of CDIs. The approach administers CDIs in an IRT-based task (Makransky et al., 2016) and estimates full CDI scores using Mayor and Mani (2018)'s computational algorithm. Results from this study showed that a correlation of .95 with the full CDI scores can be achieved with as few as 15 test items and demonstrated high reliability even when the data have relatively small sample sizes (e.g., Italian, which had around 50 samples for each age in month group) in Wordbank (Frank et al., 2017). Study 5 examined the generalisation of the MM-IRT model by comparing the performance of the model with the Malaysian data. Results from this study showed a correlation of .95 with the full CDI scores with 100 test items, even when compensation was made due to low sample size – the model was deployed at intervals of 3 months in age, rather than each month as in Study 4. The findings also suggested that the MM-IRT generalises adequately to new data, as demonstrated by similar performances across training and test sets.

Study 6 examined the viability of using tablets to assess young children's word comprehension using the two-alternative forced-choice. In general, the current data suggests that a tablet-based word recognition task can be a useful measure of word comprehension in the second year of life and serve as a supplemental and convergent measure of parent reports.

In the next chapter – the concluding chapter for the thesis, I will discuss the main findings and their implications towards the development of a mature language assessment tool that works for the multicultural and multilingual context of Malaysia.

Chapter 5

General Discussion

In the preceding chapters, I reported the first Malaysian early vocabulary data collected using a trilingual adaptation of CDIs. The data provided insights into Malaysian children's early language development by exploring the roles of internal and external factors (Chapter 3). In Chapter 4, I introduced two novel computerised assessment tools with the aim of improving the administration of CDIs, as well as supplementing them with toddler-directed tasks. In this final chapter, I summarise the main findings of both chapters and discuss the primary contributions of this thesis in light of the objectives laid out at the beginning of this thesis. I conclude the thesis by discussing potential directions for future research.

5.1 Summary of Main Findings

5.1.1 Early Vocabulary Acquisition Measured using Trilingual CDIs

In Chapter 3 of this thesis, I examined the early vocabulary acquisition of young Malaysian children, and the effects of environmental factors on the language development of these children, using a trilingual adaptation of the CDIs – MCDI-M (Low, 2010).

Examining the overall early vocabulary acquisition trajectory of young Malaysian children, Study 1 demonstrated the utility of MCDI-M in capturing the growth of word knowledge among young children and was able to capture gender differences (i.e., a girl advantage) in vocabulary production. The linguistically diverse aspect of the Malaysian version of CDI also enabled the investigation of the role of relative language exposures on children vocabulary acquisition. The results suggested that higher relative exposure to a language resulted in larger vocabulary size in that language than in the other. In parallel, the availability of information regarding SES from the sample collected online enabled an investigation into the effects of SES on relative language exposure and on early vocabulary sizes, as reported in Study 2. In contrast to other bilingualism studies in English-dominant countries, Study 2 focused on both languages (Malay and English for Malay bilinguals, or Mandarin and English for Chinese bilinguals) and for both comprehension and production. The findings suggested that parents' effort in using their ethnic language (Malay for Malay-ethnic and Mandarin for Chinese-ethnic groups) is effective in fostering their children's word acquisition of the language across all income levels. In contrast, children from high-income families were more advanced when it comes to vocabulary development in English than children from low-income families.

Based on these early findings suggesting that MCDI-M is an efficient tool to assess early vocabulary development in a multilingual setting, the next study used MCDI-M as a tool to inform about theories of early language acquisition.

5.1.2 Modulation of Language Compositions of Verbs and Nouns

Early vocabularies typically contain more nouns than verbs. Yet, the strength of this noun-bias varies across languages and cultures. Two main theories have aimed at explaining such variations; either that the relative importance of nouns vs. verbs is specific to the language itself, or that extra-linguistic factors shape early vocabulary structures. Study 3 looked into the potential role of parents' attentional patterns and the linguistic properties of the language in modulating children language composition, in particular verbs and nouns. Cultural differences in attentional structures and their influence on both parental attention patterns and children's own attention bias, may, as a result, affects children's early vocabulary acquisition at the noun bias in particular. In contrast, language-specific factors such as morphological complexity, pronoun-dropping parameters, and word order may also play a significant role in modulating the learning of verbs and nouns. Yet, previous studies typically cannot tell apart both explanations, as culture is confounded with language use, e.g., examining Mandarin vocabulary in Chinese Mandarin learners and English vocabulary in Western Caucasian English learners. The study described in Chapter 3.5, however, looked in particular at the English vocabulary composition among Chinese learners, and among Malay learners. The evidence from our study suggested that the English of Mandarin-English bilinguals contains a higher proportion of verbs than the English of Malay-English bilinguals. This modulation of the noun bias can not be attributed to language-intrinsic factors, as the focus is on the same language — English. We argue that culture influences the pattern of acquisition of verbs and nouns in young children, possibly through differential attentional structures in their learning environment.

While MCDI-M have shown useful to document early vocabulary acquisition and helped address the noun bias debates, MCDI-M suffer from two main short-comings, first, they are lengthy and take substantial time for parents to fill them in, thereby limiting their use, e.g., as a universal screening tool. Second, vocabulary is indirectly addressed, via parental reports, which may introduce reporting bias. The next chapter will aim at addressing both shortcomings; by introducing a principled way of shortening CDIs – via the application of computerised adaptive test; and then to introduce a vocabulary assessment tool that is directly responded by toddlers.

5.1.3 Novel Approaches in Early Vocabulary Assessments based on CDIs

Chapter 4 introduced two novel approaches to improve early vocabulary assessment; first a language-general, novel computational framework that greatly shorten assessment time[1], and second, a tablet-based comprehension assessment tool directly attended by toddlers to assess early vocabulary comprehension[2].

Study 4 introduced a Bayesian-inspired item response theory-based framework developed based on prior work by Mayor and Mani (2018) and Makransky et al. (2016) (MM-IRT) in order to reduce the number of items needed for language assessment. The findings

[1] Study 4 and 5.
[2] Study 6.

indicated that the validity and reliability of the MM-IRT model relied upon the availability of sufficient amount of CDI data from children with matching key demographics (e.g., language, age, and gender). Based on the findings, even with a small dataset having fewer than 50 samples (CDI administrations) for each age group in months, the MM-IRT model was able to reliably estimate children's full CDI scores with just 25 items, effectively reduced the number of items by a factor of 4, relative to the 100-item short forms of CDIs. In addition, the MM-IRT model was implemented on the MCDI-M samples to evaluate its effectiveness in the context of non-monolingually exposed children in Malaysia, which was reported in Study 5. With the application of the MM-IRT model, a substantial improvement in the MCDI-M was observed, in which the total number of words can be reduced from 1,800 in the trilingual full-CDI form to a mere 300 words (e.g., 100 conceptual items in trilingual version) in the shorter version. Lastly, Study 6 examined a tablet-based direct-measure version of CDIs that assesses toddlers' word comprehension in word recognition tasks using the two-alternative-forced-choice (2AFC) format. The results showed that toddlers (aged between 18-and 20-month-old) attempted more frequently easy than difficult trials, thus indicating their responses were non-random and reflected their understanding of the referents. Importantly, toddlers' performance was consistent with the a priori difficulty level of the items, generally correlating with previous research using the Computerized Comprehension Task (CCT; Friend & Keplinger, 2003, 2008), thus demonstrating the viability of the task in assessing toddlers' word comprehension knowledge. In addition, significant and moderate correlations between toddlers' recognition accuracy and vocabulary sizes reported by their parents, as well as good item-level agreement between parental reports and their children's performance, demonstrated acceptable convergence validity of the word recognition task, and also supported the CDI-WG as a proxy for receptive vocabulary.

5.2 Implications

This research was motivated by the need to establish an early vocabulary assessment for linguistically diverse countries such as Malaysia, as well as to introduce novel early vocabulary assessment tools that improve and build upon existing tools.

Chapter 3 illustrated the first step of early vocabulary assessment using a standardised tool that is specific to Malaysian children. It is noteworthy that, while the variables involved in Study 1, i.e., age, language exposure, gender, and language type, all played a significant role in early vocabulary acquisition, they are not an exhaustive list of factors influencing this process, as evidenced by the moderate marginal variance in the models (see 3.2.2.1 for comprehension and 3.3.2.2 for production). Children's early language development is modulated by a variety of factors, such as biological, psychological, and social factors. It is beyond the scope of this thesis to provide an exhaustive assessment of the development of Malaysian children's language. Hence, the chapter focused on key factors which have been examined in the various subsamples of Malaysian children. While the MCDI-M captured vocabulary growth and the growth rate is modulated by language exposure (in comprehension and production), it is noteworthy that the data collected is cross-sectional, not longitudinal, and any observed trend results from a collective of individuals. At the individual level, relative language exposure could change in the first few years of life, especially when the child begins schooling (Dixon, Zhao, Quiroz, & Shin, 2012), as demonstrated in a longitudinal study (Welsh & Hoff, 2020). As a result, any change in language exposure might affect vocabulary growth in each of the languages bilingual and multilingual children acquire.

Several studies have shown that other external factors, such as socio-economic status (to name a few, Blanden & Machin, 2010; Layzer & Price, 2008; Hoff et al., 2018; Wanless et al., 2011) influence early vocabulary acquisition. However, the factors involved in mediating the disparities in vocabulary sizes across SES are often understudied in particular with bilingual and multilingual children. In the current Malaysian study, mediation analyses were additionally conducted to examine the mediating role of relative language exposure. The mediation analysis provided important insights. First, a significant role of language exposure was found for English – parents from high-income families tend to use more English with their children, which results in their children building larger English vocabulary sizes. Second, the amount of ethnic language exposure provided by parents did not contribute to the vocabulary disparities as observed in English – parent's lesser use of ethnic language with their children did not result in smaller vocabulary sizes in that language compared to other children from lower income families who received more ethnic language exposure. These contrasting results reflect the linguistically diverse aspect unique to Malaysian culture – the acquisition of English as a *lingua franca* and the acquisition of an ethnic language (EL; Malay for Malay-ethnic and Mandarin for Chinese-ethnic) being the dominant language used in their community. Hence, the lack of effect of parental ethnic language exposure highlighted other language input involved in facilitating their child's vocabulary acquisition. The ecological validity of the current study is constrained in that it did not examine other sources of language input, namely quantity and quality of language input at home and outside the home. Data concerning language proficiency of parents and other people interacting with the child, such as grandparents, day care staff, and peers, in addition to other aspects of language input, such as access to books, reading activities, and access to media content, would help describe children's language experiences in a fuller picture. Nevertheless, the gap observed in English vocabulary provided a glimpse into disparities in children's vocabulary development that are associated with lower SES.

As mentioned earlier in this thesis (Section 2.3.6), the Ministry of Education Malaysia aims to improve the Malay and English language proficiency of children nationwide, as documented in the Malaysian Education Blueprint 2013-2025 (Ministry of Education Malaysia, 2013). To achieve this goal, special attention has to be paid to children from low-income families, since their parents are less capable of providing a conducive environment for the acquisition of English vocabulary, as demonstrated in the current study. Public Early Childhood Care & Education (ECCE) programmes are therefore essential for low-income families, as children from these families are less likely to attend preschool (Ministry of Education Malaysia, 2013). Governmental funding in public ECCE programmes is therefore essential in providing affordable and effective bilingual education, which strives to improve the English development of children from low-income families while maintaining the acquisition of EL. As a result, perhaps children from low-income families will be able to enter primary school on an equal footing with their more affluent peers.

On the other hand, thee investigation of bilingual children's vocabulary composition in Study 3 provided valuable insights when it comes to understanding the factors driving a noun-bias in some languages[3], but not other languages[4]. Given the current study suggested that cultural differences, e.g., in the parents' attentional patterns may shape vocabulary composition of their children, early vocabulary assessments need to be mindful of these factors when using, e.g., noun knowledge as an index of language acquisition in a multilingual and multicultural setting.

For speech therapists, language assessment in a linguistically diverse setting is not an easy task. Apart from the assessment of children's word knowledge in their languages, there are additional factors that are influential in early vocabulary acquisition, such as the quantity and quality of language input, the linguistic properties of each language, and socio-cultural influences. CDIs in their digital form provide a promising avenue for low-cost and efficient data collection and cut down on time for data analysis. Besides, the work in Chapter 4 provided encouraging results when it comes to improving the efficiency of the original CDIs.

Study 4 introduced an innovative blend of item-response theory IRT; Makransky et al. (2016) and Mayor and Mani (2018)'s (MM) model. These methods have their own limitations when applied separately – while IRT provides a dynamic selection of words, the highest attainable scores are based on the maximum amount of test items rather than full-CDI scores, whereas, while the MM model is able to estimate full-CDI scores from a subset of items, the items are selected randomly. Combining these methods proved fruitful and led to more stable and better performance, and can even be used for very small sample sizes (around 50 per month of age). The shortening of the vocabulary forms thus enables more time for assessing other aspects, such as the child's language environment and language practice at home, to better understand the language exposure composition of the child. In parallel, Study 6 demonstrated that tablets can be used to generate a direct measure of early vocabulary comprehension via children's word recognition, which can supplement the use of parental reports. The use of direct-measure assessment tools circumvents the concerns of relying solely on parental reports, thereby providing a fuller picture of children's early

[3]For example: English (Gentner, 1982), Spanish (Jackson-Maldonado et al., 1993), Italian (Caselli & Casadio, 1995) and German (Kauschke, Lee, & Pae, 2007)

[4]For example: Korean (Choi & Gopnik, 1995; Kauschke et al., 2007; Arunachalam et al., 2013), Mandarin (Tardif, 1996; Tardif, Fletcher, Zhang, et al., 2008; Chen et al., 2015), Turkish (Kauschke et al., 2007) and Cantonese (Tardif, Fletcher, Zhang, et al., 2008)

language development. While only toddlers aged 18-to-20-month-old were assessed in the study, it is in principle possible to assess older children with proper selection of words based on a priori difficulty level on the basis of the existing normative sample. Furthermore, the (often laborious) initial word selection process could be avoided by computationally selecting words based on IRT, as demonstrated in Study 4 (and 5).

5.3 Limitations & Future Directions

With this work as a basis, three directions can be taken to move towards: 1) establishing a more comprehensive view of early vocabulary acquisition by exploring the role of family structure (e.g., number of siblings, siblings order, whether both parents work full-time, whether grandparents are involved in care-giving), the role of dialects, the role of the home environment (media use and overheard speech), and the role of communities (language use and overheard speech), to name a few; 2) conducting an empirical validation of the MM-IRT model on both monolingually- and non-monolingually-exposed children; 3) examining the viability of combining the MM-IRT model with the tablet-based word comprehension task to allow for an estimation of full CDI scores (Mayor & Mani, 2018) and total vocabulary sizes (Mayor & Plunkett, 2011).

5.3.1 Future of MCDI-M

By integrating the MM-IRT model into the MCDI-M forms, the length of vocabulary assessment will be considerably reduced, allowing more time for the inclusion of other language indexes into the forms. For instance, the original American English CDIs (Fenson et al., 2007) included checklists such as: phrases understood by the child, communicative actions and gestures used by the child, actions the child does or tries to do with various objects, pretend play and imitation. While the assessment of these items does not provide a measure of vocabulary size, it provides an overall picture of the children's current language and cognitive development. Given that the MCDI-M collects vocabulary data in more than one language, it is also essential to include additional information regarding children's language environment at home or outside the home.

5.3.2 Future of MM-IRT model

Despite the encouraging results from the validation of the MM-IRT model using real-data simulations, the current model suffers from some limitations.

Previously, neither Mayor and Mani (2018)'s model nor Makransky et al. (2016)'s IRT approaches had examined their application on bilingual or multilingually-exposed children. The combination of both approaches, i.e., the MM-IRT model introduced in Study 4 was also to be established for children in a linguistically diverse context in Study 5. Here, we demonstrated, for the first time, the effectiveness of the MM-IRT model on a sample collected from a linguistically diverse society, using a real-data simulation study. Preliminary results suggest that, despite aggregating month ages in 3-month intervals, a larger dataset is still needed for the model to perform well, if the aim is to have a very short test (e.g., less than 100 items). With the preliminary results, we believe that, with a larger sample as a basis to fit the model, the MM-IRT model could reach the desired correlation of .95 with full CDI administrations, a threshold typically adopted in these contexts. Future work could involve building a larger dataset than in the current work, by collecting further data, or as a first step, by stitching together paper and online samples and empirically validating the model on a new sample, in line with Mayor and Mani (2018), by comparing parents' reported full-CDI scores and short form scores, perhaps even supplementing with the tablet task described in Study 6, for a full validation. Once empirically validated, the MM-IRT could provide a rapid way of assessing children's early vocabulary knowledge and greatly save time for other batteries of tasks assessing other aspects of children's early development.

Second, the current model provides very limited information on the vocabulary composition of the children, since the dynamically chosen lists of test items are different for each child, also with no guarantee on the number of items selected in each semantic category. Yet, given that the estimation of full-CDI scores relies on fitting item-level responses, the possibility of reverse-engineering the model for item-level prediction of other untested items is worth exploring and is open for consideration.

Finally, the current application of IRT requires separate assessment of comprehension and production, thus requiring two sessions if both are needed, unlike other conventional CDIs which assess both vocabulary measures simultaneously. Future work may consider increasing the complexity of the IRT model to more than 2 parameters, in order to compute the item discrimination and difficulty levels for both production and comprehension, thus allowing the assessment of production and comprehension simultaneously.

5.3.3 Future Integration of the MM-IRT Model and the Tablet-based 2AFC in a Linguistically Diverse Setting

While the findings in Study 6 provided supportive evidence for a mobile, efficient way of assessing word comprehension using a tablet, the viability of using such a tool in a linguistically diverse setting is uncertain as yet. Several methodological and clinical concerns need to be addressed before assessing non-monolingually-exposed children. For example, when designing a counterbalanced empirical study (such as a tablet-assessment in Study 6) involving bilingually-exposed children, the order of the languages being assessed would need to be randomly assigned to avoid confounding errors due to practice effects or fatigue (depending on the actual intervals between tests). While counterbalancing is useful when examining early language development in a group, it could be misleading when used on an individual level, such as contributing toward the clinical diagnosis of a child who has been exposed to two languages. Perhaps, in a clinical environment, speech therapists could examine the early language development of a bilingually-exposed child with a focus on one language, ideally the child's dominant language (as reported by their parents), thus reducing misdiagnosis and avoiding the methodological concerns of measuring both languages. This suggestion, only applies to children with a clear dominant knowledge of one language – however constituting the majority of the current Malaysian sample (see Table 3.2). A different strategy may be needed for children who are exposed to two languages in a similar proportion, or who are reported to possess a similar level of knowledge in both languages.

Another concern is the availability of data. The a priori difficulty levels of the items used in Study 6 were determined based on normative data of toddlers aged between 18- and 20-month-old, that comprised 251 toddlers who were monolingually-exposed to Norwegian (Simonsen et al., 2014), whereas normative data in Friend and Zesiger (2011)'s study comprised 720 toddlers who were monolingually-exposed to American English (Fenson et al., 1994) and 84 samples were monolingually-exposed to Mexican Spanish (Jackson-Maldonado, Thal, & Fenson, 2003). In the Malaysian sample, among toddlers aged 16-month-old, 6 out of 31 were monolingually-exposed to Mandarin, 10 were monolingually-exposed to Malay and the rest were bilingually-exposed with varying exposures. Even among the older, 20-month-old toddlers, 13 out of 29 were monolingually-exposed to Malay, 1 was monolingually-exposed to Mandarin and the rest were bilingually exposed. As discussed in Chapter 1, 2 and 3, early vocabulary acquisition is characterised by high individual variability, which, when combined with the fact that the Malaysian data is linguistically diverse, significantly limits the amount of data we can use to represent children's comprehension (or production) at a given age, for each language.

5.4 Conclusion

In summary, this thesis contributed to the assessment of early vocabulary development, with a focus on linguistically diverse settings. The collection and analysis of the first Malaysian early vocabulary data using a trilingual adaptation of the CDI provided insight into the early vocabulary acquisition of infants and toddlers, and also provided evidence that the Malaysian CDI is effective in capturing the vocabulary developmental trajectory observed in the other CDI studies. Future work will aim at developing the next generation of language assessment that further improves the performance of the model, also integrating it with the toddler-directed tablet-based word recognition task and extending it to multilinguals. Future challenges remain about how early vocabulary assessment can be used to establish a strong evidence-based framework that best informs young children's language development, in particular in a country as culturally- and linguistically diverse as Malaysia. In this regard, it is critical that future work expands, improves and validates the assessment tools introduced in the present thesis in order to provide reliable and efficient assessments.

References

Acarlar, F., Aksu-Koç, A., Küntay, A. C., Maviş, İ., Sofu, H., Topbaş, S., & Turan, F. (2008). Adapting mb-cdi to turkish: The first phase. In *Essays of turkish linguistics: Proceedings of the 14th international conference on turkish linguistics* (pp. 6–8).

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child development*, *72*(2), 416–430.

Albury, N. J. (2017). Mother tongues and languaging in malaysia: Critical linguistics under critical examination. *Language in Society*, *46*(4), 567.

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of vocabulary testing in bilingual children. *Language, speech, and hearing services in schools*, *49*(1), 85–97.

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1-2), 63–80.

Armon-Lotem, S., & Ohana, O. (2016, May). A CDI study of bilingual english-hebrew children – frequency of exposure as a major source of variation. *International Journal of Bilingual Education and Bilingualism*, *20*(2), 201–217. Retrieved from `https://doi.org/10.1080/13670050.2016.1179257` doi: 10.1080/13670050.2016.1179257

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the macarthur communicative development inventory of children from low and middle-income families. *Applied Psycholinguistics*, *19*(2), 209–223.

Arunachalam, S., Leddon, E. M., Song, H.-j., Lee, Y., & Waxman, S. R. (2013). Doing more with less: Verb learning in korean-acquiring 24-month-olds. *Language acquisition*, *20*(4), 292–304.

ASHA. (2020). Early intervention – what is early intervention? Retrieved from `https://www.asha.org/public/speech/early-intervention/`

Astington, J. W., & Baird, J. A. (2005). *Why language matters for theory of mind*. Oxford University Press.

Azah, N., Aziz, A., Syuhada, N., Sin, M., Batmaz, F., Stone, R., ... Chung, H. (2014). Selection of touch gestures for children's applications: repeated experiment to increase reliability.

Baker, L., Silverstein, M., Arber, S., & Timonen, V. (2012). The well-being of grandparents caring for grandchildren in china and the united states. *Contemporary grandparents: Changing family relationships in global contexts*, 51–70.

Barbu, S., Cabanes, G., & Le Maner-Idrissi, G. (2011). Boys and girls on the playground:

sex differences in social development are not stable across early childhood. *Plos one*, *6*(1), e16407.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Barton, K., & Barton, M. K. (2015). Package 'mumin'. *Version*, *1*(18), 439.

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

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*(5-6), 507–584.

Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J., ... Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of child language*, *21*(1), 85–123.

Bauer, L., & Schanzenbach, D. W. (2016). The long-term impact of the head start program. *The Hamilton Project*.

Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring. *Language, Speech, and Hearing Services in Schools*, *36*(3), 188-200. doi: 10.1044/0161-1461(2005/020)

Bergelson, E., & Aslin, R. (2017b). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, *13*(4), 481–501.

Bergelson, E., & Aslin, R. N. (2017a). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Blanden, J., & Machin, S. (2010). Intergenerational inequality in early years assessments. *Children of the 21st century–the first five years*, 153–168.

Bleses, D., Vach, W., Jørgensen, R. N., & Worm, T. (2010). The internal validity and acceptability of the Danish SI-3: A language-screening instrument for 3-year-olds. *Journal of Speech, Language, and Hearing Research*.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008a). The danish communicative developmental inventories: validity and main developmental trends. *Journal of child language*, *35*(3), 651–669.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008b). The Danish Communicative Developmental Inventories: validity and main developmental trends. *Journal of child language*, *35*(3), 651–669.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008c). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of child language*, *35*(3), 619–650.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax* (Vol. 154). Walter de Gruyter.

Bloom, L. (1993). *The transition from infancy to language: Acquiring the power of expression*. Cambridge University Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, *46*(4), 443–459.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First language*, *24*(3), 267–304.

Bornstein, M. H., Haynes, M. O., & Painter, K. M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of child language*, *25*(2), 367–393.

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: a multi-age, multidomain, multimeasure, and multisource study. *Developmental psychology*, *48*(2), 477.

Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior and Development*, *22*(1), 65–85.

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, *46*(2-3), 217-243.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, *49*(4), 422–432.

Braginsky, M. (2018). wordbankr: Accessing the Wordbank database [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=wordbankr` (R package version 0.3.0)

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67.

Brese, F., & Mirazchiyski, P. (2013). Measuring students' family background in largescale international education studies. *Hamburg, Germany: IEA-ETS Research Institute*.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. Retrieved from `https://journal.r-project.org/archive/2017/RJ-2017-066/index.html`

Bushneil, I., Sai, F., & Mullin, J. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, *7*(1), 3–15.

Cadime, I., Silva, C., Ribeiro, I., & Viana, F. L. (2018). Early lexical development: Do day care attendance and maternal education matter? *First Language*, *38*(5), 503–519.

Carey, S. (1994). Does learning a language require the child to reconceptualize the world. *Lingua*, *92*(1-4), 143–167.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159–199.

Caselli, M. C., & Casadio, P. (1995). *Il primo vocabolario del bambino: Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita* (Vol. 5). Franco Angeli.

Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of child language*, *26*(1), 69–111.

Castro, D. C., Páez, M. M., Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child development perspectives*, *5*(1), 15–21.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to english is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International journal of language & communication disorders*, *49*(6), 649–671.

Cavallaro, F., & Ng, B. C. (2014). Language in singapore: From multilingualism to english plus. *Challenging the monolingual mindset*, *156*(33), 9781783092529–005.

CDI Advisory Board. (n.d.). Adaptations in other languages. Retrieved from `http://mb-cdi.stanford.edu/adaptations.html`

Chai, J. H., Low, H. M., Wong, T. P., Onnis, L., & Mayor, J. (2021). Extra-linguistic modulation of the english noun-bias: evidence from malaysian bilingual infants and toddlers. *Journal of Cultural Cognitive Science*, *5*(1), 49–64.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. doi: 10.18637/jss.v048.i06

Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software*, *71*(5), 1–39. doi: 10.18637/jss.v071.i05

Chan, & Nicoladis, E. (2010). Predicting two mandarin-english bilingual children's first 50 words: Effects of frequency and relative exposure in the input. *International Journal of Bilingualism*, *14*(2), 237–270.

Chan, Tardif, T., Chen, J., Pulverman, R. B., Zhu, L., & Meng, X. (2011). English-and chinese-learning infants map novel labels to objects and actions differently. *Developmental psychology*, *47*(5), 1459.

Chen, J., Tardif, T., Pulverman, R., Casasola, M., Zhu, L., Zheng, X., & Meng, X. (2015). English-and mandarin-learning infants' discrimination of actions and objects in dynamic events. *Developmental psychology*, *51*(10), 1501.

Cheung, K. C.-K., & Wong, R. K.-S. (2020). Income effects on the intellectual development of children from low-income families in hong kong: The mediating role of

parental investment and parental stress.
Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental psychology*, *38*(6), 967-978. doi: 10.1037/0012-1649.38.6.967
Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in korean: A cross-linguistic study. *Journal of child language*, *22*(3), 497–529.
Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child development*, *77*(3), 712–735.
Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in spanish–english bilinguals from 22 to 30 months: implications for assessment. *Journal of Speech, Language, and Hearing Research*.
Côté, S. L., Gonzalez, A. M., & Byers-Heinlein, K. (2020). Multilingual toddlers' vocabulary development in two languages: Comparing bilinguals and trilinguals.
Couse, L. J., & Chen, D. W. (2010). A tablet computer for young children? exploring its viability for early childhood education. *Journal of research on technology in education*, *43*(1), 75–96.
Croll, E. J. (2006). The intergenerational contract in the changing asian family. *Oxford Development Studies*, *34*(4), 473–491.
Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, *34*(3), 565–571.
Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*(3), 544–560.
Dale, P. S., Reznick, J. S., & Thal, D. J. (1998). A parent report measure of language development for three-year-olds. *Infant Behavior and Development*(21), 370.
David, A., & Wei, L. (2008). Individual differences in the lexical development of french–english bilingual children. *International Journal of Bilingual Education and Bilingualism*, *11*(5), 598–618.
Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, *19*(2), 294.
Dawson, G., & Bernier, R. (2013). A quarter century of progress on the early detection and treatment of autism spectrum disorder. *Development and psychopathology*, *25*(4pt2), 1455–1472.
De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35*(6), 1189–1211.
Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, *1*(3), 111–34.
Department of Statistics Malaysia, M. (2010). Education and social characteristics of the population, 2010. Retrieved from `https://www.dosm.gov.my/v1/index.php?r=column/cthree&menu_id=cEhBV0xzWll6WTRjdkJienhoR290QT09`

Department of Statistics Malaysia, M. (2017). Ict use and access by individuals and households survey report, malaysia, 2017. Retrieved from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=395&bul_id=bHBzbWxkWElxRDlmaU81Q3R2ckRkZz09&menu_id=amVoWU54UTl0a21NWmdhMjFMMWcyZz09

Department of Statistics Malaysia, M. (2020). Current population estimates, malaysia, 2020. Retrieved from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=OVByWjg5YkQ3MWF%20ZRTN5bDJiaEVhZz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International journal of language & communication disorders*, *43*(4), 361–389.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, *32*(4), 759–786.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child development*, *74*(2), 394–412.

Dixon, L. Q., Zhao, J., Quiroz, B. G., & Shin, J.-Y. (2012). Home and community factors influencing bilingual children's ethnic language vocabulary development. *International Journal of Bilingualism*, *16*(4), 541–565.

Dunn, D. (2018). Peabody picture vocabulary test fifth edition (ppvt-5). *Minneapolis, MN*.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists.

Eriksson, M. (2017). The swedish communicative development inventory iii: Parent reports on language in preschool children. *International journal of behavioral development*, *41*(5), 647–654.

Feldman, Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the macarthur communicative development inventories at ages one and two years. *Child development*, *71*(2), 310–322.

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., ... Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2007). *MacArthur–Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the macarthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95–116.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child development*, *83*(1), 203–222.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). Ses differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, *16*(2), 234–248.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98.

Fernald, A., Pinto, J., Swingley, D., Weinberg, A., McRoberts, G., Tomasello, M., & Bates, E. (2001). Rapid gains of verbal processing by infants in the 2nd year. *Language development: The essential readings*, 49–56.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological science*, *9*(3), 228–231.

Fischel, J. E., Whitehurst, G. J., Caulfield, M. B., & DeBaryshe, B. (1989, Feb). Language growth in children with expressive language delay. *Pediatrics*, *83*(2), 218–27.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of personality and social psychology*, *78*(2), 350.

Frank, M., Braginsky, M., Marchman, V. A., & Yurovsky, D. (2021). *Variability and consistency in early language learning: The wordbank project.* Cambridge, MA: MIT Press.

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 1-18. doi: 10.1017/S0305000916000209

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677–694.

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 302–309.

Friend, M., & Keplinger, M. (2008). Reliability and validity of the computerized comprehension task (cct): data from american english and mexican spanish infants. *Journal of child language*, *35*(1), 77–98.

Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the computerized comprehension task: comprehension predicts production. *Developmental psychology*, *48*(1), 136.

Friend, M., Smolak, E., Patrucco-Nanchen, T., Poulin-Dubois, D., & Zesiger, P. (2019). Language status at age 3: Group and individual prediction from vocabulary comprehension in the second year. *Developmental psychology*, *55*(1), 9.

Friend, M., & Zesiger, P. (2011). Une réplication systématique des propriétés psychométriques du computerized comprehension task dans trois langues. *Enfance*(3), 329–344.

Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., & Vigário, M. (2016). Infant communicative development assessed with the european portuguese macarthur–bates communicative development inventories short forms. *First Language*, *36*(5), 525–545.

Fuchs, D., & Fuchs, L. S. (2009). Responsiveness to intervention: Multilevel assessment and instruction as early intervention and disability identification. *The Reading*

Teacher, *63*(3), 250–252.
Galeote, M., Checa, E., Sánchez-Palacios, C., Sebastián, E., & Soto, P. (2016). Adaptation of the macarthur-bates communicative development inventories for spanish children with down syndrome: Validity and reliability data for vocabulary. *American journal of speech-language pathology*, *25*(3), 371–380.
Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science*, *3*(2), 206–215.
Garcia, I., Barrena, A., Ezeizabarrena, M. J., Almgren, M., Arratibel, N., & Barnes, J. (2014). Assessing the communicative development of 30 to 50-months old basque children: The basque version of the macarthur-bates cdi-iii. *Psikologia*, *88*, 33–72.
Genesee, F. (1989). Early bilingual development: one language or two? *Journal of child language*, *16*, 01.
Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, cognition, and culture* (p. 301-334).
Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*(3), 318–330.
Gill, S. K. (2004). Medium of instruction policy in higher education in malaysia: Nationalism versus internationalization. *Medium of instruction policies: Which agenda whose agenda*, 135–152.
Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176.
Goldfield, B. A. (1987). The contributions of child and caregiver to referential and expressive language. *Applied Psycholinguistics*, *8*(3), 267–280.
Golinkoff, R. M., & Kerr, J. L. (1978). Infants' perception of semantically defined action role changes in filmed events. *Merrill-Palmer Quarterly of Behavior and Development*, *24*(1), 53–61.
Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339.
Gopinathan, S., Ho, W. K., & Saravanan, V. (2004). Ethnicity management and language education policy: Towards a modified model of language education in singapore schools. *Beyond rituals and riots: Ethnic pluralism and social cohesion in Singapore*, 228–257.
Greenwood, C. R., Carta, J. J., Walker, D., Watson-Thompson, J., Gilkerson, J., Larson, A. L., & Schnitz, A. (2017). Conceptualizing a public health prevention intervention for bridging the 30 million word gap. *Clinical child and family psychology review*, *20*(1), 3–24.
Grilli, E., & Remondino, F. (2020). Machine learning generalisation across different 3d architectural heritage. *ISPRS International Journal of Geo-Information*, *9*(6), 379.
Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574–586.

Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13-and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217.

Hall, S. J. (2015). A past before a blueprint: Malaysia's challenges in english language teaching.

Hampson, J., & Nelson, K. (1993). The relation of maternal language to variation in rate and style of language acquisition. *Journal of child language*, *20*(2), 313–342.

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

Hashim, A. (2009). Not plain sailing: Malaysia's language choice in policy and education. *AILA Review*, *22*(1), 36–51.

Hendrickson, K., & Friend, M. (2013). Quantifying the relationship between infants' haptic and visual response to word-object pairings. In *Proceedings of the... annual boston university conference on language development. boston university conference on language development* (Vol. 37).

Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science*, *18*(5), 723–735.

Hendrickson, K., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2017). Assessing a continuum of lexical–semantic knowledge in the second year of life: A multimodal approach. *Journal of experimental child psychology*, *158*, 95–111.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.

Hoff, E. (2013). *Language development*. Cengage Learning.

Hoff, E., Burridge, A., Ribot, K. M., & Giguere, D. (2018). Language specificity in the relation of maternal education to bilingual children's vocabulary growth. *Developmental psychology*, *54*(6), 1011.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of child language*, *39*(1), 1.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child development*, *62*(4), 782–796.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*(4), 603–629.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*(4), 701–724.

Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in spanish–english bilingual toddlers. *Bilingualism: Language and Cognition*, *17*(1), 189–202.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? links between maternal talk, processing speed and vocabulary size in spanish-learning children. *Developmental science*, *11*(6), F31–F39.

Hyde, J. S. (1981). How large are cognitive gender differences? a meta-analysis using! w^2 and d.. *American psychologist*, *36*(8), 892.

Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. (2013). Short-form versions of the spanish macarthur–bates communicative development inventories. *Applied Psycholinguistics*, *34*(4), 837–868.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in spanish-speaking infants and toddlers. *Journal of child language*, *20*(3), 523-49.

Jackson-Maldonado, D., Thal, D. J., & Fenson, L. (2003). *Macarthur inventarios del desarrollo de habilidades comunicativas: User's guide and technical manual*. Brookes Pub.

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into english as a lingua franca. *Language teaching*, *44*(3), 281–315.

Jerger, S., & Damian, M. F. (2005). What's in a name? typicality and relatedness effects in children. *Journal of Experimental Child Psychology*, *92*(1), 46–75.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of Personality and Social Psychological psychology*, *78*(5), 943-955. doi: 10.1037/0022-3514.78.5.943

Johnson, C. J., Beitchman, J. H., & Brownlie, E. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American journal of speech-language pathology*.

Kachergis, G., Marchman, V. A., Dale, P., Mankewitz, J., & Frank, M. C. (2021). Online computerized adaptive tests (cat) of children's vocabulary development in english and mexican spanish.

Kalajdzic, P., Krüger, L., & Venli, V. (2020, Mar). *Alle utdanningsinstitusjoner stenges – flere arrangementer og virksomheter får forbud.* NRK. Retrieved from https://www.nrk.no/norge/alle-utdanningsinstitusjoner-stenges-_-flere-arrangementer-og-virksomheter-far-forbud-1.14940952

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N., Almohammadi, A., … others (2021). Covid-19 first lockdown as a unique window into language acquisition: What you do (with your child) matters.

Kartushina, N., & Mayor, J. (2019). Word knowledge in six-to nine-month-old norwegian infants? not without additional frequency cues. *Royal Society open science*, *6*(9), 180711.

Kauschke, C., Lee, H., & Pae, S. (2007). Similarities and variation in noun and verb acquisition: A crosslinguistic study of children learning german, korean, and turkish. *Language and Cognitive Processes*, *22*(7), 1045-1072. doi: http://dx.doi.org/10.1080/01690960701307348

Keating, D. P., & Hertzman, C. (2000). *Developmental health and the wealth of nations: Social, biological, and educational dynamics*. Guilford Press.

Kern, S. (2003). Le compte-rendu parental au service de l'évaluation de la production lexicale des enfants français entre 16 et 30 mois. *Glossa*(85), 48–62.

Knight, S. (2013). *Forest school and outdoor learning in the early years*. Sage.

Kristoffersen, K. E., Bleses, D., Wehberg, S., Jørgensen, R. N., Eiesland, E. A., Henriksen, L. Y., & Simonsen, H. G. (2013). The use of the internet in collecting cdi data-an

example from norway. *Journal of child language*, *40*(3), 567–585.
Kutsuki, A. (2021). Do bilinguals acquire similar words to monolinguals? an examination of word acquisition and the similarity effect in japanese–english bilinguals' vocabularies. *European Journal of Investigation in Health, Psychology and Education*, *11*(1), 168–182.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).
Lavin, T. A., Hall, D. G., & Waxman, S. R. (2006). East and west: A role for culture in the acquisition of nouns and verbs. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs (p. 525).* Oxford: Oxford University Press.
Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health*, *13*(4), 198–206.
Layzer, J., & Price, C. (2008). Closing the gap in the school readiness of low-income children. *Washington, DC: US Department of Health and Human Services*.
Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. *Developmental psychology*, *40*(6), 993.
Lee, S., & Shanmuganathan, T. (2020). Reconceptualizing aunty as an address term in urban multilingual malaysia. *World Englishes*, *39*(1), 198–213.
Lee, S. K., Lee, K. S., Wong, F. F., & Ya'acob, A. (2010). The english language and its impact on identities of multilingual malaysian undergraduates. *GEMA Online® Journal of Language Studies*, *10*(1).
Lee, T. Y., Yu, J., & Nah, S. (2011). A confucianism observed in disaster films of east asia. *International Journal of Humanities and Social Sciences*, *5*(5), 385–393.
Lee, W., & Pring, T. (2016). Supporting language in schools: Evaluating an intervention for children with delayed language in the early school years. *Child Language Teaching and Therapy*, *32*(2), 135–146.
Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means (version 1.5. 2-1)[r package].*
Lim, H. W., Wells, B., & Howard, S. (2015). Rate of multilingual phonological acquisition: Evidence from a cross-sectional study of english–mandarin–malay. *Clinical linguistics & phonetics*, *29*(11), 793–811.
Lim, R. (2021). Door-to-door drive to complete census 2020. *The Star*. Retrieved 2021-07-28, from `https://www.thestar.com.my/news/nation/2021/04/22/door-to-door-drive-to-complete-census-2020`
Lo, Rosslund, A., Chai, J. H., Mayor, J., & Kartushina, N. (2021). Tablet assessment of word comprehension reveals coarse word representations in 18-20-month-old toddlers. *Infancy*.
Lo, C. H., Mani, N., Kartushina, N., Mayor, J., & Hermes, J. (2021). e-babylab: An open-source browser-based tool for unmoderated online developmental studies.
Low, H. M. (2009). *Three case studies: Early language development in mixed-language contexts of childhood multilingualism* (Unpublished doctoral dissertation). La Trobe University.

Low, H. M. (2010). Trilingual macarthur-bates communicative developmental inventories (malaysia version): An investigation of early language development in mixed-language contexts of childhood multilingualism. *Paper presented in Asia Pacific Conference of Speech, Language and Hearing 2010, Kuala Lumpur, Malaysia*.

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772.

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur–Bates Communicative Development Inventory (CDI). *Journal of Child Language*, *34*(3), 623–654.

Mack, F. R., Smith, V. G., & Straight, H. (2010). Response to intervention: Implications for the proficiency of early childhood special educators. *Journal of the International Association of Special Education*, *11*(1).

Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory–based, computerized adaptive testing version of the macarthur–bates communicative development inventory: Words & sentences (cdi: Ws). *Journal of Speech, Language, and Hearing Research*, *59*(2), 281–289.

Mancilla-Martinez, J., Pan, B. A., & Vagh, S. B. (2011). Assessing the productive vocabulary of spanish–english bilingual toddlers from low-income families. *Applied Psycholinguistics*, *32*(2), 333–357.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), F9–F16.

Marchman, V. A., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both english and spanish. *Journal of Speech, Language, and Hearing Research*.

Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, *7*(2), 212–224.

Marjanovič-Umek, L., Fekonja-Peklaj, U., & Podlesek, A. (2013). Characteristics of early vocabulary and grammar development in Slovenian-speaking infants and toddlers: a CDI-adaptation study. *Journal of child language*, *40*(4), 779–798.

Markova, G., & Smolík, F. (2014). What do you think? the relationship between person reference and communication about the mind in toddlers. *Social Development*, *23*(1), 61–79.

Marsh, P. L. Y.-R. D. B. J. L. J. S. F. D. A. D. S. F. K. P. M. R. P. T. S. . W. P., J. (2015). *Exploring play and creativity in pre-schooler's use of apps: Final project report.* Retrieved from `http://www.techandplay.org/reports/TAP_Final_Report.pdf`

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of memory and language*, *94*, 305–315.

Mayne, A. M., Yoshinaga-Itano, C., & Sedey, A. L. (1999). Receptive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta Review*, *100*, 29–52.

Mayne, A. M., Yoshinaga-Itano, C., Sedey, A. L., & Carey, A. (1998). Expressive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta*

Review, *100*, 1–28.
Mayor, J., & Mani, N. (2018). A short version of the macarthur–bates communicative development inventories with high validity. *Behavior research methods*, 1–8.
Mayor, J., & Plunkett, K. (2010). Vocabulary spurt: Are infants full of zipf? In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 836–841).
Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from cdi analysis. *Developmental Science*, *14*(4), 769–785.
Mayor, J., & Plunkett, K. (2014). Shared understanding and idiosyncratic expression in early vocabularies. *Developmental Science*, *17*(3), 412–423.
McWayne, C. M., Cheung, K., Wright, L. E. G., & Hahs-Vaughn, D. L. (2012). Patterns of school readiness among head start children: Meaningful within-group variability during the transition to kindergarten. *Journal of educational psychology*, *104*(3), 862.
Meisel, J. M. (1989). Early differentiation of languages in bilingual children. *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*, 13-40.
Meisel, J. M. (2001). The simultaneous acquisition of two first languages. *Trends in bilingual acquisition*, *1*, 11.
Michael, F. (2017). vcdextra: 'vcd' extensions and additions [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=vcdExtra` (R package version 0.7-1)
Ministry of Education Malaysia, M. (2013). Malaysia education blueprint 2013-2025 (preschool to post-secondary education). *Putrajaya: Kementerian Pendidikan Malaysia*.
Ministry of Education Malaysia, M. (2020). Quick facts 2020 malaysia educational statistics. Retrieved from `https://www.moe.gov.my/muat-turun/penerbitan-dan-jurnal/terbitan/buku-informasi/3719-quick-facts-2020/file`
Mitchell, T. M., et al. (1997). Machine learning.
Muggeo, V., Sciandra, M., Tomasello, A., & Calvo, S. (2013). Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environmental and Ecological Statistics*, *20*, 519–531.
Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? effects of input frequency and structure on children's early verb use. *Journal of child language*, *25*(1), 95–120.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, *4*(2), 133–142.
Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1–135.
Nisbett, R. E. (2004). *The geography of thought: How asians and westerners think differently...and why*. New York: Free Press.
Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467-473. doi: 10.1016/j.tics.2005.08.004
Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E. (2014). Language assessment in children with autism spectrum disorder: Concurrent validity between report-based assess-

ments and direct tests. *Research in Autism Spectrum Disorders*, *8*(9), 1100–1106.
Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children* (Vol. 2). Multilingual Matters.
Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019). Reliability of the language environment analysis recording system in analyzing french–english bilingual speech. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2491–2500.
Pae, S., & Kwak, K. (2011). Korean macarthur-bates communicative development inventories (k mb cdi). *Seoul: Mindpress*.
Pan, B. A., Rowe, M. L., Spier, E., & Tamis-Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of child language*, *31*(3), 587–608.
Park, Y., Gurel, S., Oh, J., Bettini, E. A., & Leite, W. (2015). Literacy-related school readiness skills of english language learners in head start: An analysis of the school readiness survey. *Journal of Early Childhood Research*, *13*(3), 280–294.
Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in spanish–english bilingually developing 2-year-olds. *Journal of experimental child psychology*, *108*(1), 113–125.
Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied psycholinguistics*, *18*(1), 41–58.
Pellegrini, A. D., Brody, G. H., & Sigel, I. E. (1985). Parents' book-reading habits with their children. *Journal of Educational Psychology*, *77*(3), 332.
Pérez-Pereira, M., & Resches, M. (2011). Concurrent and predictive validity of the galician cdi. *Journal of Child Language*, *38*(1), 121–140.
Persici, V., Vihman, M., Burro, R., & Majorano, M. (2019). Lexical access and competition in bilingual children: The role of proficiency and the lexical similarity of the two languages. *Journal of experimental child psychology*, *179*, 103–125.
Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child development*, *82*(6), 1834–1849.
Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, *17*(1), 57–70.
Preston, J. L., Frost, S. J., Mencl, W. E., Fulbright, R. K., Landi, N., Grigorenko, E., ... Pugh, K. R. (2010). Early and late talkers: school-age language, literacy and neurolinguistic differences. *Brain*, *133*(8), 2185–2195.
R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`
Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American psychologist*, *53*(2), 109.
Redlinger, W. E., & Park, T.-Z. (1980). Language mixing in young bilinguals. *Journal of child language*, *7*, 02.
Rescorla, L. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing disorders*, *54*(4), 587–599.

Rescorla, L., & Dale, P. S. (2013). *Late talkers: Language development, interventions, and outcomes*. Baltimore, MD: Brookes.

Rescorla, L., Ratner, N. B., Jusczyk, P., & Jusczyk, A. M. (2005). Concurrent validity of the Language Development Survey: Associations with the MacArthur—-Bates Communicative Development Inventories. *American Journal of Speech-Language Pathology*.

Rinaldi, P., Pasqualetti, P., Stefanini, S., Bello, A., & Caselli, M. C. (2019). The Italian Words and Sentences MB-CDI: Normative data and concordance between complete and short forms. *Journal of child language*, *46*(3), 546–566.

Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*, *20*(59), 141–159.

Rojas, R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., & Reese, L. (2016). Interlocutor differential effects on the expressive language skills of spanish-speaking english learners. *International journal of speech-language pathology*, *18*(2), 166–177.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and acquisition of language* (pp. 111–144). Elsevier.

Rowe, M. L. (2008, January). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. Retrieved from `https://doi.org/10.1017/s0305000907008343` doi: 10.1017/s0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, *83*(5), 1762–1774.

Sameroff, A. (1975). Transactional models in early social relations. *Human development*, *18*(1-2), 65–79.

Saravanan, V. (2001). The significance of bilingual chinese, malay and tamil children's english network patterns on community language use patterns. *Early Child Development and Care*, *166*(1), 81–91.

Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Development*, *76*(1), 87–96.

Schick, B., De Villiers, P., De Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child development*, *78*(2), 376–396.

Schneider, R., Yurovsky, D., & Frank, M. (2015). Large-scale investigations of variability in children's first words. In *Cogsci* (pp. 2110–2115).

Schults, A., & Tulviste, T. (2016). Composition of estonian infants' expressive lexicon according to the adaptation of cdi/words and gestures. *First Language*, *36*(5), 485–504.

Serrat-Sellabona, E., Aguilar-Mediavilla, E., Sanz-Torrent, M., Andreu, L., Amadó, A., & Serra, M. (2021). Sociodemographic and pre-linguistic factors in early vocabulary acquisition. *Children*, *8*(3), 206.

Setoh, P., Cheng, M., Bornstein, M. H., & Esposito, G. (2021). Contrasting lexical biases in bilingual english–mandarin speech: Verb-biased mothers, but noun-biased toddlers. *Journal of Child Language*, 1–24.

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., ... others (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678.

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The norwegian communicative development inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*(1), 3–23.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). afex: Analysis of factorial experiments. *R package version 0.13–145*.

Snedeker, J., Li, P., & Yuan, S. (2003). Cross-cultural differences in the input to early word learning. In L. E. Associates (Ed.), *Proceedings of the twenty-fifth annual conference of the cognitive science society* (p. 1094-1099).

Soli, S. D., Zheng, Y., Meng, Z., & Li, G. (2012). Clinical assessment of early language development: A simplified short form of the mandarin communicative development inventory. *International journal of pediatric otorhinolaryngology*, *76*(9), 1255–1264.

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2008). Early lexical development of finnish children: A longitudinal study. *First language*, *28*(3), 259–279.

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year in Finnish children. *Journal of Child Language*, *36*(4), 779–806.

Strouse, G. A., & Ganea, P. A. (2017). Parent–toddler behavior and language differ when reading electronic and print picture books. *Frontiers in Psychology*, *8*, 677.

Styles, S., & Plunkett, K. (2008). What is 'word understanding'for the parent of a one-year-old? matching the difficulty of a lexical comprehension task to parental cdi report.

Sun, H., Ng, S., O'Brien, B., & Fritzsche, T. (2020). Child, family, and school factors in bilingual preschoolers' vocabulary development in heritage languages. *Journal of child language*, 1.

Sun, H., Steinkrauss, R., Wieling, M., & De Bot, K. (2018). Individual differences in very young chinese children's english vocabulary breadth and semantic depth: internal and external factors. *International Journal of Bilingual Education and Bilingualism*, *21*(4), 405–425.

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy–word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202–217.

Szagun, G., Steinbrink, C., Franik, M., & Stumper, B. (2006). Development of vocabulary and grammar in young german-speaking children assessed with a german language development inventory. *First language*, *26*(3), 259–280.

Szagun, G., Stumper, B., & Schramm, S. A. (2009). Fragebogen zur frühkindlichen sprachentwicklung (frakis). *Frankfurt/M: Pearson*.

Szagun, G., Stumper, B., & Schramm, S. A. (2014). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt, Germany: Pearson Assessment.

Tadmor, U. (2009). Malay-indonesian. *The world's major languages*, 791-818.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from mandarin speakers' early vocabularies. *Developmental psychology*, *32*(3), 492.

Tardif, T., & Fletcher, P. (2008). *Chinese communicative development inventories: user's guide and manual.* Peking University Medical Press.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*(5), 1115–1144.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929.

Tardif, T., Fletcher, P., Zhang, Z. X., Liang, W. L., Zuo, Q. H., & Chen, P. (2008). User's guide and manual for the Chinese communicative development inventories (Putonghua and Cantonese). *Peking University Medical Press, Beijing, China*.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of english and mandarin. *Child Development*, *70*(3), 620–635.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of english, italian, and mandarin. *Journal of Child Language*, *24*(3), 535–565. doi: 10.1017/S030500099700319X

Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the longitudinal study of australian children. *PLOS one*, *8*(9), e73046.

Thal, & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech, Language, and Hearing Research*, *31*(1), 115–123.

Thal, D., DesJardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur–Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Language Pathology*.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in child Development*, *59*(5), 174–179.

Tupas, T. R. F. (2011). English-knowing bilingualism in singapore: Economic pragmatism, ethnic relations and class. *English language education across greater China*, 46–69.

Urek, O., Vulāne, A., Darģis, R., Tauriņa, A., Zīriņa, T., & Simonsen, H. G. (2019). Latvian cdi: methodology, developmental trends and cross-linguistic comparison. *Journal of Baltic Studies*, 1–21.

Vales, C., & Fisher, A. V. (2019). When stronger knowledge slows you down: Semantic relatedness predicts children's co-activation of related items in a visual search paradigm. *Cognitive science*, *43*(6), e12746.

van der Linden, W. J., & Glas, C. A. (2010). *Elements of adaptive testing*. New York, NY: Springer.

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of child language*, *5*(2), 311–326.

Wanless, S. B., McClelland, M. M., Acock, A. C., Chen, F.-M., & Chen, J.-L. (2011). Behavioral regulation and early academic achievement in taiwan. *Early Education & Development*, *22*(1), 1–28.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450.

Waxman, S., Fu, X., Ferguson, B., Geraghty, K., Leddon, E., Liang, J., & Zhao, M.-F. (2016). How early is infants' attention to objects and actions shaped by culture? new evidence from 24-month-olds raised in the us and china. frontiers in psychology. *Frontiers in psychology*, *7*, 97. doi: 10.3389/fpsyg.2016.00097

Weismer, S. E. (2017). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *Language disorders from a developmental perspective* (pp. 83–101). Psychology Press.

Welsh, S. N., & Hoff, E. (2020). Language exposure outside the home becomes more english-dominant from 30 to 60 months for children from spanish-speaking homes in the united states. *International Journal of Bilingualism*, 1367006920951870.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in cognitive sciences*, *12*(4), 144–151.

Williams, K. T. (2018). Expressive vocabulary test (3rd ed.) [measurement instrument]. *Bloomington, MN: NCS Pearson*.

Wu, Y. (1990). The usages of kinship address forms amongst non-kin in mandarin chinese: The extension of family solidarity. *Australian Journal of Linguistics*, *10*(1), 61–88.

Xuan, L., & Dollaghan, C. (2013). Language-specific noun bias: evidence from bilingual children. *Journal of child language*, *40*(5), 1057–1075.

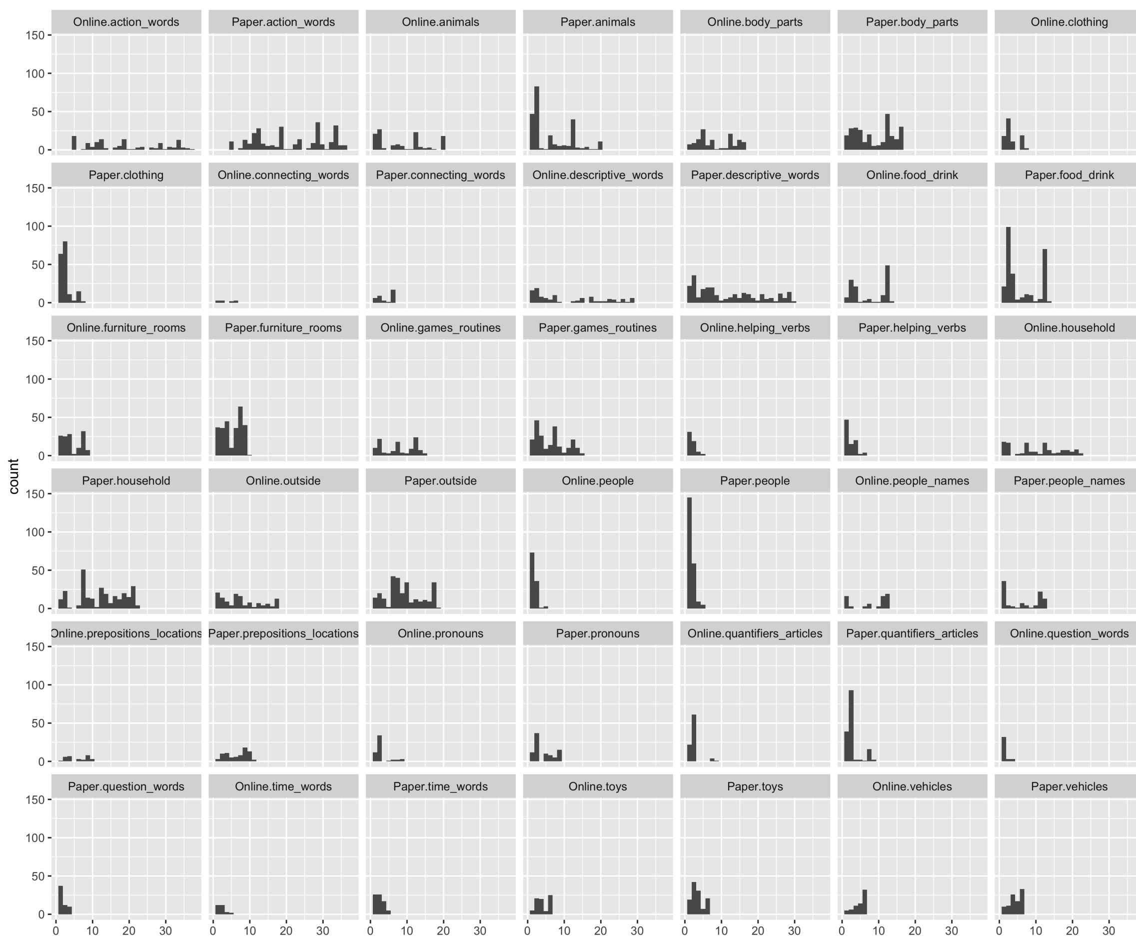
Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, *6*(1), 59–64.

Zuilkowski, S. S., McCoy, D. C., Jonason, C., & Dowd, A. J. (2019). Relationships among home literacy behaviors, materials, socioeconomic status, and early literacy outcomes across 14 low-and middle-income countries. *Journal of Cross-Cultural Psychology*, *50*(4), 539–555.

Appendix A

Appendix for Chapter 3

Figure A-1: *Histograms of the total number of occurances (count) for the total number of IRT-selected words in each category (n) across sampling type.*



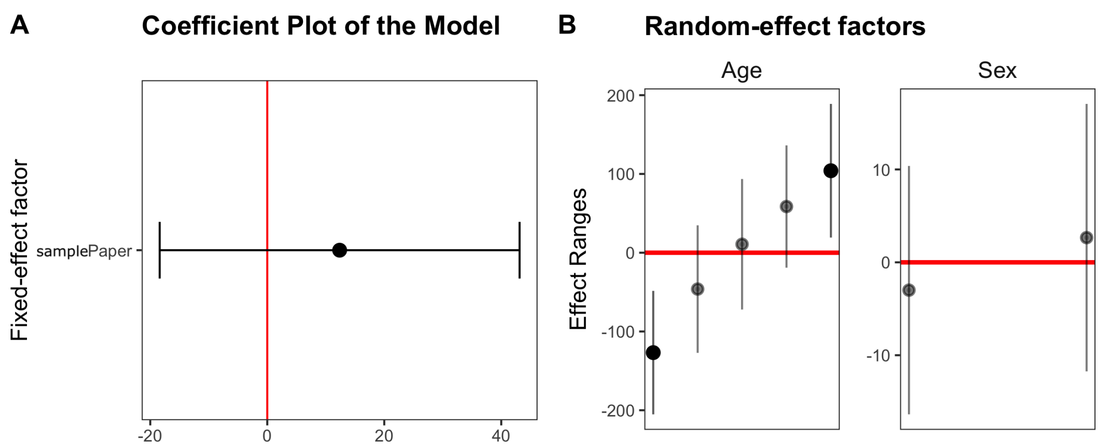


Figure A-2: *Linear mixed model fitted with the sampling type as the fixed effect factor and with the age and gender as random effects.* ***A)*** *Coefficient plot of the fixed effect. (The 'samplePaper' represents the differences between the paper and the online sample, with the online sample as the reference point)* ***B)*** *Effect ranges of random effects.*

Table A.1. Comparison of the *IRT version* and the original model (in parentheses) with different test item sizes on the American English CDI-WS and the baseline (random list), by gender.

Length	Females			Males			Baseline		
	r with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
680	.988 (.988)	.03 (.03)	.999 (.999)	.989 (.989)	.03 (.03)	.999 (.999)	1.000	.00	1.000
400	.990 (.987)	.03 (.03)	.999 (.999)	.990 (.987)	.03 (.03)	.999 (.999)	.998	.01	1.000
200	.988 (.985)	.03 (.04)	.999 (.999)	.989 (.985)	.04 (.04)	.999 (.999)	.993	.02	.999
100	.982 (.979)	.04 (.04)	.998 (.998)	.982 (.978)	.04 (.04)	.998 (.998)	.985	.04	.999
50	.976 (.968)	.05 (.05)	.997 (.997)	.976 (.966)	.05 (.05)	.997 (.997)	.967	.05	.997
25	.963 (.950)	.06 (.07)	.996 (.995)	.964 (.946)	.06 (.07)	.997 (.995)	.936	.07	.994
10	.937 (.884)	.07 (.10)	.994 (.990)	.937 (.873)	.07 (.10)	.994 (.989)	.856	.12	.985
5	.891 (.820)	.11 (.13)	.988 (.982)	.886 (.812)	.10 (.13)	.989 (.982)	.765	.17	.970

Table A.2. Comparison of the *IRT version* and the original model (in parentheses) with different test item sizes on the Danish CDI-WS and the baseline (random list), by gender.

Length	Females			Males			Baseline		
	r with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
725	.982 (.982)	.03 (.04)	.999 (.998)	.983 (.983)	.03 (.04)	.999 (.998)	1.000	.00	1.000
400	.985 (.980)	.03 (.04)	.999 (.998)	.987 (.981)	.03 (.04)	.999 (.998)	.997	.01	1.000
200	.985 (.977)	.03 (.04)	.999 (.998)	.985 (.979)	.04 (.04)	.998 (.998)	.990	.02	.999
100	.981 (.969)	.04 (.05)	.998 (.998)	.981 (.971)	.04 (.05)	.998 (.998)	.978	.03	.999
50	.974 (.957)	.04 (.06)	.998 (.997)	.974 (.956)	.05 (.05)	.998 (.997)	.955	.05	.997
25	.964 (.931)	.05 (.07)	.997 (.995)	.961 (.932)	.05 (.07)	.997 (.995)	.913	.07	.995
10	.924 (.863)	.06 (.09)	.996 (.991)	.939 (.870)	.06 (.09)	.995 (.991)	.807	.12	.986
5	.866 (.792)	.10 (.12)	.989 (.985)	.888 (.801)	.10 (.11)	.989 (.986)	.702	.16	.971

Table A.3. Comparison of the *IRT version* and the original model (in parentheses) with different test item sizes on the Beijing Mandarin CDI-WS and the baseline (random list), by gender.

Length	Females			Males			Baseline		
	r with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
799	.976 (.976)	.05 (.06)	.997 (.994)	.974 (.974)	.04 (.05)	.997 (.997)	1.000	.00	1.000
400	.981 (.975)	.05 (.06)	.997 (.994)	.979 (.973)	.05 (.05)	.997 (.997)	.997	.01	1.000
200	.980 (.971)	.05 (.07)	.997 (.994)	.978 (.970)	.06 (.06)	.996 (.996)	.993	.02	1.000
100	.969 (.964)	.06 (.07)	.996 (.994)	.974 (.968)	.06 (.06)	.996 (.996)	.983	.03	.999
50	.957 (.950)	.06 (.07)	.995 (.993)	.967 (.959)	.07 (.07)	.995 (.995)	.965	.05	.998
25	.942 (.930)	.07 (.08)	.995 (.991)	.955 (.947)	.07 (.07)	.994 (.994)	.932	.07	.995
10	.916 (.871)	.08 (.11)	.994 (.987)	.930 (.902)	.09 (.09)	.992 (.991)	.852	.11	.987
5	.873 (.790)	.10 (.13)	.990 (.979)	.893 (.826)	.10 (.13)	.989 (.983)	.754	.16	.974

Table A.4. Comparison of the *IRT version* and the original model with flexible approach in fitting of polynomial (in parentheses) with different test item sizes on the Italian CDI-WS and the baseline (random list), by gender.

Length	Females			Males			Baseline		
	r with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.	*r* with full CDI	Avg.*SE*	Rel.
670	.993 (.993)	.02 (.03)	.999 (.998)	.996 (.996)	.03 (.03)	.997 (.999)	1.000	.00	1.000
400	.992 (.992)	.03 (.04)	.999 (.998)	.995 (.994)	.04 (.03)	.997 (.999)	.998	.01	1.000
200	.987 (.989)	.04 (.04)	.998 (.998)	.990 (.990)	.05 (.04)	.996 (.998)	.992	.02	.999
100	.976 (.983)	.05 (.05)	.997 (.997)	.981 (.981)	.06 (.05)	.996 (.997)	.983	.04	.999
50	.965 (.970)	.06 (.06)	.995 (.996)	.971 (.962)	.07 (.06)	.994 (.996)	.964	.05	.997
25	.954 (.950)	.07 (.08)	.994 (.994)	.960 (.939)	.08 (.08)	.993 (.993)	.929	.08	.994
10	.943 (.877)	.08 (.11)	.993 (.987)	.931 (.862)	.08 (.11)	.992 (.986)	.840	.12	.984
5	.912 (.797)	.10 (.15)	.990 (.976)	.895 (.765)	.10 (.16)	.988 (.973)	.740	.18	.967

Table A.5. Descriptive statistics of the conceptual productive vocabulary across month-old groups for the data from the paper and online sample.

	M	*SD*	*min*	*max*	*n*
16 – 18 mo					
Paper					
Female	62.38	57.58	0	180	26
Male	54.44	65.74	0	237	34
Online					
Female	17.38	22.58	0	56	8
Male	29.33	32.23	0	81	12
19 – 21 mo					
Paper					
Female	148.62	129.80	5	452	21
Male	112.04	120.98	0	556	24
Online					
Female	97.33	82.22	1	229	6
Male	128.21	147.99	0	432	14
22 – 24 mo					
Paper					
Female	185.09	137.89	15	524	32
Male	184.44	154.99	4	514	34
Online					
Female	208.33	189.20	0	517	12
Male	176.53	156	0	437	17
25 – 27 mo					
Paper					
Female	248.33	157.78	10	575	27
Male	239.17	145.03	21	572	18
Online					
Female	242.87	137.71	0	515	15
Male	221	173.71	0	436	13
28 – 30 mo					
Paper					
Female	304.24	178.93	0	572	25
Male	272.54	165.32	17	599	28
Online					
Female	298.81	183.42	0	577	16
Male	236.88	214.99	0	588	17

Table A.6. The hand-picked 100-item Malaysian CDI-SF developed with reference to guideline in Fenson et al. (2000).

Nominals		
Animals	mouse	horse
	frog	monkey
	spider	snake
	butterfly	
	worm	
Vehicles	bicycle	train
	lorry	airplane
Toys	ball	
	book	
Food & drinks	milk	sweets
	water	ice
	banana	fruit
	vegetables	ice cream
	chicken	
Clothings	hat	trousers
	shoes	underwear
	slipper	
Body parts	cheek	mouth
	leg	hand
	eye	
Household items	tooth-brush	spoon
	medicine	bottle
	pillow	lamp
Furniture & rooms	table	stairs
	door	bed
People	baby	maternal grandmother
	daddy	older brother
	mummy	paternal grandmother
Outside things	home	movie
	star	circus
	water	roof

Non-nominals		
Games & routines	bye-bye no yes	hello bath
Action words	carry play drop open close throw sing	have kiss eat knock clean cut (with knife)
Descriptive words	hot cold high wet broken	quiet hungry pink happy dry
Words about time	night before after	
Pronouns	this he / she my	
Question words	who	
Prepositions & locations	out above behind	
Quantifiers & articles	a lot not a / an	
Helping verbs	want to can have to	
Connecting words	then	

Appendix B

Appendix for Chapter 4

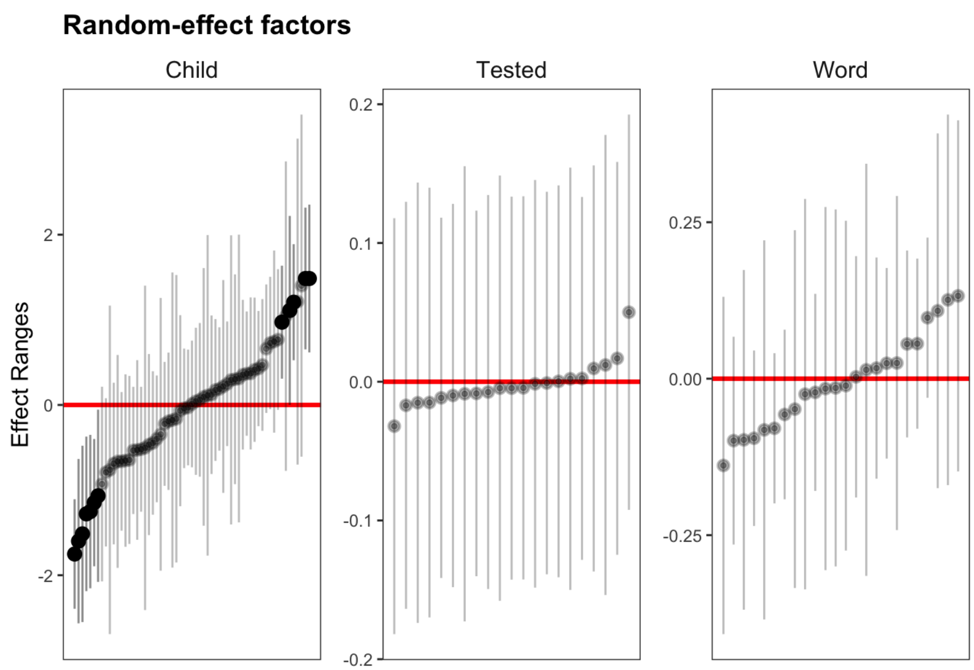


Figure B-1: *Effect range plot of random-effect factors in the generalised linear mixed model fitted to examine the effect of experimental condition on children's accuracy in the recognition task.*

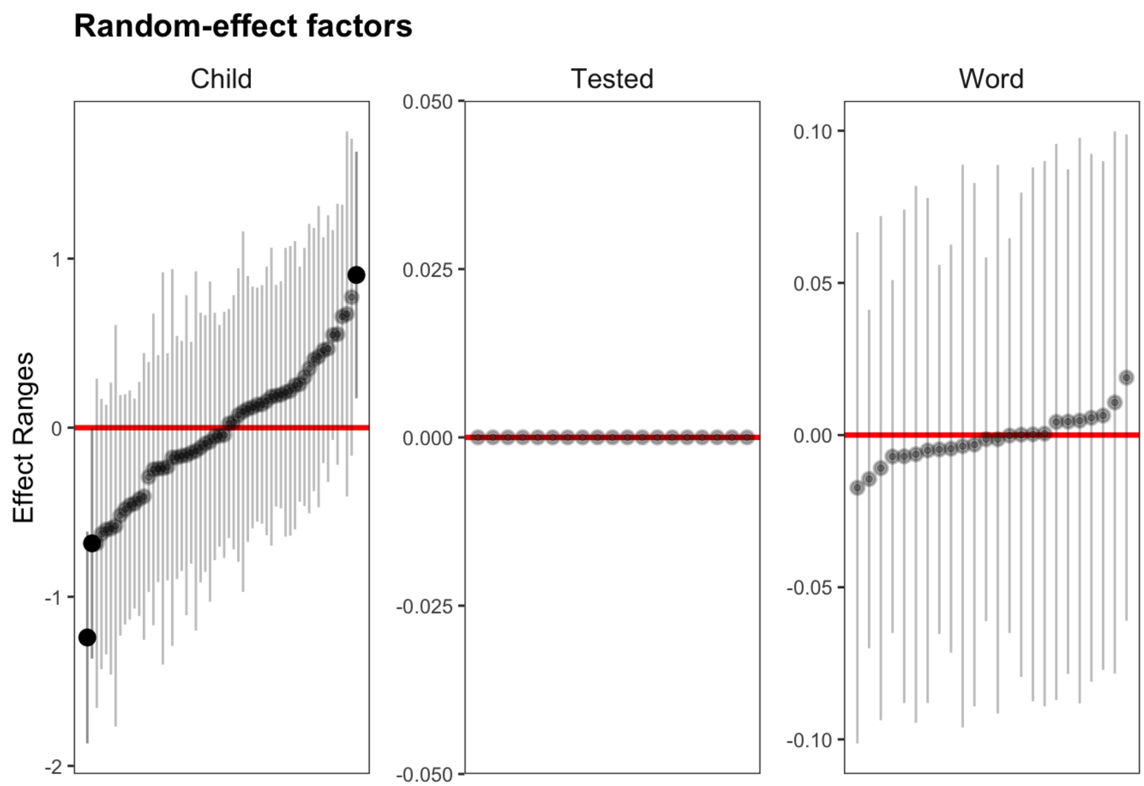
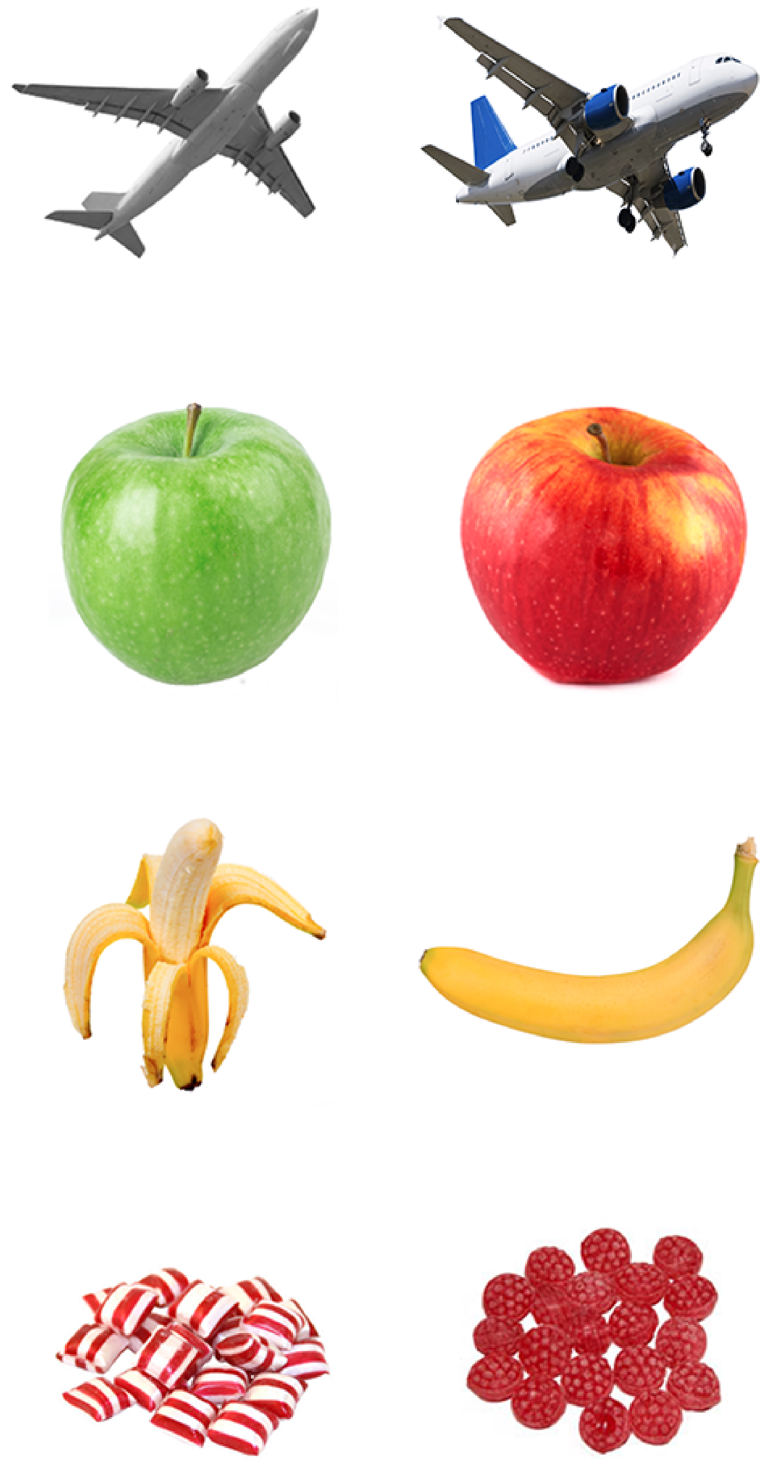


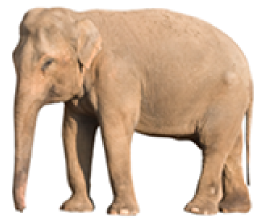
Figure B-2: *Effect range plot of random-effect factors in the generalised linear mixed model fitted to examine whether parents' insight is consistent with children's accuracy in recognition task.*

Table B.1. Analysis of Deviance table summary using Type II Wald Chi-square tests for generalised linear mixed-effects model fitted to examine whether parental reports reflect children's accuracy and whether the significant effects from the previous model interact with parental reports.

	χ^2	Df	$Pr(>\chi^2)$
Target	5.39	1	.020
Distractor	.069	1	.793
Similarity	1.814	1	.178
Animacy	5.772	1	.016
Comprehension	11.136	1	.001
Type	.001	1	.978
Difficulty	17.165	2	< .001
Target:Distractor	2.616	1	.106
Target:Similarity	.030	1	.863
Distractor:Similarity	.843	1	.359
Target:Animacy	.176	1	.675
Distractor:Animacy	1.137	1	.286
Similarity:Animacy	7.318	1	.007
Target:Type	1.257	1	.262
Distractor:Type	.195	1	.658
Animacy:Type	5.180	1	.023
Target:Difficulty	.776	2	.678
Distractor:Difficulty	6.129	2	.047
Target:Distractor:Similarity	1.334	1	.248
Target:Distractor:Animacy	.374	1	.541
Target:Similarity:Animacy	.372	1	.542
Distractor:Similarity:Animacy	.240	1	.624
Target:Distractor:Type	1.184	1	.277
Target:Animacy:Type	3.001	1	.083
Distractor:Animacy:Type	.707	1	.400
Target:Distractor:Difficulty	.098	2	.952
Target:Distractor:Similarity:Animacy	1.852	1	.174
Target:Distractor:Animacy:Type	.009	1	.926

Figure B-3: *Full list of stimuli used in the main trials.*



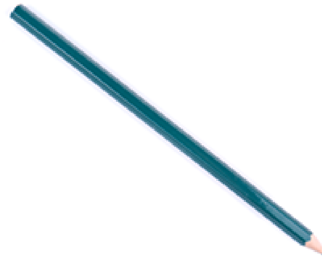




















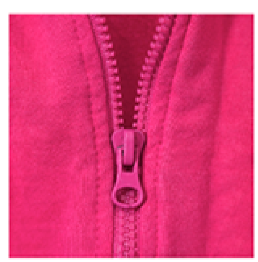


Figure B-4: *Stimuli used in the practice trials.*

Appendix C

Appendix for MCDI-M

Demographic Questions

Demographic

Start of Block: Demographic Info_SG

Q3 **Questionnaire about child's early language development**
You are invited to participate in this survey which aims to obtain information about your child's early language development. **Any answer is confidential and kept secret. There is no right or wrong answer.**

Q4 **Child's Gender:**

- ○ Male (1)
- ○ Female (2)

Q5 **Ethnicity:**

- ○ Malay (1)
- ○ Chinese (2)
- ○ Indian (3)
- ○ Others: (4) ____

Q6 **Birthday(DD/MM/YYYY):**

____

Q7
Position of the child:
______ -th among ______ siblings.

____

Q8 **Mom's Occupation:**

◯ Full-time job (1)

◯ Part-time job (2)

◯ Housewife (3)

Q9 **Did the child born full term or pre-term?**

◯ Born full term (1)

◯ Born pre-term. (Please state the week) (2) ____________________

Q10 **Child's weight at birth (kg):**

____

Q11 **Early development of children**

◯ No sign of developmental delay (1)

◯ Have sign of developmental delay (Please specify): (2) ____________________

Q12 **Age start walking (Month):**

____

Q13 **Health condition of the child:**

- ○ Generally healthy with no significant short or long term illness. (1)
- ○ Consulted doctor because of short-term illness: (2) ____
- ○ Consulted doctor or admitted into hospital because of long-term illness: (3) ____

Q14 **Do you or your husband/wife have had any close relative who suffered from developmental delays?:**

- ○ Yes. (Please indicate who and the reason.) (1) ____
- ○ No. (2)

Page Break

Q15 **Combined monthly household income (Please pick as appropriate):**

- ◯ Less than RM1,500 (31)
- ◯ RM 1,501 – 4,000 (26)
- ◯ RM 4,001 – 6,000 (32)
- ◯ RM 6,001 – 8,500 (33)
- ◯ RM 8,500 and above (34)

Q16 **Mother's highest educational level:**

- ◯ Secondary school (4)
- ◯ Colleges / Polytechnics (5)
- ◯ University (Undergraduate) (6)
- ◯ Other (7) ____

Q17 **Father's highest educational level:**

- ◯ Secondary school (4)
- ◯ Colleges / Polytechnics (5)
- ◯ University (Undergraduate) (6)
- ◯ Other (8) ____

Q18 Contact details (It will only be used to contact you in case that you have won the prize draw.)

○ Email address (2) ____

Page Break

Q19

Languages the family used with the child (May choose more than one):

Please estimates the percentage of language use.

Example:

Mom who spent half of the time speaking Malay and another half in English when interacting with the child.

Mom ... Malay 50 English 50 Mandarin
0 ... Total 100

Mom who spent all of the time speaking English when interacting with the child.

Mom ... Malay 0 English 100 Mandarin
0 ... Total 100

Mom who spent most of the time speaking English when interacting with the child.

Mom ... Malay 10 English 90 Mandarin
0 ... Total 100

	Father (1)	Mom (2)	Siblings (3)	Grandparents (4)	Relatives (5)	Neighbours (6)
Malay (1)						
English (2)						
Chinese (3)						
Tamil (4)						
Others 1 (5)						

Others 2 (6)						
Others 3 (7)						
Others 4 (8)						
Total						

Q20 Please select again the language(s) that is/are exposed to the child:

- ☐ Malay (1)
- ☐ English (2)
- ☐ Mandarin (3)
- ☐ Tamil (4)
- ☐ Others 1 (please specify): (5) ____
- ☐ Others 2 (please specify): (6) ____
- ☐ Others 3 (please specify): (7) ____
- ☐ Others 4 (please specify): (8) ____

Page Break

Q21 **Amount of time the family spent with the child:**
Please specify the amount of time spent in terms of percentage.
Please make sure the boxes in the "Total" rows all adds up to 100% (The amount in the "Total" boxes are calculated automatically, please fill in only the percentage for the amount of time spent).

Example:
The child spent more than half of the time with his/her mom. The remaining time spent was distributed evenly with his/her father and siblings. **Amount of time spent (Percentage):** Father 20 Mom 60 Siblings 20 ... Total 100

The child spent most of the time with his/her mom. **Amount of time spent (Percentage):** Father 20 Mom 80 Siblings 0 ... Total 100

The child spent half of the time with his/her mom and another half with his/her father. **Amount of time spent (Percentage):** Father 50 Mom 50 Siblings 0 ... Total 100

	Amount of time spent (Percentage): (22)
Father (2)	
Mom (3)	
Siblings (4)	
Grandparents (5)	
Relatives (6)	

Neighbours (9)	
Others 1 (7)	
Others 2 (8)	
Total	

Page Break

Q22

Part B
Trilingual Vocabulary Checklist
Has your child begun to produce any word?

- ○ Not yet (1)
- ○ Sometimes (Begin at what month (age in months)?): (2) ________________
- ○ Often (Begin at what month (age in months)?): (3) ________________

End of Block: Demographic Info_SG

Vocabulary Items Interface

Figure C-1: *Parents are prompted to select instructional language between English, Malay and Mandarin via the dropdown menu at the top right corner.*

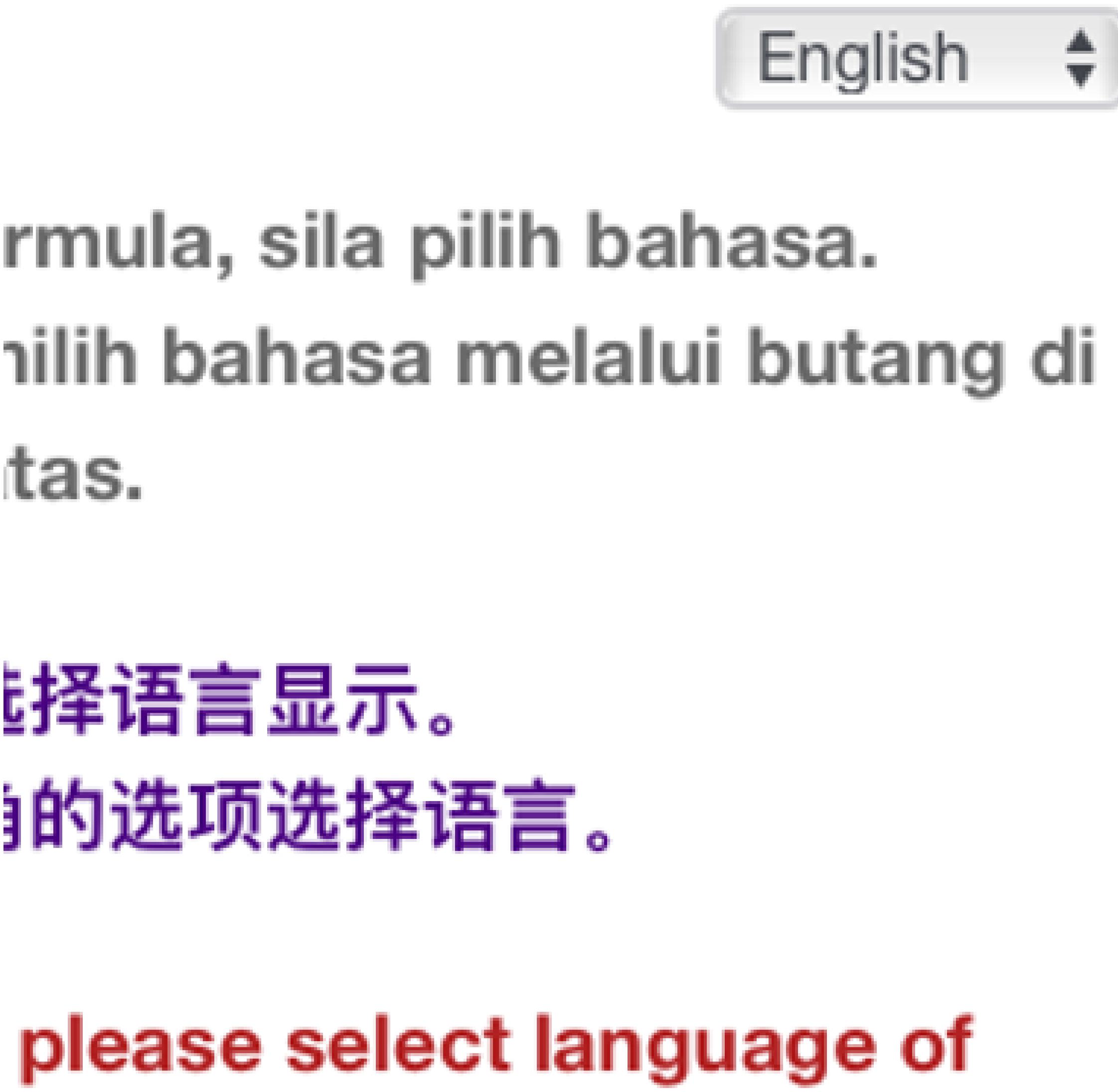


Figure C-2: *Parents are asked to select language(s) that is/are exposed to the child. The selection of Malay, English and/or Mandarin corresponds with the languages in which the vocabulary are shown in the vocabulary list.*

Figure C-3: *This is an example after the parents selected "Malay" and "English".*

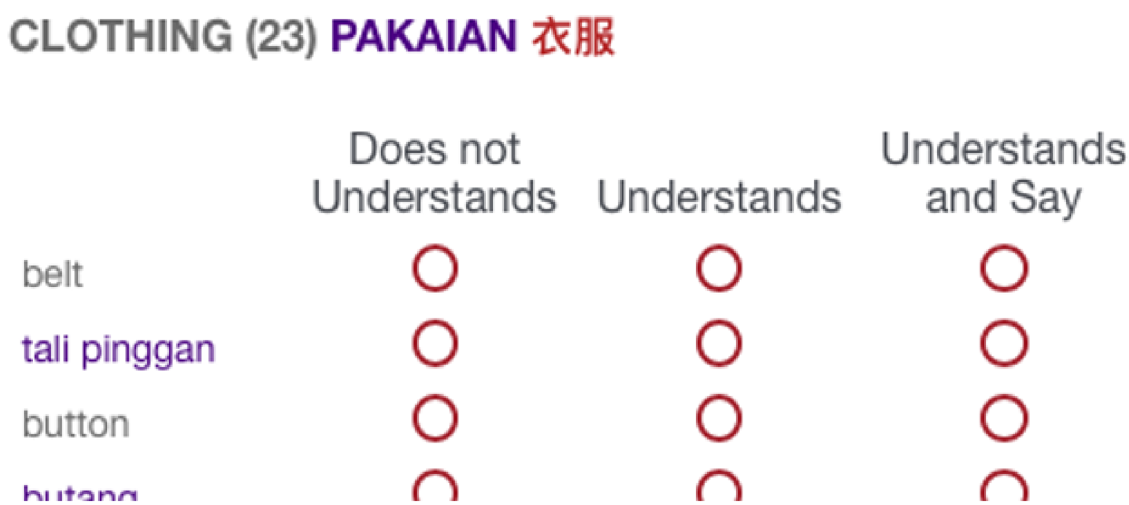


Figure C-4: *Parents are asked whether their child begin producing words. This is an example when a parent selects "Not Yet". The response correspond with the options displayed for all vocabulary item.*

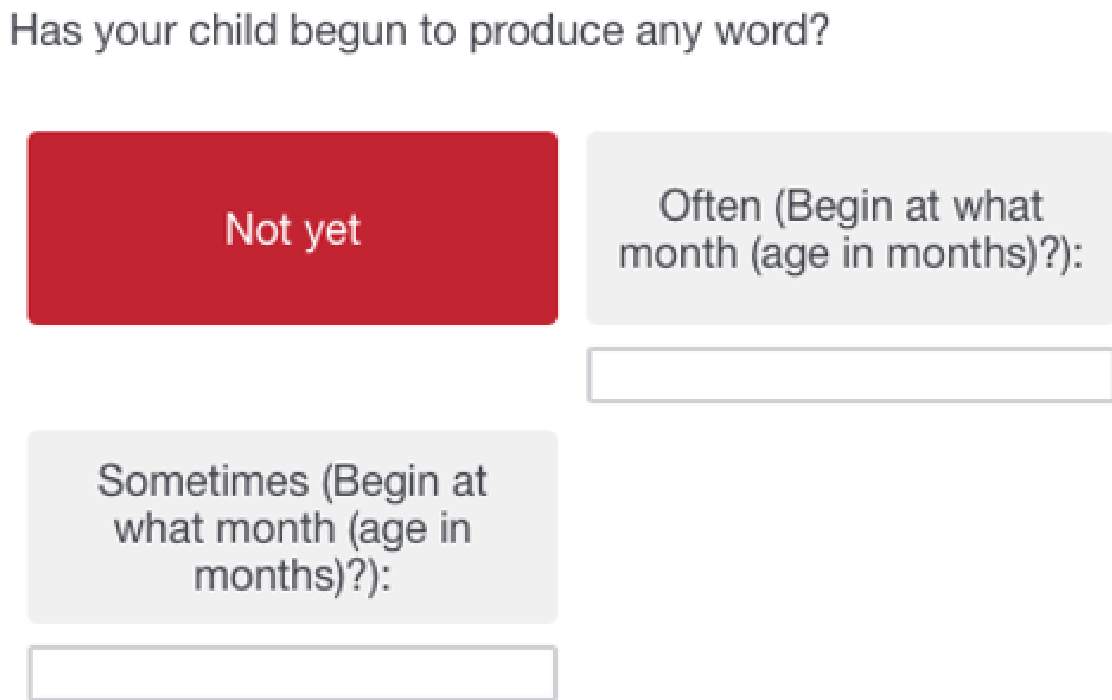


Figure C-5: *This is an example after the parents selected "Not Yet". Parents were asked to assess whether their child "Does not understands" or "Understands" a word.*

Figure C-6: *Parents are asked whether their child begin producing words. This is an example when a parent selects "Often...". The response correspond with the options displayed for all vocabulary item.*

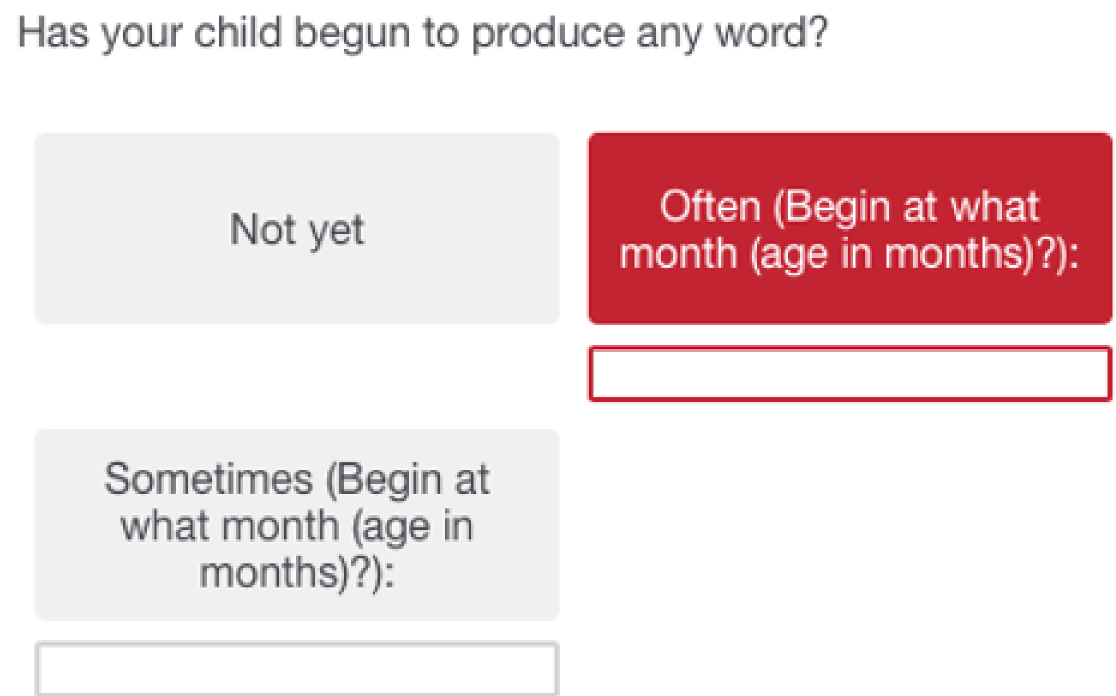
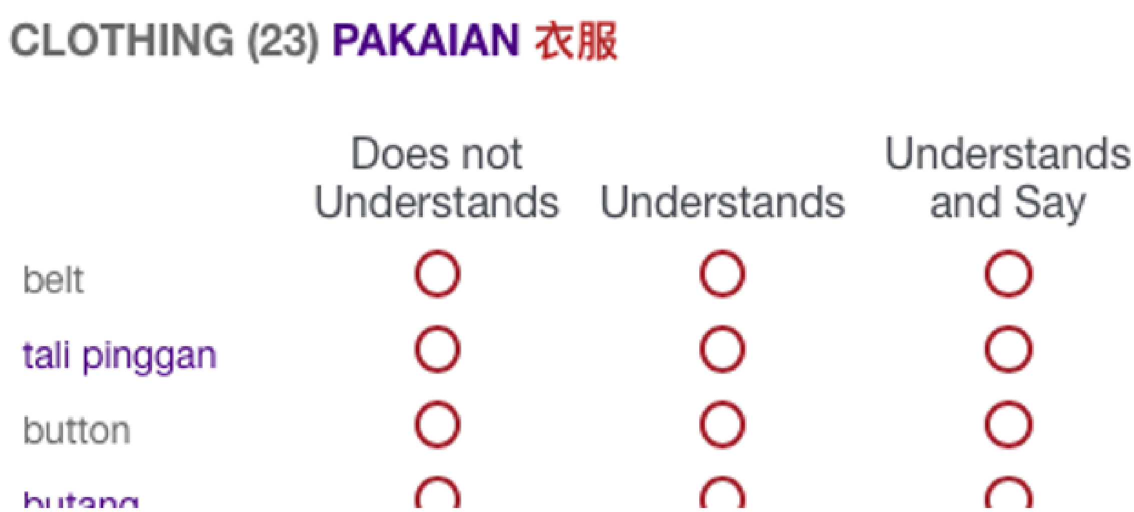


Figure C-7: *This is an example after the parents selected "Often..." or "Sometimes..." produce any words. Parents were asked to assess whether their child "Does not understands", "Understands" or "Understands and speaks" a word.*



Vocabulary Items in MCDI-M

Vocabulary Item	Semantic Category
belt / tali pinggan / 腰带	clothing
button / butang / 钮扣	clothing
dress / skirt panjang / 长裙	clothing
diaper / lampin / 尿片	clothing
handkerchief / sapu tangan / 手帕	clothing
hat / topi / 帽子	clothing
jacket / jaket / 冷衣	clothing
jeans / jean / 牛仔裤	clothing
napkin / kain lampin / 尿布	clothing
necklace / rantai leher / 项链	clothing
pants / seluar panjang / 长裤	clothing
pyjamas / baju tidur / 睡衣	clothing
shirt / baju kemeja / 衬衫	clothing
shoes / kasut / 鞋子	clothing
shorts / seluar pendek / 短裤	clothing
singlet / baju dalam / 背心	clothing
skirt / skirt / 裙子	clothing
slipper / selipar / 拖鞋	clothing
sock / sarung kaki / 袜子	clothing
trousers / seluar / 裤子	clothing
T-shirt / baju kemeja T / 恤衫	clothing
underwear / seluar dalam / 内裤	clothing
zip / zip / 拉链	clothing
a / an / satu / 一个	quantifiers_articles
a little / sedikit / 一点	quantifiers_articles
all / semua / 全部	quantifiers_articles
a lot / banyak / 很多	quantifiers_articles
another / yang lain / 另一个	quantifiers_articles
any / mana-mana satu / 任何一个	quantifiers_articles
each / tiap-tiap / 各自	quantifiers_articles
every / setiap / 每个	quantifiers_articles
more / lebih / 再多一个	quantifiers_articles
much / lebih lagi / 更多	quantifiers_articles
none / tak ada / 没有	quantifiers_articles
not / bukan / 不是	quantifiers_articles
other / yang lain / 别的	quantifiers_articles
same / sama / 一样	quantifiers_articles
some / sesetengah / 一些	quantifiers_articles
too / juga / 也	quantifiers_articles
bath / mandi / 冲凉	games_routines
breakfast / sarapan pagi / 早餐	games_routines
bye-bye / bye-bye / 拜拜	games_routines
(phone) call / (panggilan) telefon / 电话	games_routines
call (great) / panggil / 叫	games_routines
carry (me) / angkat / 抱	games_routines
dinner / makan malam / 晚餐	games_routines
don’t / jangan / 不可以	games_routines

go out / keluar / 出去	games_routines
hello / halo / 哈罗	games_routines
hurry up / cepat-cepat / 快点	games_routines
lunch / makan tengah hari / 午餐	games_routines
nap / tidur / 小睡	games_routines
night-night / selamat malam / 晚安	games_routines
no / tidak / 不要	games_routines
no more (gone) / tak ada lagi / 没有了	games_routines
(peek-a-) boo / (main) cak / 嚓	games_routines
please / tolong / 请	games_routines
shhh / shhh / 嘘	games_routines
shopping / beli-belah / 买东西	games_routines
thank you / terima kasih / 谢谢	games_routines
turn around / pusing belakang / 转过来	games_routines
wait / tunggu / 等一下	games_routines
yes / ya / 是	games_routines
ball / bola / 球	toys
balloon / belon / 气球	toys
book / buku / 书	toys
bubbles / buih / 泡泡	toys
doll / anak patung / 洋娃娃	toys
game / permainan / 游戏	toys
pen / pen / 笔	toys
pencil / pensil / 铅笔	toys
present / hadiah / 礼物	toys
puzzle / pasel / 拼图	toys
story / cerita / 故事	toys
toy / alat mainan / 玩具	toys
ankle / pergelangan kaki / 脚踝	body_parts
arm / tangan / 手臂	body_parts
belly button / puat / 肚脐	body_parts
bottom / buntut / 屁股	body_parts
cheek / pipi / 脸颊	body_parts
chin / dagu / 下巴	body_parts
ear / telinga / 耳朵	body_parts
eye / mata / 眼睛	body_parts
face / muka / 脸	body_parts
finger / jari / 手指	body_parts
foot / kaki / 脚	body_parts
hair / rambut / 头发	body_parts
hand / tangan / 手	body_parts
head / kepala / 头	body_parts
knee / lutut / 膝盖	body_parts
leg / kaki / 脚	body_parts
lips / bibir / 嘴唇	body_parts
mouth / mulut / 嘴巴	body_parts
nail / kuku / 指甲	body_parts
nose / hidung / 鼻子	body_parts

shoulder / bahu / 肩膀	body_parts
teeth / gigi / 牙齿	body_parts
thumb / ibu jari / 姆指	body_parts
toe / jari kaki / 脚趾	body_parts
tongue / lidah / 舌头	body_parts
tummy / perut / 肚子	body_parts
angry / marah / 生气	descriptive_words
asleep / dah tidur / 睡了	descriptive_words
awake / dah bangun / 醒了	descriptive_words
bad / jahat / 坏	descriptive_words
better / lebih baik / 更好	descriptive_words
big / besar / 大	descriptive_words
black / hitam / 黑色	descriptive_words
blue / biru / 蓝色	descriptive_words
broken / rosak / 坏了	descriptive_words
brown / perang / 褐色	descriptive_words
careful / berhati-hati / 小心	descriptive_words
clean / bersih / 干净	descriptive_words
cold / sejuk / 冷	descriptive_words
cute / comel / 可爱	descriptive_words
dark / gelap / 暗	descriptive_words
dirty / kotor / 肮脏	descriptive_words
dry / kering / 干	descriptive_words
empty / kosong / 空	descriptive_words
fast / pantas / 快	descriptive_words
fine / bagus / 不错	descriptive_words
first / pertama / 第一	descriptive_words
full / penuh / 满了	descriptive_words
gentle / lemah-lembut / 温和	descriptive_words
good / baik / 好	descriptive_words
green / hijau / 青色	descriptive_words
happy / gembira / 高兴	descriptive_words
hard / keras / 硬	descriptive_words
heavy / berat / 重	descriptive_words
high / tinggi / 高	descriptive_words
hot / panas / 热	descriptive_words
(spicy) hot / pedas / 辣	descriptive_words
hungry / lapar / 饿	descriptive_words
hurt / terluka / 受伤	descriptive_words
last / (ter)akhir / 最后	descriptive_words
little / sedikit / 一点点	descriptive_words
long / panjang / 长	descriptive_words
loud / kuat / 大声	descriptive_words
naughty / nakal / 顽皮	descriptive_words
new / baru / 新	descriptive_words
nice / baik / 好的	descriptive_words
noisy / bising / 吵	descriptive_words
old / tua / 老	descriptive_words

pink / merah jambu / 粉红色	descriptive_words
poor / miskin / 穷	descriptive_words
pretty / cantik / 漂亮	descriptive_words
purple / ungu / 紫色	descriptive_words
quiet / senyap / 安静	descriptive_words
red / merah / 红色	descriptive_words
sad / sedih / 伤心	descriptive_words
scared / takut / 怕怕	descriptive_words
sick / sakit / 生病	descriptive_words
sleepy / mengantuk / 想睡	descriptive_words
slow / lambat / 慢	descriptive_words
soft / lembut / 软	descriptive_words
sticky / lekit / 黏	descriptive_words
stuck / tersekat / 卡住	descriptive_words
thirsty / dahaga / 口渴	descriptive_words
tiny / kecil / 很小	descriptive_words
tired / letih / 疲倦/累	descriptive_words
wet / basah / 湿	descriptive_words
white / putih / 白色	descriptive_words
windy / berangin / 有风	descriptive_words
yellow / kuning / 黄色	descriptive_words
auntie / makcik / 阿姨	people_names
daddy / bapa / 爸爸	people_names
maid's name / nama pembantu rumah / 女佣的名字	people_names
maternal grandfather / datuk / 外公	people_names
maternal grandmother / nenek / 外婆	people_names
mummy / emak / 妈妈	people_names
own name / nama sendiri / 自己的名字	people_names
older brother / abang / 哥哥	people_names
older sister / kakak / 姐姐	people_names
paternal grandfather / datuk / 爷爷	people_names
paternal grandmother / nenek / 奶奶	people_names
uncle / pakcik / 叔叔	people_names
younger brother / adik lelaki / 弟弟	people_names
younger sister / adik perempuan / 妹妹	people_names
airplane / kapal terbang / 飞机	vehicles
bicycle / basikal / 脚踏车	vehicles
boat / sampan / 船	vehicles
bus / bas / 巴士	vehicles
car / kereta / 车	vehicles
fire engine / kereta bomba / 救火车	vehicles
helicopter / helikopter / 直升机	vehicles
lorry / lori / 罗里	vehicles
motorcycle / motosikal / 摩托车	vehicles
ship / kapal / 轮船	vehicles
taxi / taksi / 的士	vehicles
train / keretapi / 火车	vehicles
trishaw / beca / 三轮车	vehicles

van / van / 货车	vehicles
am / ialah / 是	helping_verbs
can / boleh / 能	helping_verbs
have to / mesti / 必须	helping_verbs
let me / biar saya / 让我	helping_verbs
need to / perlu / 需要	helping_verbs
try to / cuba / 试	helping_verbs
want to / nak / 要	helping_verbs
will / akan / 会	helping_verbs
beach / pantai / 沙滩	outside
camping / berkehmah / 露营	outside
church / gereja / 教堂	outside
circus / sarkas / 马戏团	outside
clinic / klinik / 诊疗所	outside
cloud / awan / 云	outside
drain / longkang / 水沟	outside
farm / ladang / 农场	outside
flag / bendera / 旗	outside
flower / bunga / 花	outside
garden / kebun / 花园	outside
gate / pagar / 篱笆	outside
grass / rumput / 草	outside
home / rumah / 家	outside
hospital / hospital / 医院	outside
house / rumah / 屋子	outside
ladder / tangga / 梯子	outside
moon / bulan / 月亮	outside
mosque / masjid / 回教堂	outside
movie / wayang gambar / 电影	outside
mud / lumpur / 泥	outside
oil station / stesen minyak / 油站	outside
outside / luar / 外面	outside
park / taman / 公园	outside
party / parti / 庆祝会	outside
playground / taman permainan / 游乐场	outside
(swimming) pool / kolan renang / 游泳池	outside
rain / hujan / 雨	outside
rock / batu / 石头	outside
roof / bumbung / 屋顶	outside
sandpit / tempat pasir / 沙	outside
school / sekolah / 学校	outside
shop / kedai / 店	outside
sky / langit / 天空	outside
slide / papan gelongsor / 滑梯	outside
star / bintang / 星星	outside
stick / kayu pokok / 树枝	outside
stone / batu / 石子	outside
road / jalan / 马路	outside

sun / matahari / 太阳	outside
supermarket / pasaraya / 超级市场	outside
swing / buaian / 秋千	outside
temple / kuil / 庙	outside
town / bandar / 城	outside
tree / pokok / 树	outside
wall / dinding / 墙	outside
water / air / 水	outside
wind / angin / 风	outside
zoo / zoo / 动物园	outside
he / she / dia / 他 / 她/ 它	pronouns
his / hers / dia punya / 他的 / 她的	pronouns
I / aku / 我	pronouns
me / saya / 我	pronouns
my / saya punya / 我的	pronouns
myself / saya sendiri / 自己	pronouns
that / itu / 那个	pronouns
their / mereka punya / 他/她们的	pronouns
these / semua ini / 这些	pronouns
they / mereka / 他/她们	pronouns
this / ini / 这个	pronouns
those / semua itu / 那些	pronouns
our / kami punya / 我们的	pronouns
we / kami / 我们	pronouns
you / kamu / 你	pronouns
your / kamu punya / 你的	pronouns
yourself / kamu sendiri / 你自己	pronouns
after / kemudian / 过后	time_words
afternoon / petang / 下午	time_words
before / sebelum / 之前	time_words
day / hari / 日	time_words
later / nanti / 以后	time_words
morning / pagi / 早上	time_words
night / malam / 晚上	time_words
now / sekarang / 现在	time_words
time / masa / 时间	time_words
today / hari ini / 今天	time_words
tomorrow / esok / 明天	time_words
tonight / malam ini / 今晚	time_words
yesterday / semalam / 昨天	time_words
bite / gigit / 咬	action_words
blow / tiup / 吹	action_words
break / pecah / 打破	action_words
bring / bawa / 带	action_words
build / bina / 建造	action_words
bump / hentam / 碰	action_words
buy / beli / 买	action_words
carry / angkat / 抱	action_words

catch / tangkap / 捉	action_words
chase / kejar / 追	action_words
clap / tepuk tangan / 拍手	action_words
clean / cuci / 打扫	action_words
climb / panjat / 爬	action_words
close / tutup / 关	action_words
cook / masak / 煮	action_words
cover / tutup / 盖	action_words
cry / menangis / 哭	action_words
cut (with scissors) / gunting / 剪	action_words
cut (with knife) / potong / 切	action_words
dance / menari / 跳舞	action_words
draw / lukis / 画	action_words
drink / minum / 喝	action_words
drive / memandu / 驾车	action_words
drop / jatuh / 掉	action_words
dry / menge-ringkan / 擦干	action_words
eat / makan / 吃	action_words
fall / jatuh / 跌到	action_words
feed / suap / 喂	action_words
find / cari / 找	action_words
finish / habis / 做完	action_words
fit / mencan-tumkan / 装	action_words
fix / membaiki / 修理	action_words
get / ambil / 取	action_words
give / beri / 给	action_words
go / pergi / 走	action_words
hate / benci / 讨厌	action_words
have / ada / 有	action_words
hear / dengar / 听	action_words
help / tolong / 帮	action_words
hide / sembunyi / 躲	action_words
hit / pukul / 打	action_words
hold / pegang / 拿着	action_words
hug / peluk / 拥抱	action_words
jump / lompat / 跳	action_words
kick / tendang / 踢	action_words
kiss / cium / 亲亲	action_words
knock / ketuk / 敲	action_words
lick / jilat / 舔	action_words
like / suka / 喜欢	action_words
listen / dengar / 听	action_words
look / tengok / 看	action_words
love / sayang / 爱	action_words
make / buat / 做	action_words
open / buka / 开	action_words
paint / (men)cat / 油漆	action_words
pick / angkat / 拾起来/ 选	action_words

play / main / 玩	action_words
pour / tuang / 倒(水)	action_words
pretend / pura-pura / 假装	action_words
pull / tarik / 拉	action_words
push / tolak / 推	action_words
put / letak / 放	action_words
read / baca / 读	action_words
ride / menunggang / 骑	action_words
run / lari / 跑	action_words
say / cakap / 说	action_words
see / tengok / 看	action_words
shake / goyang / 摇	action_words
share / kongsi / 分享	action_words
show / tunjuk / 给(我) 看	action_words
sing / nyanyi / 唱歌	action_words
sit / duduk / 坐	action_words
skate / meluncur / 溜冰	action_words
sleep / tidur / 睡觉	action_words
slide / menggelongsor / 滑下来	action_words
smile / senyum / 笑	action_words
spill / tumpah / 溢出	action_words
stand / berdiri / 站	action_words
stay / tinggal / 留	action_words
stop / henti / 停	action_words
sweep / sapu / 扫	action_words
swim / berenang / 游泳	action_words
swing / buai / 摇摇	action_words
take / ambil / 拿	action_words
talk / cakap / 讲	action_words
taste / rasa / 试/ 尝(食物)	action_words
tear / koyak / 撕破	action_words
tell / beritahu / 告诉	action_words
think / fikir / 想	action_words
throw / buang / 丢	action_words
tickle / meggelikan / 搔痒	action_words
touch / sentuh / 摸	action_words
wait / tunggu / 等	action_words
wake / bangun / 起来/醒	action_words
walk / jalan / 走	action_words
wash / cuci / 洗	action_words
watch / tengok / 观看	action_words
wipe / lap / 擦	action_words
wish / harap / 许愿	action_words
work / kerja / 做工	action_words
write / tulis / 写	action_words
and / dan / 和	connecting_words
because / sebab / 因为	connecting_words
but / tetapi / 但是	connecting_words

if / kalau / 如果	connecting_words
so / oleh itu / 所以	connecting_words
then / kemudian / 就	connecting_words
apple / epal / 苹果	food_drink
banana / pisang / 香蕉	food_drink
biscuit / biskut / 饼干	food_drink
bread / roti / 面包	food_drink
butter / mentega / 牛油	food_drink
cake / kek / 蛋糕	food_drink
carrot / lobak / 红萝卜	food_drink
chicken / ayam / 鸡	food_drink
chocolate / coklat / 巧克力	food_drink
coffee / kopi / 咖啡	food_drink
cornflake / emping jagung / 谷粮	food_drink
drink / minimum / 饮品	food_drink
egg / telur / 鸡蛋	food_drink
grape / anggur / 葡萄	food_drink
fish / ikan / 鱼	food_drink
food / makanan / 食物	food_drink
French fries / kentang goreng / 薯条	food_drink
fruit / buah-buahan / 水果	food_drink
hamburger/burger / burger / 汉堡包	food_drink
ice / ais / 冰	food_drink
ice cream / ais krim / 冰淇淋	food_drink
jam / jem / 果酱	food_drink
jelly / agar-agar / 果冻	food_drink
juice / jus / 果汁	food_drink
meat / daging / 肉	food_drink
melon / buah / 瓜	food_drink
milk / susu / 牛奶	food_drink
noodles / mi / 面	food_drink
nut / kacang / 花生	food_drink
orange / oren / 橙	food_drink
papaya / betik / 木瓜	food_drink
porridge / bubur / 粥	food_drink
rice / nasi / 饭	food_drink
sauce / sos / 酱	food_drink
soup / sup / 汤	food_drink
sugar / gula / 糖	food_drink
sweets / gula-gula / 糖果	food_drink
tea / teh / 茶	food_drink
vegetables / sayur / 蔬菜	food_drink
vitamin / vitamin / 维他命	food_drink
water / air / 水	food_drink
yogurt / yogurt / 酸奶	food_drink
aircon / hawa dingin / 冷气机	furniture_rooms
baby chair / kerusi tinggi / 高椅	furniture_rooms
bathroom / bilik mandi / 冲凉房	furniture_rooms

bed / katil / 床	furniture_rooms
bedroom / bilik tidur / 睡房	furniture_rooms
chair / kerusi / 椅子	furniture_rooms
cooker / pemasak / 锅	furniture_rooms
door / pintu / 门	furniture_rooms
drawer / laci / 抽屉	furniture_rooms
fan / kipas / 风扇	furniture_rooms
fridge / peti sejuk / 冰橱	furniture_rooms
kitchen / dapur / 厨房	furniture_rooms
living room / bilik tetamu / 客厅	furniture_rooms
microwave oven / ketuhar / 微波炉	furniture_rooms
potty / tong kencing / 尿壶	furniture_rooms
room / bilik / 房间	furniture_rooms
shower / mandi hujan / 花洒	furniture_rooms
sink / sinki / 洗手盆	furniture_rooms
sofa / sofa / 沙发	furniture_rooms
stairs / tangga / 楼梯	furniture_rooms
stool / bangku / 凳子	furniture_rooms
stove / tempat masak / 火炉	furniture_rooms
table / meja / 桌子	furniture_rooms
toilet / tandas / 厕所	furniture_rooms
tv/ television / televisyen / 电视机	furniture_rooms
washing machine / mesin cuci / 洗衣机	furniture_rooms
window / tingkap / 窗	furniture_rooms
baby / bayi / 婴儿	people
boy / budak lelaki / 男孩子	people
child / anak / 小孩子	people
clown / pelawak / 小丑	people
doctor / doktor / 医生	people
fireman / ahli bomba / 救火员	people
friend / kawan / 朋友	people
girl / budak perempuan / 女孩子	people
man / lelaki / 男人	people
nurse / jururawat / 护士	people
people / orang / 人	people
pet's name / nama haiwan belaan / 宠物的名字	people
policeman / polis / 警察	people
postman / posmen / 油差	people
teacher / cikgu / 老师	people
woman / perempuan / 女人	people
about / tentang / 关于	prepositions_locations
above / di atas / 上面	prepositions_locations
around / sekitar / 在.... 周围	prepositions_locations
at / di / 在	prepositions_locations
away / jauh / 远离	prepositions_locations
back / di belakang / 后面	prepositions_locations
behind / di belakang / 在.....背后	prepositions_locations
beside / di tepi / 旁边	prepositions_locations

by / oleh / 由	prepositions_locations
down / di bawah / 下面	prepositions_locations
for / untuk / 为	prepositions_locations
here / di sini / 这边	prepositions_locations
in / dalam / 里面	prepositions_locations
inside / di dalam / 在.....里面	prepositions_locations
next to / di sebelah / 靠近	prepositions_locations
on top of / di atas / 上面	prepositions_locations
out / luar / 外面	prepositions_locations
there / di sana / 那边	prepositions_locations
to / ke / 到	prepositions_locations
under / di bawah / 下面	prepositions_locations
up / ke atas / 向上	prepositions_locations
with / dengan / 跟	prepositions_locations
animal / binatang / 动物	animals
ant / semut / 蚂蚁	animals
bear / beruang / 熊	animals
bee / lebah / 蜜蜂	animals
bird / burung / 鸟	animals
buffalo / kerbau / 水牛	animals
butterfly / rama-rama / 蝴蝶	animals
cat / kucing / 猫	animals
chicken / ayam / 鸡	animals
cow / lembu / 牛	animals
deer / kancil / 鹿	animals
dog / anjing / 狗	animals
duck / itik / 鸭	animals
elephant / gajah / 大象	animals
fish / ikan / 鱼	animals
fox / musang / 狐狸	animals
frog / katak / 青蛙	animals
giraffe / zirafah / 长颈鹿	animals
goat / kambing / 羊	animals
goose / angsa / 鹅	animals
hamster / hamster / 仓鼠	animals
hen / ayam betina / 母鸡	animals
hippopotamus / badak air / 犀牛	animals
horse / kuda / 马	animals
house fly / lalat / 苍蝇	animals
kangaroo / kanggaru / 袋鼠	animals
kitten / anak kucing / 小猫	animals
lamb / anak kambing / 小羊	animals
lion / singa / 狮子	animals
lizard / biawak / 壁虎	animals
monkey / monyek / 猴子	animals
mosquito / nyamuk / 蚊子	animals
mouse / tikus / 老鼠	animals
octopus / sotong / 八爪鱼	animals

owl / burung hantu / 猫头鹰	animals
penguin / burung penguin / 企鹅	animals
orang-utan / orang utan / 人猿	animals
puppy / anak anjing / 小狗	animals
rabbit / arnab / 兔子	animals
rooster / ayam jantan / 公鸡	animals
shark / ikan yu / 鲨鱼	animals
sheep / biri-biri / 绵羊	animals
snail / siput / 蜗牛	animals
snake / ular / 蛇	animals
spider / labah-labah / 蜘蛛	animals
teddy bear / anak beruang / 玩具熊	animals
tiger / harimau / 老虎	animals
tortoise / kura-kura / 乌龟	animals
turtle / penyu / 海龟	animals
wolf / serigala / 狼	animals
worm / ulat / 虫	animals
zebra / kuda belang / 斑马	animals
how / bagaimana / 怎样	question_words
what / apa / 什么	question_words
when / bila / 几时	question_words
where / mana / 哪里	question_words
which / mana satu / 哪个	question_words
who / siapa / 谁	question_words
why / mengapa / 为什么	question_words
bin / tong / 桶	household
blanket / selimut / 被单	household
bottle / botol / 水罐	household
bowl / mangkuk / 碗	household
box / kotak / 盒子	household
broom / penyapu / 扫把	household
brush / berus / 刷子	household
bucket / baldi / 桶	household
camera / kamera / 相机	household
can / tin / 罐头	household
chopsticks / penyepit / 筷子	household
clock / jam / 时钟	household
coins / syiling / 钱币	household
comb / sikat / 梳子	household
cup / cawan / 杯	household
fork / garpu / 叉	household
glass / kaca / 玻璃	household
hammer / penukul / 锤子	household
handphone / telefon bimbit / 手提电话	household
jar / balang / 瓶子	household
keys / kunci / 锁匙	household
knife / pisau / 刀	household
lamp / lampu / 灯	household

light / cahaya / 光	household
medicine / ubat / 药	household
mirror / cermin / 镜子	household
money / duit / 钱	household
mop / pengelap lantai / 拖把	household
nail / paku / 钉子	household
paper / kertas / 纸	household
picture / lukisan / 图画	household
pillow / bantal / 枕头	household
plant / tumbuh-tumbuhan / 植物	household
plate / pinggan / 盘	household
purse / dompet / 钱包	household
radio / radio / 收音机	household
rubbish / sampah / 垃圾	household
scissors / gunting / 剪刀	household
soap / sabun / 肥皂	household
spoon / sudu / 汤匙	household
telephone / telefon / 电话	household
tissue paper / kertas tisu / 纸巾	household
tooth-brush / berus gigi / 牙刷	household
towel / tuala / 毛巾	household
umbrella / payung / 雨伞	household
vacuum cleaner / hampagas / 吸尘机	household
VCD / VCD / 光碟	household
watch / jam tangan / 手表	household

Appendix D

Appendix for Published Works

JSLHR

Research Article

A Bayesian-Inspired Item Response Theory–Based Framework to Produce Very Short Versions of MacArthur–Bates Communicative Development Inventories

Jun Ho Chai,[b,*] Chang Huan Lo,[b,*] and Julien Mayor[a]

Purpose: This study introduces a framework to produce very short versions of the MacArthur–Bates Communicative Development Inventories (CDIs) by combining the Bayesian-inspired approach introduced by Mayor and Mani (2019) with an item response theory–based computerized adaptive testing that adapts to the ability of each child, in line with Makransky et al. (2016).
Method: We evaluated the performance of our approach—dynamically selecting maximally informative words from the CDI and combining parental response with prior vocabulary data—by conducting real-data simulations using four CDI versions having varying sample sizes on Wordbank—the online repository of digitalized CDIs: American English (a very large data set), Danish (a large data set), Beijing Mandarin (a medium-sized data set), and Italian (a small data set).
Results: Real-data simulations revealed that correlations exceeding .95 with full CDI administrations were reached with as few as 15 test items, with high levels of reliability, even when languages (e.g., Italian) possessed few digitalized administrations on Wordbank.
Conclusions: The current approach establishes a generic framework that produces very short (less than 20 items) adaptive early vocabulary assessments—hence considerably reducing their administration time. This approach appears to be robust even when CDIs have smaller samples in online repositories, for example, with around 50 samples per month-age.

The MacArthur–Bates Communicative Development Inventories (CDIs) are one of the most widely used sets of parent report instruments for assessing young children's early language development (Fenson et al., 2007). Originally developed in American English (Fenson et al., 1993), CDIs have since been adapted into nearly 100 languages (e.g., Danish, Mandarin, and Italian), including language variations (e.g., British English, Australian English, and Singaporean English). Adaptations have also been developed in a number of sign languages, including American Sign Language and British Sign Language (see CDI Advisory Board, 2015, for a list of available adaptations).

CDIs typically consist of three forms: The CDI: Words and Gestures (CDI:WG)—targeting children approximately 8–18 months of age—assesses both comprehension and production of early vocabulary, as well as production of communicative gestures; the CDI: Words and Sentences (CDI:WS)—targeting children approximately 16–30 months of age—assesses productive vocabulary and morphosyntactic skills; and the CDI-III—a short form targeting children approximately 30–37 months of age—assesses productive vocabulary, syntactic maturity, and language use (Dale et al., 1998; Fenson et al., 2007).

These assessment tools rely on parents' knowledge about their children's language and allow a representative picture of children's early language development (Fenson et al., 2000). Beyond being cost-effective, CDIs are also reliable and valid, not only with children who are developing typically (Fenson et al., 1993, 2007; Law & Roy, 2008; Pan et al., 2004; Rescorla et al., 2005) but also with children with developmental disabilities (Galeote et al., 2016; Luyster et al., 2007; Mayne et al., 1999, 1998; Thal et al., 2007).

Through the application of CDIs in various languages, similarities have been observed in lexical development trajectories among children speaking different languages (Bleses et al., 2008; Braginsky et al., 2019; Frank et al., in press).

[a]University of Oslo, Norway
[b]University of Nottingham Malaysia, Semenyih, Selangor
Correspondence to Julien Mayor: julien.mayor@psykologi.uio.no
*Jun Ho Chai and Chang Huan Lo share first authorship.
Editor-in-Chief: Stephen M. Camarata
Received June 25, 2020
Revision received July 5, 2020
Accepted July 7, 2020
https://doi.org/10.1044/2020_JSLHR-20-00361

Disclosure: The authors have declared that no competing interests existed at the time of publication.

The evidence suggests that most children produce their first words between 12 and 20 months of age (Bleses et al., 2008; Devescovi et al., 2005; Fernald et al., 1998), that their vocabulary acquisition rate increases rapidly after 18 months of age (Bates & Goodman, 1997; Fernald et al., 2006, 1998), and that there is a strong relationship between lexical and grammatical development (e.g., Bates & Goodman, 1997; Caselli et al., 1999; Conboy & Thal, 2006; Devescovi et al., 2005; Marjanovič-Umek et al., 2013; Stolt et al., 2009). CDIs have also been used as additional criteria for identifying late language emergence; for example, starting from 24 months of age, a child is typically considered to be a late talker or a late language learner if they have an expressive vocabulary at or below the 10th percentile on the CDI (Dale et al., 2003; Desmarais et al., 2008; Ellis Weismer, 2007; Rescorla & Dale, 2013).

Despite the many advantages and widespread applications of CDIs, completion of the forms requires a significant amount of time and that the parent should be literate. The American English CDI:WS, for example, includes a vocabulary checklist of 680 words, organized into 22 semantic categories (e.g., vehicles, toys, people, action words, descriptive words, and question words). Under circumstances when a rapid assessment is desirable (whether in a battery of tests or in multilingual environments) or when parents have low literacy skills, the applicability of CDIs becomes limited. To address these drawbacks, Fenson et al. (2000) developed the first short-form versions (CDI:SF) of the CDI:WG and CDI:WS with items drawn from the original full forms. The former consists of an 89-item checklist, whereas the latter consists of two 100-item checklists to allow for repeated administrations. As with the full CDIs, these short forms have demonstrated high validity and reliability, and are at the same time highly correlated with the full forms, thus making them a useful alternative when time or parental literacy is limited (Fenson et al., 2000). Nevertheless, due to their brevity, these short forms may not be as precise as the full forms and may fail to capture individual differences. The short-form CDI:WS suffers from a ceiling effect after 27–28 months and even more so when children have a large vocabulary. Furthermore, it takes much time and effort to develop such forms for each language in order to maintain a good balance of items from different semantic categories, as well as items with different levels of difficulty.

With the objective to develop a short-form version of CDI:WS that maintains the accuracy and precision of the full form and is tailored to each child, Makransky et al. (2016) applied computerized adaptive testing (CAT; van der Linden & Glas, 2010), whose principle is based on item response theory (IRT; Embretson & Reise, 2000). In their approach (hereafter referred to as "CDI:CAT"), items in the American English CDI:WS norming sample[1] are fitted to an IRT model. During CAT, 10 items with maximal item information are initially sampled at random from the full CDI. The algorithm then selects subsequent items that reflect the ability parameter of the child that is estimated at each point (i.e., item) in the test. Based on the results obtained from CDI:CAT simulations with 5, 10, 25, 50, 100, 200, 400, and 680 (the full form) items, it was found that, at 50 items and above, CDI:CAT performed well, with correlations above .95 with the full CDI, average *SE* below .20, and reliability coefficients above .96 (above what Makransky et al. described as a minimal threshold for test acceptability). Although this may be a viable solution to reducing the lengths of the full CDI, the performance of CDI:CAT with novel empirical data, as pointed out by Makransky et al., may be lower due to a systematic or random error. It is also possible that the respondents would respond differently as items in CDI:CAT are presented in a semantically unstructured order as opposed to the semantic grouping adopted in CDIs.

Recently, Mayor and Mani (2019) presented a language-general approach that takes advantage of the richness of Wordbank (Frank et al., 2017), an open repository for cross-linguistic CDI data from over 75,000 children across 29 languages. Their approach combines a subset of items drawn randomly from the full forms with (prior) CDI data sampled from language-, gender-, and age-matching children on Wordbank. Real-data simulations conducted using CDI:WS data of American English (Fenson et al., 2007), German (Szagun et al., 2014), and Norwegian (Simonsen et al., 2014) revealed that, at 50 items, correlations reached .97, with average *SE*s of .05, and reliability coefficients of .99, suggesting that their approach, which takes into account children's age and gender, outperforms CDI:CAT. Empirical validation with 25- and 50-item checklists administered to the parents of German-speaking children further demonstrated good performance, with correlations of .96, average *SE*s of .14, and a reliability of .98, above Makransky et al.'s (2016) recommended thresholds, even when parents showed inconsistencies (about 10%–15% of responses) in responding in the full and short forms. However, to capture the full extent of the large variations in vocabulary acquisition (e.g., within- and between-age variations, gender differences; Fenson et al., 2007), Mayor and Mani's approach requires a considerably large sample size on Wordbank—for example, the German CDI:WS data set, the smallest data set used in the study of Mayor and Mani, has over 70 children in each age group. Thus, it is unclear how their approach would perform with smaller sample sizes, for example, for languages having fewer computerized forms on Wordbank.

Another obvious limitation to this approach is that items are *randomly* selected during the test. Consequently, the sampled items may be minimally informative, that is, items that are either too easy (e.g., "cat" is produced by over 95% of 30-month-olds in American English) or too difficult (e.g., "snowman" is produced by just 1% of 16-month-olds), and hence may inform little about a child's language ability.

To address these issues, our approach aims to produce language-general, short-form versions of CDIs in which items are selected to be maximally informative. Building on Mayor and Mani's (2019) work, we applied a

[1]The American English CDI-WS norming sample consists of 1,461 children between 16 and 30 months of age.

principled selection of items in place of random selection. More specifically, we applied IRT-based CAT in real-data simulations, as in Makransky et al. (2016). In IRT, items may differ in discrimination value, which determines how well each item discriminates the level of knowledge across individuals (Fraley et al., 2000). For instance, items with high difficulty level may discriminate two children having a high degree of knowledge but may not for weaker children. Thus, by applying IRT, the risk of sampling minimally informative items can be circumvented. To validate our approach, we selected four CDI:WS versions for which their sample sizes on Wordbank vary: American English (a very large data set; Fenson et al., 2000), Danish (a large data set; Bleses et al., 2008), Beijing Mandarin (a medium-sized data set; Tardif et al., 2008), and Italian (a small data set; Rinaldi et al., 2019). This, in turn, helped to evaluate the possibility of applying a CAT approach to languages for which only small samples are represented on Wordbank and to establish short language assessments that are reliable. An evaluation of performance was conducted across different age groups and genders.

The next section details the two main components of our approach, that is, the IRT-based selection of test items coupled with a full CDI score estimation based on Mayor and Mani's (2019) model. We then present the result and discuss the implications of our findings for researchers and practitioners intending to use short forms for quick and cost-effective assessments of young children's vocabulary.

Method

IRT-Based Item Selection

Prior to an assessment, items on the CDI are fitted to a two-parameter IRT model. Item parameters are estimated using the *mirt* function from Chalmers's (2012) mirt package in R. Each item is assigned two parameters: a *discrimination* parameter and a *difficulty* parameter. These item parameters are computed with marginal maximum likelihood estimation using the expectation–maximization algorithm (Bock & Aitkin, 1981). During CAT assessments, items with maximum information are prioritized and tested. The estimation of the ability parameter of each child is conducted using the weighted likelihood estimation method (Warm, 1989) and is subsequently used to select items that can maximally inform about the knowledge level of the child. The test items are selected using Chalmers's (2016) mirtCAT package in R. This principled selection of test items on the basis of both the child's estimated ability and the properties of the test items is expected to improve the relevance of the items sampled in the algorithm. Based on children's responses on the items administered in CATs, the next step computes estimates of their full CDI scores.

CDI Score Estimation

The method of fitting each of the item-based histograms of full CDI scores to a normal distribution and the estimation of CDI scores closely resembled those described in Mayor and Mani (2019). First, language-, gender-, and age-matched children are retrieved from Wordbank (Frank et al., 2017) using Braginsky's (2018) wordbankr package in R. Then, for each test item i that is reported by the parent as either known or not known by the child, a histogram of full CDI scores of all other children with the same response is extracted from Wordbank. The resulting histogram is fitted with a normal distribution using maximum likelihood estimation. A polynomial is fitted with the parameters (mean and standard deviation) extracted from the fitted histogram to smoothen out random fluctuations. Unlike Mayor and Mani whose polynomial fitting was conducted with a degree of three, the present model took a more flexible approach, in which the degree of the polynomial adjusts to the breadth of the distribution of the vocabulary counts.[2] Once normalized, this histogram can be seen as the probability distribution of full CDI scores given the child's response for item i. The histograms of all test items are subsequently log-summed, and we retrieve the mode of the resulting histogram. Finally, a linear transformation of the mode, ensuring that the full range of CDI values associated with language-, age-, and gender-matching children can be reached, produces the estimate of the child's full CDI score.

Real-Data Simulations

The data used in this study were from the American English (Fenson et al., 2000), Danish (Bleses et al., 2008), Beijing Mandarin (Tardif et al., 2008), and Italian (Rinaldi et al., 2019) CDI:WS, retrieved from Wordbank (Frank et al., 2017). The American English data set was categorized as *very large sized*, with more than 200 samples for each age (in months); the Danish data set was categorized as *large sized*, with between 100 and 200 samples for each age; the Beijing Mandarin data set was categorized as *medium sized*, with between 50 and 100 samples for each age; whereas the Italian data set was categorized as *small sized*, with fewer than 50 samples for each age. These versions of the CDI:WS were selected for having relatively homogeneous sample sizes across all ages.

Using real-data simulations, we compared the performance of the present model (the *IRT version*) with the original model by Mayor and Mani (2019) in estimating full CDI scores. We adopt the following standard for test acceptability, as introduced by Makransky et al. (2016): correlations above .95 with the full CDI, average *SE* below .20, and reliability coefficients ($1–SE^2$) above .96. In addition, the sum of words that are reported as known on a random selection of items from the CDI, scaled up to the full CDI size, was used as a baseline measure and compared with

[2]That is, to this end, the median absolute deviation (MAD) of productive vocabulary is computed for each age, in months. When MAD < 100, a linear function is fitted to improve generalization. When MAD > 100, a cubic polynomial is fitted to obtain a finer model.

the IRT version. Results for the original model and the baseline measure were averaged over 10 simulations, whereas those for the IRT version were based on single simulations, since the item selection process establishes individual parameters for each child, consequently constraining the selection of words for that child. In line with previous work using real-data simulations (i.e., Makransky et al., 2016; Mayor & Mani, 2019), correlations between the estimates (based on 5, 10, 25, 50, 100, 200, 400, and all items on each CDI) and the full CDI scores, average standard errors, and reliability ($1–SE^2$) were reported across different age groups and genders.

Further evaluation of the performance of the IRT version was conducted using established CDI:SFs. CDI:SF scores were obtained by summing raw vocabulary count based on responses on CDI:SF items and scaling these scores up to the instrument size to fit the range of full CDI scores. Likewise, correlation coefficients of these scores with the full CDI scores, reliability, and average standard errors were computed and compared to those obtained from the IRT version.

Results

Model Selection

Two changes were made to the original model (Mayor & Mani, 2019): the application of IRT in item selection and the flexible approach to polynomial fitting. Preliminary comparisons between correlation coefficients of the original model, the original model equipped with flexible polynomial fitting, the original model with IRT (but without flexible polynomials), and the IRT version (with both flexible polynomial fitting and IRT) are shown in Figure 1. Correlations were compared using two CDIs having different sample sizes: the very large-sized American English CDI data set and the medium-sized Beijing Mandarin CDI data set, in order to select the final model. When applied to the medium-sized data set, the combination of flexible polynomial fitting and IRT led to the largest improvements. For the very large-sized data set, the mere application of IRT improved the model the most, although performance was comparable to the level of performance attained with the maximal model (IRT and flexible polynomials). With the merit that improvements were observed with the smaller data set when both flexible polynomial fitting and IRT were applied, we selected the IRT version as the final model.

American English CDI:WS

The original model and the IRT version were used in real-data simulations on the very large-sized American English CDI:WS data set for children between 16 and 30 months of age, for each gender, using a different number of test items (5, 10, 25, 50, 100, 200, 400, and 680, the full CDI size). Figure 2 shows the correlations between the estimated scores and the full CDI scores using 5, 10, 25, 50, and 100 items, along with the average standard errors, and the reliability of both the IRT version and the original model, as well as Makransky et al.'s (2016) values for comparison.[3]

In terms of correlations, the IRT version performed better than the original model, with correlations above .9, provided that the test has more than just five items. In terms of both average standard errors and reliability, the IRT version had values similar to the original model at 100 items but outperformed the latter at 50 items and below. At 25 items, correlations greater than the .95 cutoff point, as suggested by Makransky et al. (2016), were achieved. Additional real-data simulations revealed that a correlation of .95 was already achieved at 14 items, with an average *SE* of .07 and a reliability of .995.

To further examine the effectiveness of the IRT version across ages, an analysis was conducted per age group (see Appendix Table A2). Improvements in correlations were observed for all age groups when compared to the original model. It is noteworthy that, at 25 items, correlations across all age groups were already higher than .95, whereas in the original model, the youngest age group (16–18 months) required at least 50 items to achieve correlations of .95 and above. In line with Makransky et al. (2016) and Mayor and Mani (2019), a marked reduction in performance was observed for both the youngest and oldest age groups, when the test featured less than 10 items.

Danish CDI:WS

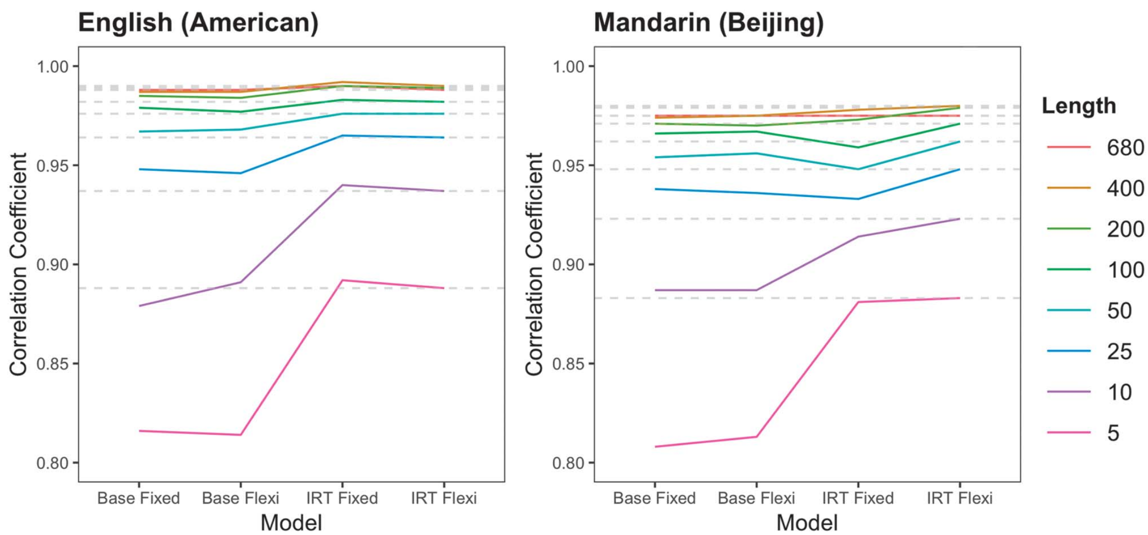
Real-data simulations were conducted using the large-sized Danish CDI:WS (Bleses et al., 2008) data set. Figure 3 depicts the performance of the IRT version and the original model for children between 16 and 30 months of age when having five to 100 items, with random lists as the baseline measure for comparisons.[4] Similar to the American English data, consistent improvements in correlations, average standard errors, and reliability were observed for the IRT version. With the Danish CDI:WS data set, the IRT version was again able to achieve correlations of above .95 at 25 items, as it did with the American English CDI:WS data set, whereas the original model required at least 50 items to achieve a similar performance. Furthermore, the IRT version had better correlations, average standard errors, and reliability than the baseline measure at 50 items and below. Additional real-data simulations revealed that a correlation of .95 was already achieved at 17 items, with an average *SE* of .06 and a reliability of .997.

Beijing Mandarin CDI:WS

Real-data simulations were run on the medium-sized Beijing Mandarin CDI:WS (Tardif et al., 2009) data set for children between 16 and 30 months of age, and the

[3]For the full list of values at all test lengths, across gender, see Appendix Table A1.
[4]For the full list of values at all test lengths, across gender, see Appendix Table A3.

Figure 1. A comparison between correlation coefficients of the original model (Base Fixed), the original model with flexible polynomial fitting (Base Flexi), the original model with item response theory (IRT; IRT Fixed), and the IRT version with flexible polynomial fitting (IRT Flexi).



results[5] are illustrated in Figure 4 for tests with 100 items and below. In terms of correlations, the IRT version performed better than the original model, with similar or better average standard errors and reliability. With a relatively smaller sample size, the IRT version achieved correlations of .95 at 25 items for male and at 50 items for female. When compared to the baseline measure, the IRT version had better correlations at 25 items and below, whereas the average standard errors and reliability were similar at 25 items and only better at 10 items and below. Additional real-data simulations revealed that a correlation of .95 was achieved at 23 items for male, with an average *SE* of .09 and a reliability of .992, and at 36 items for female, with an average *SE* of .08 and a reliability of .993.

Italian CDI:WS

Real-data simulations were conducted on the small-sized Italian CDI:WS (Caselli et al., 1995) data set for children between 18 and 30 months of age. For this particular data set, the original model (with fixed degree of polynomial fit of 3) was unable to reliably estimate scores. Thus, we could only compare the performance of the IRT version with the original model with flexible polynomial fitting. As shown in Figure 5, the IRT version outperformed the original version in terms of correlations at 25 items and below, with better or similar average standard errors and reliability.[6] Correlations of above .95 were already achieved with the IRT version starting at 25 items, whereas the original model achieved the same for females, but not for males (starting at 50 items). Correlations of the IRT version were better than the baseline measure at 50 items and below, whereas average standard errors were better at 25 items and below and reliability at 10 items and below. Additional real-data simulations revealed that a correlation of .95 was already achieved at 15 items, with an average *SE* of .08 and a reliability of .993.

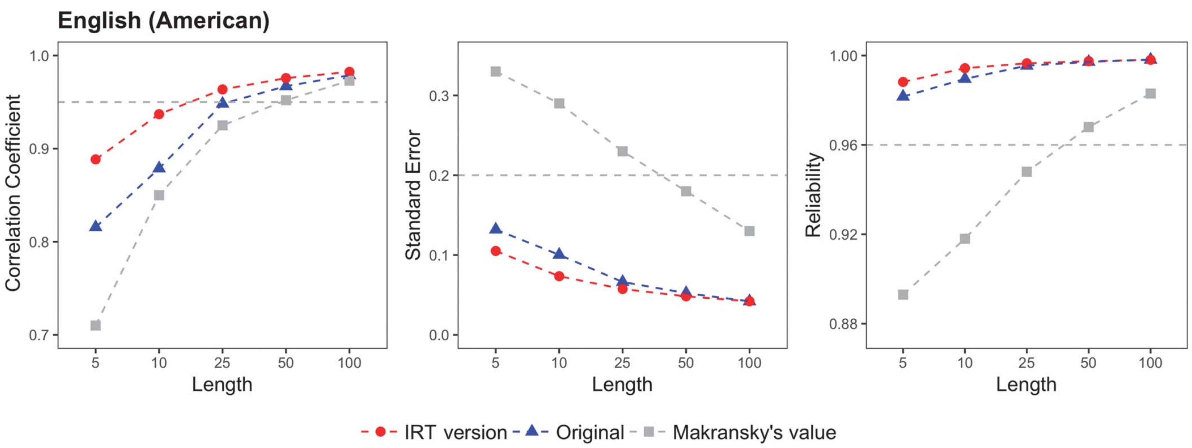
Comparisons With Established Short-Form Versions of CDIs

Using 100-item tests (110-item tests for Beijing Mandarin), the performance of the IRT version was compared with the American English (Fenson et al., 2000), Danish (Bleses et al., 2010), Beijing Mandarin (Tardif et al., 2008), and Italian (Rinaldi et al., 2019) CDI:SFs, as well as random lists as the baseline measure. For a more detailed evaluation, comparisons were made across different age groups. Overall, all three approaches met the criterion for test acceptability suggested in Makransky et al. (2016) across all age groups and CDIs.

For the very large-sized American English CDI data set, comparisons were made using Form A of the American English CDI:SF (Fenson et al., 2000). Table 1 reports correlations, average standard errors, and reliability scores for the IRT version, CDI:SF, and the baseline measure, across five age groups. The IRT version performed better than CDI:SF in terms of correlations between 16 and 24 months, whereas CDI:SF performed

[5]For the full list of values at all test lengths, across gender, see Appendix Table A4.
[6]For the full list of values at all test lengths, across gender, see Appendix Table A5.

Figure 2. Comparison of the item response theory (IRT) version and the original model with different test lengths (5, 10, 25, 50, 100) on the American English MacArthur–Bates Communicative Development Inventories: Words and Sentences, with Makransky et al.'s (2016) values for reference. The gray dashed lines at .95 on correlation, .20 on standard error, and .96 on reliability represent the cutoff points suggested by Makransky et al. The *x*-axes are not linear.



better between 25 and 30 months, though they both had similar average standard errors and reliability. The baseline measure outperformed the IRT version between 22 and 30 months as well as CDI:SF across all age groups, with better correlations, average standard errors, and reliability.

Comparisons made using the large-sized Danish CDI data set, across five age groups, among the IRT version, the Danish CDI:SF (Bleses et al., 2010), and the baseline measure are reported in Table 2. The IRT version had better correlations than CDI:SF and the baseline measure between 16 and 24 months. After 24 months, CDI:SF performed best in terms of correlations. Overall, the average standard errors and reliability of the IRT version were similar, if not slightly poorer, when compared to both CDI:SF and the baseline measure.

For the medium-sized Beijing Mandarin CDI data set, 110 items were administered in accordance with the number of items in the Beijing Mandarin CDI:SF (Tardif et al., 2008). The results reported across five age groups in Table 3 indicated poorer performance of the IRT version in terms of the correlations, except for the youngest age

Figure 3. Comparison of the item response theory (IRT) version and the original model with different test lengths (5, 10, 25, 50, 100) on the Danish MacArthur–Bates Communicative Development Inventories: Words and Sentences, with random list as the baseline measure. The gray dashed lines at .95 on correlation, .20 on standard error, and .96 on reliability represent the cutoff points suggested by Makransky et al. (2016). The *x*-axes are not linear.

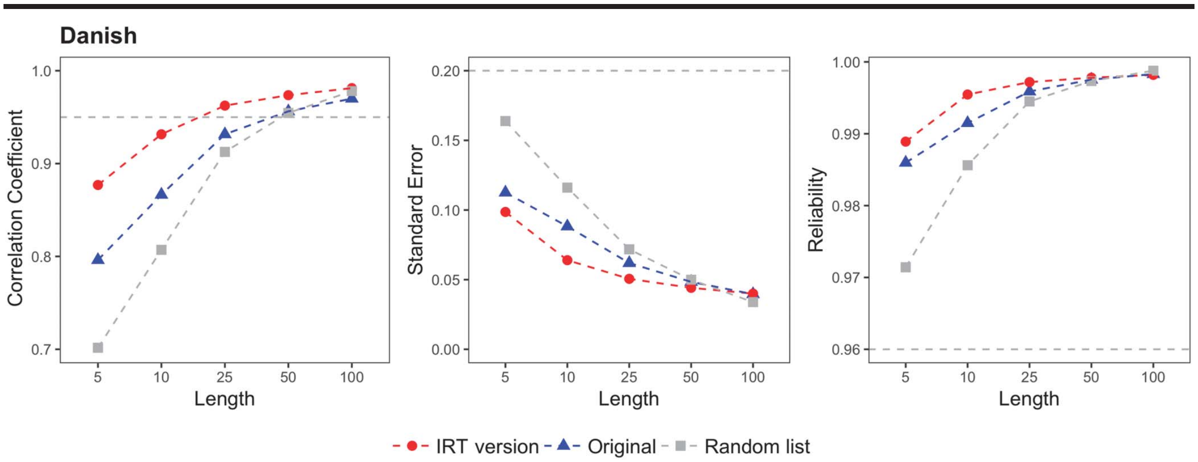
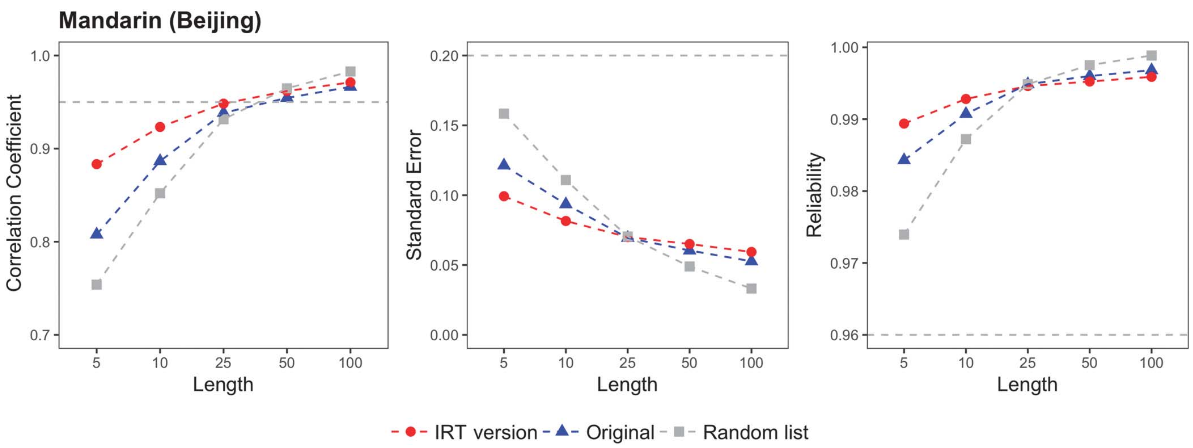


Figure 4. Comparison of the item response theory (IRT) version and the original model with different test lengths (5, 10, 25, 50, 100) on the Beijing Mandarin MacArthur–Bates Communicative Development Inventories: Words and Sentences, with random list as the baseline measure. The gray dashed lines at .95 on correlation, .20 on standard error, and .96 on reliability represent the cutoff points suggested by Makransky et al. (2016). The *x*-axes are not linear.



group, that is, 16–18 months. Average standard errors and reliability scores were also poorer than both CDI:SF and the baseline measure.

The final comparisons were made using the small-sized Italian CDI data set, among the IRT version, the Italian CDI:SF (Rinaldi et al., 2019), and the baseline measure, across four age groups. As reported in Table 4, the IRT version had better correlations than CDI:SF and the baseline measure between 18 and 24 months. Between 25 and 30 months, CDI:SF had the highest correlations. Average standard errors and reliability were similar across all three approaches.

Discussion

CDIs are a cost-effective, reliable, and valid set of parent report instruments for assessing children's early language development from 8 up to 37 months of age. However, due to their size, the administration of CDIs is time-consuming and require that parents be literate, thus

Figure 5. Comparison of the item response theory (IRT) version and the original model with flexible polynomial fitting with different test lengths (5, 10, 25, 50, 100) on the Italian MacArthur–Bates Communicative Development Inventories: Words and Sentences, with random list as the baseline measure. The gray dashed lines at .95 on correlation, .20 on standard error, and .96 on reliability represent the cutoff points suggested by Makransky et al. (2016). The *x*-axes are not linear.

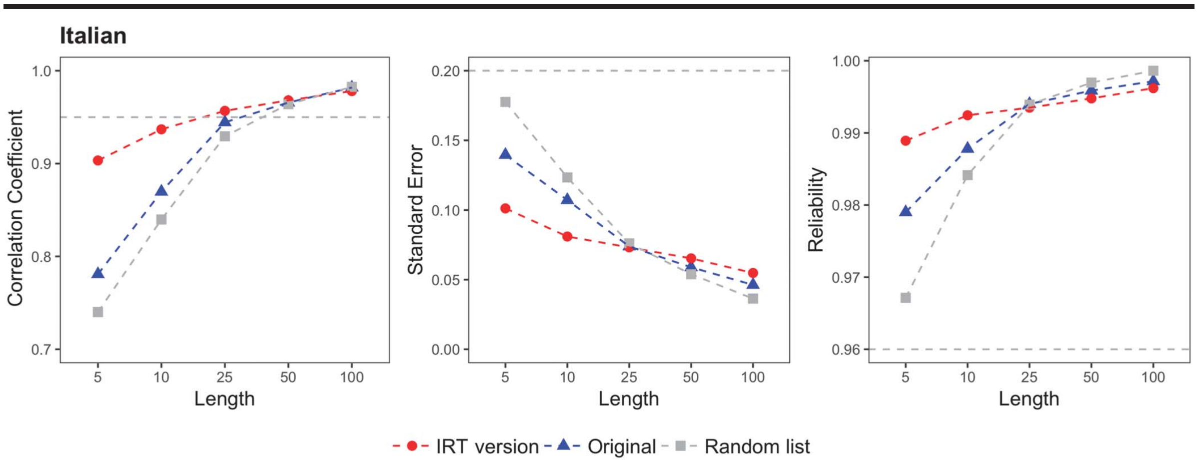


Table 1. Comparisons between the item response theory (IRT) version of the model, Fenson et al.'s (2000) American English MacArthur–Bates Communicative Development Inventories: Short Form (CDI:SF), and the baseline measure using 100 test items, by age group.

Age (months)	IRT version			CDI:SF			Baseline		
	r with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
16–18	.982	.03	.999	.954	.04	.998	.975	.03	.999
19–21	.990	.03	.999	.973	.05	.997	.985	.04	.999
22–24	.985	.04	.998	.984	.05	.997	.988	.04	.998
25–27	.978	.05	.997	.986	.06	.997	.988	.04	.999
28–30	.978	.05	.997	.985	.04	.998	.987	.04	.999

Note. Avg. *SE* = average standard error; Rel. = reliability.

restricting the applicability of CDIs when rapid assessments are desirable or when parents have low literacy skills. To deal with these drawbacks, researchers have sought to reduce the lengths of CDIs using different approaches, including developing short forms in different languages (e.g., Bleses et al., 2010; Fenson et al., 2000); administering CDIs as CAT (Makransky et al., 2016); and, more recently, estimating full CDI scores based on CDI data from language-, gender-, and age-matching children on Wordbank (Mayor & Mani, 2019).

The present approach, that is, the IRT version, combined Mayor and Mani's (2019) Bayesian-inspired approach with an IRT-based CAT that dynamically selects test items that are maximally informative based on both the child's ability and the properties of the test items (as in Makransky et al., 2016). To evaluate the IRT version, real-data simulations were conducted using four CDI:WS versions with varying sample sizes on Wordbank: American English (a very large data set; Fenson et al., 2000), Danish (a large data set; Bleses et al., 2008), Beijing Mandarin (a medium-sized data set; Tardif et al., 2008), and Italian (a small data set; Rinaldi et al., 2019). Results obtained were subsequently compared with three other approaches: Mayor and Mani's model (in a novel implementation, in R), a baseline measure (i.e., the sum of responses obtained directly from a set of items sampled randomly from the full forms), and CDI:SF. For the American English CDI:WS, Makransky et al.'s (2016) results were also included in the comparisons.

Overall, the IRT version achieved correlations with the full CDI above .95, average *SE* below .20, and reliability above .96 (a criterion for test acceptability suggested in Makransky et al., 2016) with fewer than 17 items for American English, Danish, and Italian. For the Mandarin data set, this criterion was only met from 23 items for males and 36 items for females.

To explain the uneven performance between both genders in the Mandarin data set, we further inspected the data set and found much lower variation (quantified by MAD) in the female data than in the male data. More specifically, starting from 23 months, the female data were more left-skewed than the male data; that is, a majority of females had high CDI scores, whereas males' scores continued to vary until about 27 months, when a majority, like females, began to have high CDI scores. The implication is twofold: First, it may be that, for girls, a larger sample size is needed for a better representation of the population; second, many items in the Mandarin CDI appear to be too easy, in particular, for girls older than 23 months, hence reaching a ceiling earlier than boys. Despite this exception, our results suggest that a 25-item test can reliably estimate a child's CDI scores in most cases, regardless of gender and language. Analyses conducted per age group on the American English data set extend this finding, further suggesting that a 25-item checklist is suitable for use with children across all age groups (16–30 months).

Comparisons with Mayor and Mani (2019) revealed that the IRT version had similar or better performance in

Table 2. Comparisons between the item response theory (IRT) version of the model, Bleses et al.'s (2010) Danish MacArthur–Bates Communicative Development Inventories: Short Form (CDI:SF), and the baseline measure using 100 test items, by age group.

Age (months)	IRT version			CDI:SF			Baseline		
	r with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
16–18	.986	.02	.999	.968	.02	1.000	.972	.02	1.000
19–21	.978	.05	.997	.969	.03	.999	.973	.03	.999
22–24	.990	.04	.999	.983	.04	.998	.982	.04	.998
25–27	.981	.05	.997	.984	.05	.997	.983	.04	.998
28–30	.971	.06	.997	.985	.05	.997	.98	.04	.998

Note. Avg. *SE* = average standard error; Rel. = reliability.

Table 3. Comparisons between the item response theory (IRT) version of the model, Tardif et al. (2008) Beijing Mandarin MacArthur–Bates Communicative Development Inventories: Short Form (CDI:SF), and the baseline measure using 110 test items, by age group.

Age (months)	IRT version			CDI:SF			Baseline		
	r with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
16–18	.986	.06	.995	.980	.04	.999	.979	.04	.999
19–21	.984	.05	.998	.990	.05	.998	.990	.05	.998
22–24	.963	.07	.995	.981	.04	.998	.986	.04	.998
25–27	.961	.06	.997	.979	.04	.998	.983	.04	.999
28–30	.970	.06	.996	.981	.03	.999	.976	.03	.999

Note. Avg. *SE* = average standard error; Rel. = reliability.

terms of correlations, average standard errors, and reliability, across all four CDIs and both genders, regardless of the number of test items. In other words, the scores established by the IRT version matches more closely the full CDI scores. When compared against the baseline measure, the IRT version performed better in terms of correlations, average standard errors, and reliability for all short tests, that is, having 50 items and below. It is noteworthy, however, that the baseline measure—summing a random selection of words—performed well across all four CDIs, already achieving correlations of above .95 with good average standard errors and reliability, starting at just 50 items. At 100 items, the baseline measure's performance was also comparable to CDI:SFs. Such impressive performance should be attributed to the high internal consistency of CDIs (e.g., Bleses et al., 2008; Fenson et al., 1994; Tardif et al., 2009).

The final comparisons were made with CDI:SFs and the baseline measure across age groups, with 100 test items (or 110 for the Beijing Mandarin CDI:SF). While the IRT version typically outperformed CDI:SFs in the younger age groups, that is, between 16 and 24 months (with the exception of the Mandarin CDI:SF), both the baseline measure and CDI:SFs performed better in the older age groups, that is, between 25 and 30 months. Nevertheless, the performance of the IRT version was still comparable to the baseline measure and CDI:SFs, with all three approaches meeting Makransky et al.'s (2016) suggested criterion for test acceptability across all age groups. An important point to note here is that the development of CDI:SF for even just one language is labor intensive, whereas the IRT version has the advantage of being cost-effective in that it is generalizable; that is, it can be directly applied to CDIs of any languages, as long as sufficient CDI data are available online. Crucially, our objective is to develop a brief test that allows for rapid assessments—a 100-item checklist may still be considered too long in cases requiring multiple forms to be completed (e.g., in a multilingual environment, a clinical setting) or intimidating when parents have low literacy. The IRT version*,* on the other hand, is able to provide reliable estimates with just 14–25 items, gaining a factor of 4–7 compared to CDI:SFs.

The results reported here are based on real-data simulations. A full assessment of the psychometric properties of the IRT version should be conducted with new participants, in particular, to establish its test–retest reliability and its validity using an array of validity tests. With new participants, we also expect reduced level of performance as a result of parents responding differently to the same item in the full and short forms. This was demonstrated in Mayor and Mani (2019), with parents responding more positively in both the 25- and 50-item checklists than in the full CDI. In addition, as opposed to the more structured full forms that organize items into different semantic categories, our approach presents items in a semantically unstructured order, which may in turn affect parents' response behavior. Therefore, the essential next steps include

Table 4. Comparisons between the item response theory (IRT) version of the model, Rinaldi et al.'s (2019) Italian MacArthur–Bates Communicative Development Inventories: Short Form (CDI:SF), and the baseline measure using 100 test items, by age group.

Age (months)	IRT version			CDI:SF			Baseline		
	r with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
18–21	.981	.04	.998	.972	.03	.999	.975	.03	.999
22–24	.990	.03	.999	.983	.04	.998	.982	.04	.998
25–27	.981	.04	.998	.984	.05	.997	.983	.05	.998
28–30	.971	.05	.998	.985	.05	.997	.980	.05	.998

Note. Avg. *SE* = average standard error; Rel. = reliability.

investigating the differences in parents' response behavior and validating the model on new participants.

Finally, the reliability of our generic approach relies upon the availability of CDI data from children with matching key demographics (e.g., language, age, and gender). Based on our findings, even with a small data set having less than 50 samples available online for each age group (in months), our approach is able to reliably estimate children's full CDI scores with just 25 items, effectively reducing administration time to a mere couple of minutes. Thus, it is vital that data collected on children's vocabulary be shared publicly to enable access to and reuse of these data that will allow for the establishment of computerized adaptive tests that are tailored to each child.

Acknowledgments

This work was supported by the following Malaysian grants awarded to J. M.: MOSTI-06-02-12-SF0351 and FRGS-NFHY0004.

References

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12*(5–6), 507–584. https://doi.org/10.1080/016909697386628

Bleses, D., Vach, W., Jørgensen, R. N., & Worm, T. (2010). The internal validity and acceptability of the Danish SI-3: A language-screening instrument for 3-year-olds. *Journal of Speech, Language, and Hearing Research, 53*(2), 490–507. https://doi.org/10.1044/1092-4388(2009/08-0132)

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). The Danish communicative developmental inventories: Validity and main developmental trends. *Journal of Child Language, 35*(3), 651–669. https://doi.org/10.1017/S0305000907008574

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459. https://doi.org/10.1007/BF02293801

Braginsky, M. (2018). *wordbankr: Accessing the Wordbank database* (R package Version 0.3.0) [Computer software]. https://CRAN.R-project.org/package=wordbankr

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science, 3,* 52–67. https://doi.org/10.1162/opmi_a_00026

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*(2), 159–199. https://doi.org/10.1016/0885-2014(95)90008-X

Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*(1), 69–111. https://doi.org/10.1017/S0305000998003687

CDI Advisory Board. (2015). *Adaptations in other languages.* http://mb-cdi.stanford.edu/adaptations.html

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software, 71*(5), 1–39. https://doi.org/10.18637/jss.v071.i05

Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development, 77*(3), 712–735. https://doi.org/10.1111/j.1467-8624.2006.00899.x

Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*(3), 544–560. https://doi.org/10.1044/1092-4388(2003/044)

Dale, P. S., Reznick, J. S., & Thal, D. J. (1998). A parent report measure of language development for three-year-olds. *Infant Behavior & Development, 21,* 370. https://doi.org/10.1016/S0163-6383(98)91583-1

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361–389. https://doi.org/10.1080/13682820701546854

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*(4), 759–786. https://doi.org/10.1017/S0305000905007105

Ellis Weismer, S. (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honour of Robin S. Chapman* (pp. 83–101). Erlbaum. https://doi.org/10.4324/9781315092041-3.

Embretson, S. E., & Reise, S. P. (2000). *Multivariate applications books series. Item response theory for psychologists.* Erlbaum.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–185. https://doi.org/10.2307/1166093

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S. J., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* Singular.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Brookes. https://doi.org/10.1037/t11538-000.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*(1), 95–116. https://doi.org/10.1017/S0142716400001053

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*(2), 350–365. https://doi.org/10.1037/0022-3514.78.2.350

Frank, M. C., Braginsky, M., Marchman, V. A., & Yurovsky, D. (in press). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://langcog.github.io/wordbank-book/

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Galeote, M., Checa, E., Sánchez-Palacios, C., Sebastian, E., & Soto, P. (2016). Adaptation of the MacArthur–Bates Communicative Development Inventories for Spanish children with Down syndrome: Validity and reliability data for vocabulary. *American Journal of Speech-Language Pathology, 25*(3), 371–380. https://doi.org/10.1044/2015_AJSLP-15-0007

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child and Adolescent Mental Health, 13*(4), 198–206. https://doi.org/10.1111/j.1475-3588.2008.00503.x

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur–Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*(3), 623–654. https://doi.org/10.1017/S0305000907008094

Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory–based, computerized adaptive testing version of the MacArthur–Bates Communicative Development Inventory: Words & Sentences (CDI:WS). *Journal of Speech, Language, and Hearing Research, 59*(2), 281–289. https://doi.org/10.1044/2015_JSLHR-L-15-0202

Marjanovič-Umek, L., Fekonja-Peklaj, U., & Podlesek, A. (2013). Characteristics of early vocabulary and grammar development in Slovenian-speaking infants and toddlers: A CDI adaptation study. *Journal of Child Language, 40*(4), 779–798. https://doi.org/10.1017/S0305000912000244

Mayne, A. M., Yoshinaga-Itano, C., & Sedey, A. L. (1999). Receptive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta Review, 100*(5), 29–52.

Mayne, A. M., Yoshinaga-Itano, C., Sedey, A. L., & Carey, A. (1998). Expressive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta Review, 100*(5), 1–28.

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods, 51*(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

Pan, B. A., Rowe, M. L., Spier, E., & Tamis-Lemonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language, 31*(3), 587–608. https://doi.org/10.1017/S0305000904006270

Rescorla, L., & Dale, P. S. (2013). *Late talkers: Language development, interventions, and outcomes*. Brookes.

Rescorla, L., Ratner, N. B., Jusczyk, P., & Jusczyk, A. M. (2005). Concurrent validity of the Language Development Survey: Associations with the MacArthur–Bates Communicative Development Inventories: Words and Sentences. *American Journal of Speech-Language Pathology, 14*(2), 156–163. https://doi.org/10.1044/1058-0360(2005/016)

Rinaldi, P., Pasqualetti, P., Stefanini, S., Bello, A., & Caselli, M. C. (2019). The Italian Words and Sentences MB-CDI: Normative data and concordance between complete and short forms. *Journal of Child Language, 46*(3), 546–566. https://doi.org/10.1017/S0305000919000011

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language, 34*(1), 3–23. https://doi.org/10.1177/0142723713510997

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2009). Associations between lexicon and grammar at the end of the second year in Finnish children. *Journal of Child Language, 36*(4), 779–806. https://doi.org/10.1017/S0305000908009161

Szagun, G., Stumper, B., & Schramm, S. A. (2014). *Fragebogen zur fruhkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)* [Questionnaire on Early Language Development (FRAKIS) and FRAKIS–K (Short Form Version)]. Pearson Assessment.

Tardif, T., Fletcher, P., Liang, W.-L., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language, 36*(5), 1115–1144. https://doi.org/10.1017/S0305000908009185

Tardif, T., Fletcher, P., Zhang, Z.-X., & Liang, W.-L. (2008). *The Chinese Communicative Development Inventory (Putonghua and Cantonese versions): Manual, forms, and norms*. Peking University Medical Press.

Thal, D., DesJardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur–Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Language Pathology, 16*(1), 54–64. https://doi.org/10.1044/1058-0360(2007/007)

van der Linden, W. J., & Glas, C. A. (2010). *Elements of adaptive testing*. Springer. https://doi.org/10.1007/978-0-387-85461-8.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Appendix (p. 1 of 2)

Comparisons of the IRT Version and the Original Model (in Parentheses) With Different Test Item Sizes on the CDI:WS and the Baseline

Table A1. Comparison of the item response theory version and the original model (in parentheses) with different test item sizes on the American English MacArthur–Bates Communicative Development Inventories (CDI): Words and Sentences and the baseline (random list), by gender.

	Females			Males			Baseline		
Length	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
680	.988 (.988)	.03 (.03)	.999 (.999)	.989 (.989)	.03 (.03)	.999 (.999)	1.000	.00	1.00
400	.990 (.987)	.03 (.03)	.999 (.999)	.990 (.987)	.03 (.03)	.999 (.999)	.998	.01	1.00
200	.988 (.985)	.03 (.04)	.999 (.999)	.989 (.985)	.04 (.04)	.999 (.999)	.993	.02	.999
100	.982 (.979)	.04 (.04)	.998 (.998)	.982 (.978)	.04 (.04)	.998 (.998)	.985	.04	.999
50	.976 (.968)	.05 (.05)	.997 (.997)	.976 (.966)	.05 (.05)	.997 (.997)	.967	.05	.997
25	.963 (.950)	.06 (.07)	.996 (.995)	.964 (.946)	.06 (.07)	.997 (.995)	.936	.07	.994
10	.937 (.884)	.07 (.10)	.994 (.990)	.937 (.873)	.07 (.10)	.994 (.989)	.856	.12	.985
5	.891 (.820)	.11 (.13)	.988 (.982)	.886 (.812)	.10 (.13)	.989 (.982)	.765	.17	.97

Note. Avg. *SE* = average standard error; Rel. = reliability.

Table A2. Correlations from the item response theory version and the original model (in parentheses) on the American English MacArthur–Bates Communicative Development Inventories: Words and Sentences, by age group (in months).

Length	16–18	19–21	22–24	25–27	28–30
680	.97 (.97)	.99 (.99)	1.00 (1.00)	1.00 (1.00)	.98 (.98)
400	.98 (.96)	.99 (.99)	1.00 (1.00)	.99 (1.00)	.98 (.98)
200	.99 (.96)	.99 (.99)	.99 (.99)	.99 (.99)	.98 (.98)
100	.98 (.95)	.99 (.98)	.99 (.99)	.98 (.99)	.98 (.98)
50	.98 (.94)	.99 (.97)	.98 (.98)	.97 (.98)	.97 (.96)
25	.96 (.92)	.98 (.95)	.97 (.96)	.96 (.96)	.95 (.94)
10	.92 (.81)	.95 (.87)	.95 (.90)	.94 (.90)	.92 (.89)
5	.87 (.74)	.92 (.82)	.92 (.84)	.89 (.85)	.84 (.82)

Note. Avg. *SE* = average standard error; Rel. = reliability.

Table A3. Comparison of the item response theory version and the original model (in parentheses) with different test item sizes on the Danish MacArthur–Bates Communicative Development Inventories (CDI): Words and Sentences and the baseline (random list), by gender.

	Females			Males			Baseline		
Length	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
725	.982 (.982)	.03 (.04)	.999 (.998)	.983 (.983)	.03 (.04)	.999 (.998)	1.000	.00	1.000
400	.985 (.980)	.03 (.04)	.999 (.998)	.987 (.981)	.03 (.04)	.999 (.998)	.997	.01	1.000
200	.985 (.977)	.03 (.04)	.999 (.998)	.985 (.979)	.04 (.04)	.998 (.998)	.990	.02	.999
100	.981 (.969)	.04 (.05)	.998 (.998)	.981 (.971)	.04 (.05)	.998 (.998)	.978	.03	.999
50	.974 (.957)	.04 (.06)	.998 (.997)	.974 (.956)	.05 (.05)	.998 (.997)	.955	.05	.997
25	.964 (.931)	.05 (.07)	.997 (.995)	.961 (.932)	.05 (.07)	.997 (.995)	.913	.07	.995
10	.924 (.863)	.06 (.09)	.996 (.991)	.939 (.870)	.06 (.09)	.995 (.991)	.807	.12	.986
5	.866 (.792)	.10 (.12)	.989 (.985)	.888 (.801)	.10 (.11)	.989 (.986)	.702	.16	.971

Note. Avg. *SE* = average standard error; Rel. = reliability.

Appendix (p. 2 of 2)

Comparisons of the IRT Version and the Original Model (in Parentheses) With Different Test Item Sizes on the CDI:WS and the Baseline

Table A4. Comparison of the item response theory version and the original model (in parentheses) with different test item sizes on the Beijing Mandarin MacArthur–Bates Communicative Development Inventories (CDI): Words and Sentences and the baseline (random list), by gender.

Length	Females			Males			Baseline		
	r with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
799	.976 (.976)	.05 (.06)	.997 (.994)	.974 (.974)	.04 (.05)	.997 (.997)	1.000	.00	1.000
400	.981 (.975)	.05 (.06)	.997 (.994)	.979 (.973)	.05 (.05)	.997 (.997)	.997	.01	1.000
200	.980 (.971)	.05 (.07)	.997 (.994)	.978 (.970)	.06 (.06)	.996 (.996)	.993	.02	1.000
100	.969 (.964)	.06 (.07)	.996 (.994)	.974 (.968)	.06 (.06)	.996 (.996)	.983	.03	.999
50	.957 (.950)	.06 (.07)	.995 (.993)	.967 (.959)	.07 (.07)	.995 (.995)	.965	.05	.998
25	.942 (.930)	.07 (.08)	.995 (.991)	.955 (.947)	.07 (.07)	.994 (.994)	.932	.07	.995
10	.916 (.871)	.08 (.11)	.994 (.987)	.930 (.902)	.09 (.09)	.992 (.991)	.852	.11	.987
5	.873 (.790)	.10 (.13)	.990 (.979)	.893 (.826)	.10 (.13)	.989 (.983)	.754	.16	.974

Note. Avg. *SE* = average standard error; Rel. = reliability.

Table A5. Comparison of the item response theory version and the original model with flexible approach in fitting of polynomial (in parentheses) with different test item sizes on the Italian MacArthur–Bates Communicative Development Inventories (CDI): Words and Sentences and the baseline (random list), by gender.

	Females			Males			Baseline		
Length	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.	*r* with full CDI	Avg. *SE*	Rel.
670	.993 (.993)	.02 (.03)	.999 (.998)	.996 (.996)	.03 (.03)	.997 (.999)	1.000	.00	1.000
400	.992 (.992)	.03 (.04)	.999 (.998)	.995 (.994)	.04 (.03)	.997 (.999)	.998	.01	1.000
200	.987 (.989)	.04 (.04)	.998 (.998)	.990 (.990)	.05 (.04)	.996 (.998)	.992	.02	.999
100	.976 (.983)	.05 (.05)	.997 (.997)	.981 (.981)	.06 (.05)	.996 (.997)	.983	.04	.999
50	.965 (.970)	.06 (.06)	.995 (.996)	.971 (.962)	.07 (.06)	.994 (.996)	.964	.05	.997
25	.954 (.950)	.07 (.08)	.994 (.994)	.960 (.939)	.08 (.08)	.993 (.993)	.929	.08	.994
10	.943 (.877)	.08 (.11)	.993 (.987)	.931 (.862)	.08 (.11)	.992 (.986)	.840	.12	.984
5	.912 (.797)	.10 (.15)	.990 (.976)	.895 (.765)	.10 (.16)	.988 (.973)	.740	.18	.967

Note. Avg. *SE* = average standard error; Rel. = reliability.

This is a postprint version of the article published by **Journal of Cultural Cognitive Science** in 2021,

doi: 10.1007/s41809-021-00078-5. The published version can be accessed here:

https://doi.org/10.1007/s41809-021-00078-5

Extra-linguistic modulation of the English noun-bias: Evidence from Malaysian bilingual infants and toddlers.

Jun Ho Chai[a], Hui Min Low[b], Tze Peng Wong[a], Luca Onnis[c] and Julien Mayor[d]

[a]University of Nottingham Malaysia, [b]Universiti Sains Malaysia, [c]University of Genoa, [d]University of Oslo

Author Note

Correspondence to Jun Ho Chai: junhoc94@gmail.com

ORCID

Jun Ho Chai, 0000-0003-4316-9407

Hui Min Low, 0000-0002-8595-9215

Tze Peng Wong, 0000-0001-6593-2309

Luca Onnis, 0000-0001-6843-6554

Julien Mayor, 0000-0001-9827-5421

Word count: X

Abstract

Early vocabularies typically contain more nouns than verbs. Yet, the strength of this noun-bias varies across languages and cultures. Two main theories have aimed at explaining such variations; either that the relative importance of nouns vs. verbs is specific to the language itself, or that extra-linguistic factors shape early vocabulary structures. To address this debate, the present study compares the relative distribution of verbs and nouns within the same language – English – between Malay-English and Mandarin-English bilingual infants and toddlers. The English receptive lexicons of Mandarin-English bilingual children contained more verbs than those of Malay-English bilinguals, suggesting that the noun-bias is modulated by factors external to English. We discuss the potential role of socio-cultural differences on the composition of children early vocabularies.

Keywords: Language Development; Infancy; Language specificity; Socio-cultural influences; Cross-cultural studies

This is a postprint version of the article published by **Journal of Cultural Cognitive Science** in 2021, doi: 10.1007/s41809-021-00078-5. The published version can be accessed here: https://doi.org/10.1007/s41809-021-00078-5

Extra-linguistic modulation of the English noun-bias: Evidence from Malaysian bilingual infants and toddlers.

Over the past decades, researchers have documented that young children tend to acquire nouns before verbs across a range of languages: English (Gentner, 1982), German, Italian (Caselli et al., 1995) or Spanish (Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993) and that they are able to learn novel nouns at a quicker pace than novel verbs, in English (Childers & Tomasello, 2002). Gentner (1982) suggested this noun bias to be universal, as infants may find it simpler to label concrete objects than to learn verbs, that typically describe relationships between objects. Hence, the acquisition of nouns would be a precursor to verb learning.

Yet, numerous studies have shown that noun learning does not always outperform verb learning. While multiple languages are considered "noun-friendly" (e.g., English, Spanish) – children's early vocabularies contain more nouns than verbs – studies on the acquisition of Korean, Cantonese and Mandarin as first languages have suggested these languages to be "verb-friendly" (Arunachalam et al., 2013; Chen et al., 2015; Choi & Gopnik, 1995; Tardif, Gelman, & Xu, 1999; Tardif et al., 2008) – with nouns and verbs equally prevalent in early vocabularies (Choi & Gopnik, 1995; Tardif, 1996; Chen et al., 2015, Tardif, Gelman, & Xu, 1999; Tardif, Shatz, & Naigles, 1997) – and with more verbs than among noun-friendly languages (Choi & Gopnik, 1995; Tardif et al., 2008; Tardif et al., 1997). These findings suggest that the strength of the noun bias varies across languages and cultures, rather than being a language universal phenomenon (Lavin et al., 2006).

Tardif et al. (1999) showed that 20-month-old toddlers of Mandarin-speaking mothers tend to use verbs more often whereas toddlers of English-speaking mothers tend to use nouns more often. Correspondingly, English-speaking mothers also tend to use nouns more frequently than verbs when talking to their children whereas Mandarin-speaking mothers tend to favour verbs over nouns. While this correspondence suggests infants exposed to a noun-friendly language acquire an early lexicon rich in nouns, it does not explain why some languages are noun-friendlier than others.

Researchers have suggested that attentional patterns differ across culture, hence impacting the composition of early vocabularies: Westerners tend to focus their attention on objects—typically referred to with nouns—whereas Asians tend to focus more on the relationship between objects—often described using verbs (Nisbett & Miyamoto, 2005). Additional studies provided further demonstration of culture-dependent attentional structures (Ji, Peng, & Nisbett, 2000; Nisbett, 2004). Similarly, Waxman et al. (2016) discovered that when a video was being shown, 24-month-old Chinese infants attended to more actions-related elements whereas American infants focused more on the objects involved. Such differences in attentional pattern can also modulate children's abilities to learn words. In a study by Chan, Tardif, Chen, Pulverman, Zhu and Meng (2011), Mandarin-learning infants at 18 months were better at associating novel words to actions, whereas English-learning infants did better in mapping novel words to objects. Together, these studies suggest that differing attentional patterns across cultures can manifest as a learner's bias that may be modulating infant lexical development. Such differences are external to the linguistic properties of a language – we will refer to those as *extra-linguistic* factors.

Other extra-linguistic factors may also be driving differences in early lexical composition across languages and cultures. Using the "Human Simulation Paradigm" (Gillette, Gleitman, Gleitman, & Lederer, 1999), Snedeker, Li and Yuan (2003) investigated the ability of adults to correctly guess the target words (equal proportion of nouns and verbs) from silent videos depicting play session between a mother and an infant. They showed that both English- and Mandarin-speaking adults were better at identifying nouns than verbs from silent videos of English infant-directed speech. Yet, when exposed to silent videos of Mandarin infant-directed speech, both Mandarin- and English-speakers had similar performance when identifying nouns and verbs. Snedeker et al.'s (2003) results suggest that extra-linguistic information may account for the presence of a noun-bias in the vocabularies of English-learning children, while early vocabularies of Mandarin-learning children would be more balanced.

In parallel, other researchers have argued that factors *intrinsic* to a language can also impact the early word acquisition process (e.g., morphological transparency, pronoun-dropping parameter, word order; Tardif, Shatz, & Naigles, 1997; Tardif et al., 1999; Tardif, 1996). For example, Mandarin (but not English) is a pronoun-dropping language that allows the omission of pronouns, making verbs more likely to appear at the salient front or end of a sentence, depending whether the subject or object was omitted (Tardif, Shatz, & Naigles, 1997), in turn putting the emphasis on the verb.

Italian, despite also being a pronoun-dropping language, is a noun-friendly language. Yet, its rich verbs inflections relative to nouns, favours the learning of nouns over verbs, according to Tardif, Shatz, & Naigles (1997). English, with limited noun inflections and richer verbs inflections (Gentner, 1982) can also be seen as favouring the acquisition of nouns, making English a noun-friendly language. In contrast, Mandarin hardly has any verb or noun inflections, thus reducing the asymmetry between noun and verb inflection complexity, in turn making Mandarin a verb-friendly language. Additionally, verbs in Mandarin are often enhanced by participles, further enhancing the notion of verb in the sentence.

Our brief review of the literature suggests that both language-intrinsic features and socio-cultural influences may impact children's acquisition patterns of nouns and verbs. Crucially, these language differences appear to correlate with the attentional structure displayed across cultures. Hence, a direct comparison of participant vocabularies across languages usually will not allow for an assessment of the relative contribution of language-intrinsic factors (syntax, morphology) and language-extrinsic factors (attentional patterns, extra-linguistic context) on the noun- (or verb-) bias.

One promising avenue is then to assess bilingual children, as this allows researchers to evaluate potential differences in the noun-bias of two languages within a single learning environment. Xuan and Dollaghan (2013) collected parental reports on Mandarin-English bilingual children raised in the USA. They found that the expressive lexicons in Mandarin contained more verbs than in English, while an analysis of the 50 most frequent words in English contained more nouns than in Mandarin. They concluded that the noun-bias is language-specific, as other potentially confounding factors such as socio-economic status did not vary within subject. Yet, as bilingual children in the study may have learned their languages from two different speakers (or in two learning contexts, e.g., Mandarin at home, and English in a day-care), they may have been provided with distinct extra-linguistic cues when learning both languages. Both language-intrinsic and language-extrinsic factors may be playing a role in creating differences in the lexicons of children across language groups.

To dissociate language specificity from extra-linguistic cues, bilinguals should ideally be learning both languages from the same caregiver. Chan and Nicoladis (2010) addressed this issue, as they followed longitudinally two Mandarin-English bilingual children who were learning both languages from the same person, thereby providing enhanced control over extra-linguistic cues when assessing language-specific factors in the noun bias. They found both children to use more nouns than expected from the analysis of the salient position in their input utterances, for both of their languages. The authors suggested that the parents, immigrants to Canada, were acculturating to a western style of communication, hence potentially reducing the highlighting of verbs in their non-verbal behaviour that would otherwise be observed in monolingual Mandarin-speaking parents (Snedeker, Li, & Huan, 2003). While results from Chan and Nicoladis (2010) highlighted a correspondence between the input children were exposed to and their developing lexicon, their results also suggested that the social-cultural contexts in which children are being raised (in that case, Chinese immigrants acculturating to a western culture) modulate the noun-bias in languages they are learning.

The aim of the present contribution is to further examine the role of socio-cultural influences on language acquisition patterns, in particular on the noun bias. While previous studies with bilingual children have taken the approach of assessing differences across their languages learnt within a single learning environment (e.g., Chan & Nicoladis, 2010; Xuan & Dollaghan, 2013), we adopt the distinct, and complementary, approach of considering the impact of socio-cultural differences in their learning environment when learning a common language.

To this end, we focus on the English learnt by two groups of bilingual children from two distinct socio-cultural environments – raised in Chinese-ethnic families and in Malay-ethnic families in Malaysia[1]. Malay is an Austronesian language with a morphological complexity between that of Mandarin and English; Malay relies on both noun and verb inflections of word stems to produce complex meaning, but unlike nouns, which can be used in its bare stem forms, most verbs have to be inflected in order to denote actions (Tadmor, 2009). Yet, similar to Mandarin, Malay does not inflect verbs for tenses nor nouns for marking the singular-plural distinction but relies on separate markers to do so. Consequently, we expected that language-intrinsic factors would make Malay more noun-friendly than Mandarin yet more verb-friendly than English (a secondary aim of the present study being to verify this hypothesis). Crucially, differences in the ratio of verbs to nouns in the English lexicons across these two groups of children, as indexed by parental reports, would thus suggest that extra-linguistic factors can modulate the English noun bias.

Methods

Participants

Participants were 514 Malaysian infants and toddlers (248 females and 266 males). Participants came from middle and upper-middle socio-economic backgrounds and were born full term. Since the focus of the study is not on bilingualism *per se*, but rather in comparing

[1] Note; all infants and toddlers in the study are *Malaysian*, i.e., citizens of Malaysia, a multicultural country. Some of them speak Mandarin, whereas others speak *Malay* (Bahasa Malaysia).

socio-cultural influences on language structure, we adopted a generous inclusion criterion; infants classified as bilinguals were defined as having non-zero exposure to two languages. Relative exposure to English, as reported by their parents, was similar across bilingual groups ($t(38.77) = 1.59$, $p = .12$).

All Malay learners were ethnically Malay and all Mandarin learners ethnically Chinese. 117 children were exposed to Malay only, with an age range between 7 and 45 months of age (*M* = 24.80, *SD* = 8.43); 22 were exposed to Mandarin only, with an age range between 15 and 45 months of age (*M* = 23.00, *SD* = 5.15); 297 were exposed to both Malay and English, with ages ranging from 6 to 48 months (*M* = 24.00, *SD* = 8.43); and 78 were exposed to both Mandarin and English with an age range of 7 to 45 months of age (*M* = 22.50, *SD* = 8.43). While our age range extends beyond traditional limits for the application of MacArthur-Bates Communicative Development Inventories (CDIs; Fenson et al., 1996), analyses restricted to children younger than 36 months of age led to the same pattern of results.

To increase sample size and assess the noun-friendliness of Malay in comparison to other languages, we extracted additional Mandarin and English production data from WordBank (Frank et al., 2016). 85 monolingual Mandarin-speaking Chinese infants and toddlers (48 males and 37 females) were sampled randomly from WordBank. Their ages ranged from 16 to 30 months of age (*M* = 23.25, *SD* = 4.25). 85 English-speaking monolingual American infants and toddlers (45 males and 40 females) were also sampled randomly from the Wordbank with ages ranging from 16 to 30 months (*M* = 23.07, *SD* = 4.07). The American CDI data was collected by Thal, Marchman, Tomblin, Rescorla, and Dale (2013) and Fenson and Marchman (2007). The Chinese data was collected by Tardif, Fletcher, Liang, and Kaciroti (2009). Our sample size and age range did not allow for a comparison of the noun-friendliness of these languages in comprehension (with just 10 Malay monolinguals between 12 and 16 months of age – the common age range between our sample and that of comprehension data associated with Mandarin and English in Wordbank).

Apparatus and Materials

Data was collected using a trilingual adaptation of the MacArthur-Bates Communicative Developmental Inventories (Fenson et al., 1996). This adaptation, thereafter referred to as the MCDI-M (Multilingual Communicative Development Inventories – Malaysia version), was developed by Low (2010) to assess simultaneously three languages frequently encountered in Malaysia; Mandarin, Malay and English. This adaptation contains 600 words spanning different categories, in each of the three languages. Parents were asked to assess both production and comprehension of the words on the MCDI-M. Ethics approval was granted by The University of Nottingham Institutional Ethics Board (JM190315).

Coding

Nouns were counted based on the definition of Caselli et al. (1995). Nouns were stringently defined; only noun categories that represent concrete objects (animal names, vehicles, toys, food and drink, clothing, body parts, small household items, furniture and rooms and outside things) were included. Nominal categories (names for people, people and locations) were excluded from analyses. Verbs were counted from the action words category. Table 1 reports the number of words in each category from the MCDI-M.

When performing comparisons with data from WordBank, we identified the common set of nouns and verbs across all three language data sets (Malay, Mandarin and English), as the exact set of words in a CDI modulates the ratio of verbs to nouns V/N. Words from the same taxonomic level were included for comparison (e.g., candy and sweets), while words from different hierarchical levels (e.g., cereals and cornflakes) were excluded from the final sample which included 78 verbs and 156 nouns.

--- Insert Table 1 around here ---

Analyses

Our Malaysian infants and toddlers constituted four language groups: a monolingual Malay language group, a monolingual Mandarin language group, a bilingual Malay-English group, and a bilingual Mandarin-English group. We computed a verb-to-noun ratio[2] for each participant (V/(N+V)), based on parental reports for vocabulary in both comprehension (Table 2) and production (Table 3). To account for a potential over-representation of null V/(N+V) ratios (young participants may not have verbs in their vocabularies yet - hence a V/(N+V) ratio of zero), we built zero-inflated negative binomial models (ZINBM) on the V/(N+V) ratio in early lexicons, using the *glmmTMB* function (Brooks et. al., 2017) in *R* (R Core Team, 2016). The V/(N+V) ratios were converted to percentages and rounded to the nearest integers to conduct the ZINBM. Zero-inflation models essentially decompose the analysis into two parts: the first part evaluates factors and covariates that predict the occurrence of zeros (we refer to

[2] We will loosely refer to V/(N+V) as a "verb-to-noun ratio" thereafter. V/(N+V) was preferred over V/N in order to include in the analyses infants whose early vocabularies may still not include any nouns yet (and thus where V/N would not be defined).

this as being the zero-inflated part of the model). The second part evaluates factors and covariates predicting the distribution of non-zero values (we will refer thereafter to the latter part of the analyses as the "distributional" part of the model). Since the full model with age, language group and interactions did not converge, we implemented ZINBM multiple linear regression models – with age (coded as a continuous variable) and language group (comparisons in pairs, see Q1 – Q3 for language pairs; and Tables 4 – 14 for reference groups) as fixed-effect variables for both the zero-inflation[3], and the non-zero distribution of V/(N+V) ratios.

Analyses were conducted to address the following questions:

1. Is the English V/(N+V) ratio among Malay-English bilinguals different from the English V/(N+V) ratio among Mandarin-English bilinguals? A significant difference between language groups would suggest that extra-linguistic factors can modulate the language noun bias (in our case, in English).
2. Is the Malay V/(N+V) ratio among Malay monolinguals different from that of Malay-English bilinguals? Similarly, is the Mandarin V/(N+V) of Mandarin monolinguals different from that of Mandarin-English bilinguals? Significant differences would suggest that languages of bilingual learners influence each other in terms of noun-friendliness and hence that a language noun bias can be dissimilar for monolinguals and bilinguals.

[3] Younger children are expected to know fewer verbs than older children; children learning a noun-friendly language are expected to know fewer verbs than children learning a verb-friendly language.

3. Is the Malay V/(N+V) ratio among Malay learners different from that of the Mandarin V/(N+V) ratio among Mandarin learners? A significant difference between language groups would suggest that the noun-friendliness of Malay is different from the noun-friendliness of Mandarin.

In addition, our analyses will establish how noun-friendly is Malay, when compared to both Mandarin (verb-friendly) and English (noun-friendly) using supplemental data from WordBank (Frank, Braginsky, Yurovsky, & Marchman, 2016) in Mandarin[4] and American English.

Results

A two-step procedure was applied to analyse the data in *R* (R Core Team, 2020). First, a test of zero-inflation was run on the distributions of verb-to-noun ratios using the *zero.test* function from the *vcdExtra* package (Michael, 2017). In the presence of zero-inflation, zero-inflated negative binomial models (ZINBM) were then used to investigate differences in verb-to-noun ratios between language groups. Linear mixed models (LMM) were used when distributions were not zero-inflated.

Q1 – Does the English V/(N+V) ratio among Mandarin-English bilinguals differ from the English V/(N+V) ratio of the Malay-English bilinguals?

[4] There are two versions of Mandarin CDIs in WordBank, one created in Beijing and the other in Taiwan. In the present study we have used the Beijing version as a comparison.

Our first research question was to find out whether the English V/(N+V) ratio among Mandarin-English bilinguals differed from the English V/(N+V) ratio of the Malay-English bilinguals. The tests for the presence of zero-inflation were significant for both production and comprehension ($p < .001$), hence zero-inflation models were used.

In comprehension, the zero-inflation part of the ZINBM did not reveal an effect of age ($p = .94$) but an effect of language group ($p = .01$). Mandarin-English bilingual children were more likely than Malay-English children to possess verbs in their vocabularies. Crucially, the "distributional" part of the model (when children possess at least one verb in their vocabulary) did not reveal an effect of age ($p = .15$) but revealed an effect of language group on the English V/(N+V) ratio ($p = .011$). It is noteworthy that, while the English of both bilingual groups contained more nouns than verbs (see Table 3), the English of Mandarin-English bilinguals was richer in verbs than the English of Malay-English bilinguals (see Table 4). Thus, both the zero-inflation and the distributional parts of the model indicated significant differences between groups in comprehension (see Figure 1), suggesting that extra-linguistic factors can modulate the English noun bias.

In production, the ZINBM analyses revealed that both age ($p < .001$) and language group ($p = .002$) were significant predictors of excessive zeros. Older children and Mandarin-English bilinguals were more likely to produce verbs. Once vocabularies contained verbs (as revealed by the "distributional analyses") neither language group ($p = .25$) nor age ($p = .89$) were significant predictors of the English V/(N+V) ratio (see Table 5 and Figure 2). The English of Mandarin-English bilinguals was comparable to the English of Malay-English bilinguals, in terms of the proportion of verbs in their productive vocabularies. Thus, in production the zero-inflation but not the distributional parts of the model indicated significant differences between groups. Considering comprehension and production together, there is more evidence than not suggesting that extra-linguistic factors can modulate the English noun bias.

Q2 – Do verb-to-noun ratios differ between monolinguals and bilinguals?

Our second research question was whether the verb-to-noun ratios would differ between monolinguals and bilinguals. To address this question, we compared the Malay V/(N+V) of Malay monolinguals to that of Malay-English bilinguals, and the Mandarin V/(N+V) of Mandarin monolinguals to that of Mandarin-English bilinguals.

For the Malay learners, the tests for the presence of zero-inflation were significant for both production and comprehension ($p < .001$), hence zero-inflation models were used. In comprehension, age (p = .003) but not language group (p = .82), was a significant predictor of excessive zeros. Similarly, the distributional part of the analysis revealed an effect of age (p < .001) but not of language group (p = .34), see Table 6. In other words, the noun-friendliness of Malay, in comprehension, did not differ between monolinguals and bilinguals.

As for production, age (p < .001) but not language group (p = .14), was a significant predictor of excessive zeros. In contrast, language group (p = .02) but not age (p = .17) was a significant predictor of the Malay verb-to-noun ratio, in production (see Table 7). Unexpectedly, the Malay of Malay-English bilinguals was richer in verbs than the Malay of Malay monolinguals, in production (see Table 7).

For Mandarin learners, the tests for the presence of zero-inflation were significant for production ($p < .001$) but not for comprehension ($p = 1$), hence a zero-inflation model was used for production, whereas a linear model was used for comprehension. In comprehension, age ($p = .004$) but not language group ($p = .90$), was a significant predictor of the Mandarin verb-to-noun ratio (see Table 8). The verb-friendliness of Mandarin, in comprehension, did not differ between monolinguals and bilinguals.

In production, neither age ($p = .08$) nor language group ($p = .34$), were significant predictors of excessive zeros. Similarly, neither language group ($p = .06$) nor age ($p = .56$) predicted the Mandarin verb-to-noun ratio (see Table 9). As for comprehension, the verb-friendliness of Mandarin, in production, did not differ between monolinguals and bilinguals.

Q3 – How noun-friendly is Malay (in comparison to Mandarin and English)?

Our third question aimed at establishing the noun-friendliness of Malay, first in comparison to Mandarin, using the present sample of Malaysian data, then in comparison with both Mandarin and English, with data retrieved from WordBank. The common set of verbs and nouns was used when comparing the Malaysian sample, collected with the MCDI-M, with data retrieved from WordBank.

First, we compared the Malay verb-to-noun ratio of Malay monolinguals to that of the Mandarin of Mandarin monolinguals, in our sample. Due to its more complex morphological complexity, we expected that the Malay verb-to-noun ratio to be smaller than in Mandarin, in line with the proposal that richer morphological complexity of a lexical category slows down its learning (Gardner, 1982; Tardif, Shatz, & Naigles, 1997). The tests for the presence of zero-inflation were significant for both production and comprehension data ($p < .001$), hence zero-inflation models were used. In comprehension, age ($p = .043$) but not language group ($p = .997$), was a significant predictor of excessive zeros – older participants were more likely to have verbs in their vocabulary. In contrast, language group ($p = .030$), but not age ($p = .215$), was a significant predictor of the verb-to-noun ratio (see Table 10). The Mandarin of Mandarin monolinguals was richer in verbs than the Malay of Malay monolinguals, in comprehension.

As for production, age ($p = .017$) but not language group ($p = .321$), was a significant predictor of excessive zeros. In contrast, neither language group ($p = .586$) nor age ($p = .252$) were significant predictors of the verb-to-noun ratio (see Table 11). The verb-to-noun ratio of Malay among monolinguals was comparable to that of the Mandarin of Mandarin monolinguals, in production.

Further analyses were conducted to evaluate the noun friendliness of Malay relative to both Mandarin and English (see Table 14). Given that our current sample did not include any English monolinguals, we supplemented the dataset with vocabulary data extracted from Wordbank, in both Mandarin and English. The test for the presence of zero-inflation was significant ($p < .001$), hence a zero-inflation model was used. The model revealed that age ($p < .001$) was a significant predictor of excessive zeros. Older participants were more likely to produce verbs.

The distributional part of the analysis (when vocabularies possess at least one verb) revealed an effect of age and that Malay verb-to-noun ratios were significantly larger than English ($p = .002$) but significantly lower than Mandarin ($p = .002$) (see Table 14 and Figure 3). Additional Tukey-adjusted pairwise comparison revealed that the Mandarin verb-to-noun ratios were significantly larger than English ($p < .001$). Mandarin and Malay are more verb-friendly than English, in production, while Malay was found to be less verb-friendly than Mandarin.

To examine whether Malay and Mandarin retained their relative verb-friendliness among bilinguals, we compared the Malay verb-to-noun ratio of Malay-English bilinguals to that of the Mandarin of Mandarin-English bilinguals. In comprehension, the model revealed that age (*p* = .021) but not language group (*p* = .996) was significant predictor of excessive zeros. Younger bilingual children were less likely to possess verbs in their vocabularies, yet. In contrast, both language group (p = .001) and age (p < .001) were significant predictors of the verb-to-noun ratio (see Table 12). The Mandarin of Mandarin bilinguals was richer in verbs than the Malay of Malay bilinguals, in comprehension – consistent with findings of their monolingual peers.

In production, the model revealed that age (*p* < .001) but not language group (*p* = .256) was a significant predictor of excessive zeros. Younger bilingual children were more likely not to possess verbs in their vocabularies yet. In contrast, language group (p = .006) but not age (p = .625) was a significant predictor of the verb-to-noun ratio (see Table 13). The Mandarin of Mandarin bilinguals was richer in verbs than the Malay of Malay bilinguals, in production – in contrast with findings of their monolingual peers, where Malay and Mandarin were similar in terms of noun-friendliness, but consistent with the findings of the comparisons made with Wordbank samples (see above).

Summary

In sum, our results suggest that in comprehension, but not in production, the English lexicon of Mandarin-English bilinguals is more verb-friendly than that of Malay-English bilinguals. This addressed the first question and suggests that extra-linguistic factors can modulate the noun bias. With the exception of Malay productive vocabularies, the noun bias does not appear to differ between monolinguals and bilinguals, overall (addressing the second question). Finally, we found Malay to be more verb-friendly than English yet more noun-friendly than Mandarin for both bilingual participants and monolingual participants (with the exception of productive vocabularies among monolinguals in the MCDI-M sample), addressing the third question. Overall, the pattern of results suggests that extra-linguistic factors modulate the noun bias in a unidirectional manner – modulating English rather than the children's ethnic language.

Discussion

The present study looked into verb/noun distributions in the early lexicon of bilingually exposed infants and toddlers. Our first aim was to investigate whether what is known in the literature as a language-specific noun bias (in our case English) can be modulated by language-extrinsic factors. Our analysis with young children revealed that their English was less noun-friendly in comprehension (that is, the ratio of verbs to nouns was higher) among Mandarin-English bilinguals than it was among Malay-English bilinguals. This modulation of the noun bias cannot be attributed to language-intrinsic factors, as the focus is on the same language — English.

While our results suggest a modulation of the noun-friendliness of a language due to factors external to that language, we can only speculate about the mechanisms underlying such changes. A first interpretation is that the attentional patterns in parent-infant interactions modulate the structure of early lexicons. Chinese speakers tend to analyse visual scenes in a more holistic manner than English speakers, who tend to have focal attention towards individual objects (Nisbett, 2004; Nisbett & Miyamoto, 2005; Tardif et al., 1999; Waxman et al., 2016). Holistic processing tends to reveal relations between objects, typically referred to with verbs, whereas focal attention emphasises individual objects, typically referred to with nouns. Following this stream of reasoning, infants raised in Chinese-ethnic families are exposed to attentional patterns favouring the acquisition of verbs in all languages that they are learning, including English.

Another potential explanation is that some features of a language (in our case the verb-friendliness) may bleed into the other language of a bilingual infant or toddler via processes of cross-language transfer. One could imagine that, in an immature system, languages are not properly differentiated yet, and that the structure of a language is heavily influenced by the structure of the other language being learnt by the bilingual infant (e.g., Volterra & Taeschner, 1978; Redlinger & Park, 1980; Meisel, 1989). As the ethnicity of families is confounded with the language environment (Mandarin-English children are raised in Chinese families while Malay-English children are raised in Malay families), one cannot firmly advocate between both explanations from our findings. Yet, much of the evidence in favour of a unitary language system hypothesis comes from observations of code-switching behaviour during childhood. This perspective is based on the *production* of speech, whereas our results suggest that the structure of the English lexicon changes in light of the other language infants and toddlers are exposed to, in *comprehension* too. The explanation that infants confuse both of their languages, bringing the verb/noun distributions of each language towards the other, is at odds with strong evidence that infants and toddlers discriminate languages from a very early age (e.g., Werker & Byers-Heinlein, 2008; Genesee, 1989; Meisel, 2001; Bosch & Sebastián- Gallés, 2003). Furthermore, and addressing the second question in the study, the Malay and the Mandarin verb-to-noun ratios did not differ between monolinguals and bilinguals (with the exception of productive vocabularies in Malay), suggesting an asymmetry in the noun-bias modulation: while the English lexicon changes, the other language being learnt by children remains similar. In other words, it seems to suggest that transfer occurs more in the direction from the stronger/more established L1 (here, Mandarin or Malay) to the L2 (here, English) which functions as a *lingua franca*. As such, English, as the second language, is more susceptible of socio-cultural modulations than the other languages.

A third explanation is that the characteristics of parental input, in English, differ across groups. One could expect that the English of parents of English-Mandarin bilingual children contains more verbs that the English of parents of English-Malay bilingual children. Future work will aim at identifying the source of this modulation of the noun bias by comparing the word type and token produced by the parents during parent-child interaction with the verbs and nouns compositions of their children. A recent study has evaluated the verb to noun ratio of speech from parent-child interactions in a Singaporean bilingual population. Setoh, Cheng, Bornstein & Esposito (2021) found that most Singaporean Chinese mothers either used more verbs than nouns, or used similar number of verbs and nouns, in English and in Mandarin, thus suggesting that differences in the composition of parental input could be driving differences in the composition of their child's vocabulary. However, this finding falls short of explaining how Mandarin-English bilingual adults used more verbs in English than, say, Malaysian-English bilinguals adults in the first place. Thus, any of the above-mentioned explanations (or a combination thereof) may account for the emergence of differences in the English lexicons of young and adult bilinguals.

The focus on multilinguals exposed to languages having differing compositions – in terms of verbs and nouns – opens the door to further investigation into the roots of the noun bias, and into factors that may modulate this bias. The particular set of languages spoken in Malaysia (and Singapore) analysed in the present study, with Mandarin being verb-friendly, English being noun-friendly, and Malay in-between (as addressed by the third research question in the present manuscript), offers a unique opportunity to test competing theories about the origin of the noun bias.

In sum, our results suggest that the degree of noun-friendliness of a language can be influenced by factors external to that language. We argue that a likely explanation has the culture in which an infant is raised influencing the pattern of acquisition of verbs and nouns, possibly through differential attentional structures in her learning environment. Future research will investigate the link between attentional patterns in adults of different cultural communities and the verb/noun distributions in the lexicons of their children.

References

Arunachalam, S., Leddon, E. M., Song, H.-j., Lee, Y., & Waxman, S. R. (2013). Doing more with less: Verb learning in Korean-acquiring 24-month-olds. *Language acquisition*, 20(4), 292–304.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, *46* (2-3), 217–243.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal, 9(2), 378–400. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10* (2), 159–199.

Chan, W. H., & Nicoladis, E. (2010). Predicting two Mandarin-English bilingual children's first 50 words: Effects of frequency and relative exposure in the input. *International Journal of Bilingualism*, *14*(2), 237-270.

Chan, C. C., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L., & Meng, X. (2011). English-and Chinese-learning infants map novel labels to objects and actions differently. *Developmental psychology*, *47*(5), 1459.

Chen, J., Tardif, T., Pulverman, R., Casasola, M., Zhu, L., Zheng, X., & Meng, X. (2015). English-and Mandarin-learning infants' discrimination of actions and objects in dynamic events. *Developmental psychology*, 51(10), 1501. Chicago

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental psychology*, *38* (6), 967–978. doi: 10.1037/0012-1649.38.6.967

Choi, S., & Gopnik, A. (1995). *Early acquisition of verbs in Korean: a cross-linguistic study.* (Vol. 22) (No. 3). doi: 10.1017/S0305000900009934

Fenson, L., Bates, E., Dale, P. S., Hartung, J. P., Pethick, S. J., Reilly, J. S., . . . Thal, D. J. (1996). *The MacArthur communicative development inventories*. Singular publishing Group, Incorporated.

Fenson, L., & Marchman, V. A. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. Paul H. Brookes Publishing Company.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, 1–18. doi: 10.1017/S0305000916000209

Genesee, F. (1989). Early bilingual development: one language or two? *Journal of child language*, *16* (01), 161–179.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *In S. Kuczaj (Ed.), Language development: Language, cognition, and culture.*, 301–334.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of child language*, *20* (3), 523–49.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of Personality and Social Psychological psychology*, *78* (5), 943–955. doi: 10.1037/0022-3514.78.5.943

Lavin, T. A., Hall, D. G., & Waxman, S. R. (2006). East and west: A role for culture in the acquisition of nouns and verbs. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (p. 525). Oxford: Oxford University Press.

Low, H.M. (2010). Trilingual MacArthur-Bates Communicative Developmental Inventories (Malaysia version): An investigation of early language development in mixed-language contexts of childhood multilingualism. Paper presented in Asia Pacific Conference of Speech, Language and Hearing 2010, Kuala Lumpur, Malaysia.

Michael, F. (2017). vcdextra: 'vcd' extensions and additions [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=vcdExtra (R package version 0.7-1)

Meisel, J. M. (1989). Early differentiation of languages in bilingual children. *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*, 13–40.

Meisel, J. M. (2001). The simultaneous acquisition of two first languages. *Trends in bilingual acquisition*, *1*, 11.

Nisbett, R. E. (2004). *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. New York: Free Press.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9* (10), 467–473. doi: 10.1016/j.tics.2005.08.004

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Redlinger, W. E., & Park, T.-Z. (1980). Language mixing in young bilinguals. *Journal of child language*, *7* (02), 337–352.

Setoh, P., Cheng, M., Bornstein, M. H., & Esposito, G. (2021). Contrasting lexical biases in bilingual English–Mandarin speech: Verb-biased mothers, but noun-biased toddlers. *Journal of Child Language*, 1-24.

Snedeker, J., Li, P., & Yuan, S. (2003). Cross-cultural differences in the input to early word learning. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (pp. 1094-1099). Lawrence Erlbaum Associates.

Tadmor, U. (2009). Malay-Indonesian. *The world's major languages*, 791-818.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental psychology*, *32* (3), 492.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in mandarin (putonghua) and cantonese. *Journal of Child Language*, *36* (05), 1115–1144.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "Noun Bias" in Context: A Comparison of English and Mandarin. *Child Development*, *70* (3), 620–635. doi: 10.1111/1467-8624.00045

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24* (3), 535–565. doi: 10.1017/S030500099700319X

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, 44(4), 929.

Thal, D. J., Marchman, V. A., Tomblin, J. B., Rescorla, L. A., & Dale, P. S. (2013). Late-talking toddlers: Characterization and prediction of continuing delay. *Late talkers: Language development, interventions and outcomes*, 169–201.

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of child language*, *5* (02), 311–326.

Waxman, S., Fu, X., Ferguson, B., Geraghty, K., Leddon, E., Liang, J., & Zhao, M.-F. (2016). How Early is Infants' Attention to Objects and Actions Shaped by Culture? New Evidence from 24-Month-Olds Raised in the US and China. *Frontiers in psychology*, 7, 97. doi: 10.3389/fpsyg.2016.00097

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in cognitive sciences*, 12 (4), 144–151.

Xuan, L., & Dollaghan, C. (2013). Language-specific noun bias: evidence from bilingual children. *Journal of child language*, *40*(5), 1057-1075

Table 1. Word count for each word categories (common nouns and verbs) from the MCDI-M.

Word categories	
Animals	52
Vehicles	14
Toy	12
Food and Drink	42
Clothing	23
Body parts	26
Small household items	48
Furniture and rooms	27
Outside things	49
Action words	101

Table 2. Means, standard deviations and ranges of the Malay, English and Mandarin verbs and nouns count in comprehension, across the four language exposure groups (Malay monolinguals, Mandarin monolinguals, Malay & English bilinguals and Mandarin & English bilinguals).

Ethnicity	Exposure	Language	Category	*M*	*SD*	Min	Max
Malay	Malay only	Malay	Noun	165.22	84.38	4	293
			Verb	61.19	30.94	0	101
		Mandarin	Noun	0.06	0.40	0	4
			Verb	0.01	0.09	0	1
		English	Noun	16.44	54.45	0	293
			Verb	4.13	17.75	0	101
	Malay and English	Malay	Noun	160.09	94.50	0	293
			Verb	60.78	34.35	0	101
		Mandarin	Noun	0.06	0.55	0	8
			Verb	0.01	0.14	0	2
		English	Noun	68.51	92.17	0	293
			Verb	19.45	32.68	0	101
Chinese	Mandarin only	Malay	Noun	1.82	4.15	0	17
			Verb	0.23	0.61	0	2
		Mandarin	Noun	121.09	61.42	11	247
			Verb	54.81	23.21	5	94
		English	Noun	20.95	37.40	0	159
			Verb	1.18	3.11	0	11
	Mandarin and English	Malay	Noun	0.76	4.07	0	35
			Verb	0.18	0.70	0	5
		Mandarin	Noun	102.01	96.86	0	293
			Verb	45.62	38.07	0	102
		English	Noun	96.61	102.62	0	293
			Verb	35.31	38.95	0	101

Table 3. Means, standard deviations and ranges of the Malay, English and Mandarin verbs and nouns count in production, across the four language exposure groups (Malay monolinguals, Mandarin monolinguals, Malay & English bilinguals and Mandarin & English bilinguals).

Ethnicity	Exposure	Language	Category	*M*	*SD*	Min	Max
Malay	Malay only	Malay	Noun	99.66	83.50	0	281
			Verb	32.87	32.11	0	101
		Mandarin	Noun	0.06	0.40	0	4
			Verb	0.00	0.00	0	0
		English	Noun	5.94	32.60	0	274
			Verb	0.92	9.33	0	101
	Malay and English	Malay	Noun	87.79	91.24	0	293
			Verb	31.50	34.86	0	101
		Mandarin	Noun	0.05	0.48	0	7
			Verb	0.00	0.06	0	1
		English	Noun	34.72	61.21	0	282
			Verb	7.15	18.24	0	100
Chinese	Mandarin only	Malay	Noun	1.18	2.99	0	10
			Verb	0.23	0.61	0	2
		Mandarin	Noun	55.27	65.47	0	247
			Verb	16.33	18.87	0	56
		English	Noun	11.91	21.25	0	81
			Verb	0.64	1.56	0	6
	Mandarin and English	Malay	Noun	0.31	1.77	0	15
			Verb	0.05	0.27	0	2
		Mandarin	Noun	49.49	76.85	0	276
			Verb	21.68	31.89	0	98
		English	Noun	61.51	88.96	0	276
			Verb	18.88	31.93	0	100

Table 4. Estimates of the zero-inflated negative binomial model for the English verb-to-noun ratio between both language groups, Malay-English and Mandarin-English bilinguals, in comprehension.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.42	.13	26.86	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.20	.08	-2.55	.011
Age (months)	-.01	.00	-1.44	.149
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-2.87	.77	-3.72	< .001
Language group (Mandarin-English bilinguals used as ref.)	1.61	.62	-2.55	.009
Age	-.001	.02	-.08	.935

Table 5. Estimates of the zero-inflated negative binomial model for the English verb-to-noun ratio between both language groups, Malay-English and Mandarin-English bilinguals, in production.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.08	.24	12.71	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.13	.11	-1.15	.252
Age (months)	-.001	.01	-.15	.885
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	.65	.75	-.87	.382
Language group (Mandarin-English bilinguals used as ref.)	1.51	.49	-3.07	.002
Age (months)	-.11	.03	-4.04	< .001

Table 6. Estimates of the zero-inflated negative binomial model for the Malay verb-to-noun ratio between both language groups, Malay monolinguals and Malay-English bilinguals, in comprehension.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.54	.06	63.95	< .001
Language group (Malay-English bilinguals used as ref.)	-.03	.03	-.95	.343
Age (months)	-.01	.00	-3.70	< .001
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-.18	1.03	-.17	.862
Language group (Malay-English bilinguals used as ref.)	-.18	.82	-.22	.823
Age (months)	-.17	.06	-2.97	.003

Table 7. Estimates of the zero-inflated negative binomial model for the Malay verb-to-noun ratio between both language groups, Malay monolinguals and Malay-English bilinguals, in production.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.15	.10	30.79	< .001
Language group (Malay-English bilinguals used as ref.)	-.12	.05	-2.35	.019
Age (months)	.01	.004	1.38	.169
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	1.65	.71	2.31	.021
Language group (Malay-English bilinguals used as ref.)	-.58	.39	-1.48	.139
Age (months)	-.15	.03	-4.61	< .001

Table 8. Estimates of the linear model for the Mandarin verb-to-noun ratio between both language groups, Mandarin monolinguals and Mandarin-English bilinguals, in comprehension.

	Estimate	*SE*	*t*	*p*
Intercept	45.79	4.47	10.23	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.38	3.05	-.12	.902
Age (months)	-.54	.18	-2.99	.004

Table 9. Estimates of the zero-inflated negative binomial model for the Mandarin verb-to-noun ratio between both language groups, Mandarin monolinguals and Mandarin-English bilinguals, in production.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.62	.26	14.04	< .001
Language group (Mandarin-English bilinguals used as ref.)	-.24	.13	-1.87	.061
Age (months)	-.01	.01	-.58	.563
Zero-inflation	**Estimate**	***SE***	***z***	***p***
Intercept	1.63	2.31	.71	.481
Language group (Mandarin-English bilinguals used as ref.)	.87	.90	.96	.337
Age (months)	-.18	.10	-1.77	.077

Table 10. Estimates of the zero-inflated negative binomial model for the verb-to-noun ratio between Malay in Malay monolinguals and Mandarin in Mandarin monolinguals, in comprehension.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.44	.11	30.68	< .001
Language group (Malay monolingual used as ref.)	.15	.07	2.17	.030
Age (months)	-.01	.004	-1.24	.215
Zero-inflation part	Estimate	*SE*	*z*	*p*
Intercept	9.31	5.66	1.65	.100
Language group (Malay monolingual used as ref.)	-15.43*	3798.86*	-.004	.997
Age (months)	-.75	.37	-2.03	.043

Table 11. Estimates of the zero-inflated negative binomial model for the difference of verb-to-noun ratio between Malay in Malay monolinguals and Mandarin in Mandarin monolinguals, in production.

Distributinal part	Estimate	*SE*	*z*	*p*
Intercept	3.04	.21	14.75	< .001
Language group (Malay monolingual used as ref.)	-.07	.12	-.55	.586
Age (months)	.01	.01	1.15	.252
Zero-inflation part	Estimate	*SE*	*z*	*p*
Intercept	1.95	1.57	1.24	.214
Language group (Malay monolingual used as ref.)	-.74	.74	-.99	.321
Age (months)	-.15	.06	-2.39	.017

Table 12. Estiamtes for the zero-inflated negative binomial model for the difference of verb-to-noun ratio between Malay in Malay-English bilinguals and Mandarin in Mandarin-English bilinguals, in comprehension.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.59	.06	62.37	< .001
Language group (Malay-English bilinguals used as ref.)	.14	.04	3.23	.001
Age (months)	-.01	.002	-4.36	< .001
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-.82	1.10	-.74	.458
Language group (Malay-English bilinguals used as ref.)	-17.01*	3264.22*	-.01	.996
Age (months)	-.13	.06	-2.31	.021

* Large estimate and *SE* due to lack of zero inflation in the Mandarin-English bilingual groups.

Table 13. Estimates of the zero-inflated negative binomial model for the difference of verb-to-noun ratio between Malay in Malay-English bilinguals and Mandarin in Mandarin-English bilinguals, in production.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept	3.42	.12	27.67	< .001
Language group (Malay-Englis h bilinguals used as ref.)	.18	.06	2.75	.006
Age (months)	-.001	.004	-.49	.625
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept	-.96	.99	-.97	.331

Language group (Malay-English bilinguals used as ref.)	-.73	.64	-1.14	.256
Age (months)	-.15	.03	-4.31	< .001

Table 14. Estimates of the zero-inflated negative binomial model for the verb-to-noun ratio between Malay in Malay monolinguals, Mandarin in Mandarin (Beijing) monolinguals from Wordbank and English in American English monolinguals from Wordbank, in production.

Distributional part	Estimate	*SE*	*z*	*p*
Intercept (Malay Monolingual)	3.22	.12	26.04	< .001
Mandarin (Beijing)	.15	.05	3.02	.002
English (American)	-.17	.05	-3.14	.002
Age (months)	.02	.005	4.97	< .001
Zero-inflation part	**Estimate**	***SE***	***z***	***p***
Intercept (Malay Monolingual)	2.67	1.71	1.56	.119
Mandarin (Beijing)	-18.42*	3356.68*	-.01	.995
English (American)	-.91	.64	-1.43	.153
Age (months)	-.22	.08	-2.72	.007

* Large estimate and *SE* due to lack of zero inflation in the Mandarin groups.

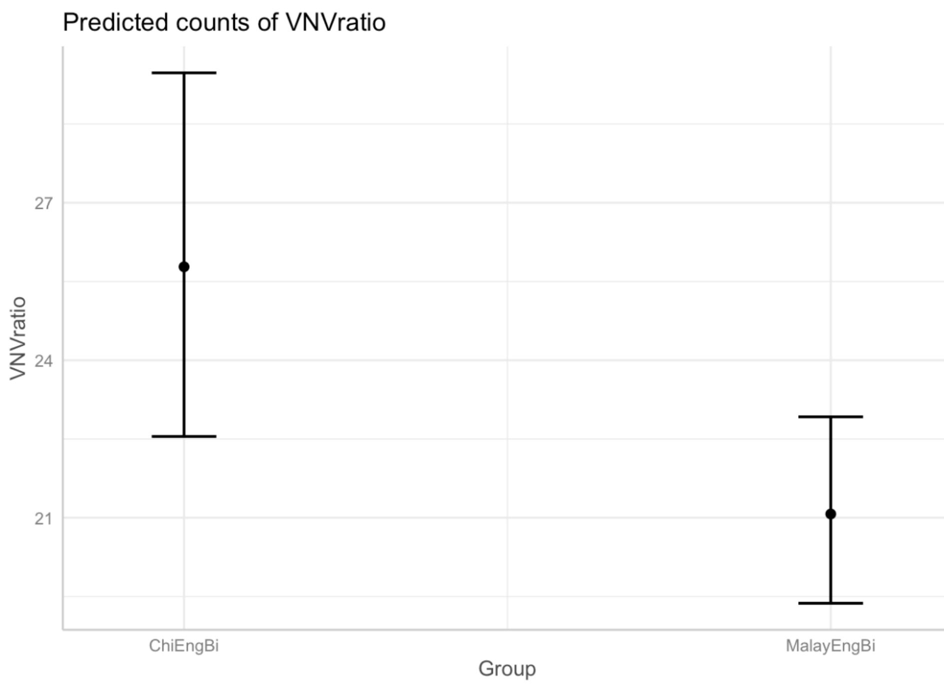


Figure 1. Predicted English verb-to-noun percentage ratios in comprehension for Malay-English and Mandarin-English bilingual children using zero-inflation negative binomial model, by group. Note that the reporting of sample means were replaced with predicted means stemming from the model to account for the other factors in the model.

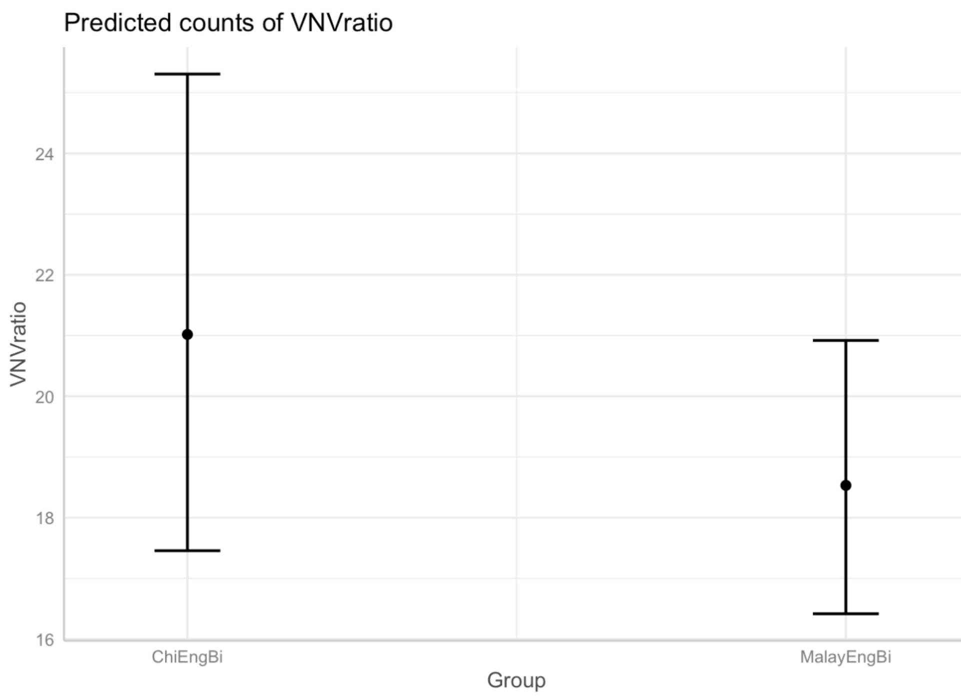


Figure 2. Predicted English verb-to-noun percentage ratios in production for Malay-English and Mandarin-English bilingual children using zero-inflation negative binomial model, by group. Note that the reporting of sample means were replaced with predicted means stemming from the model to account for the other factors in the model.

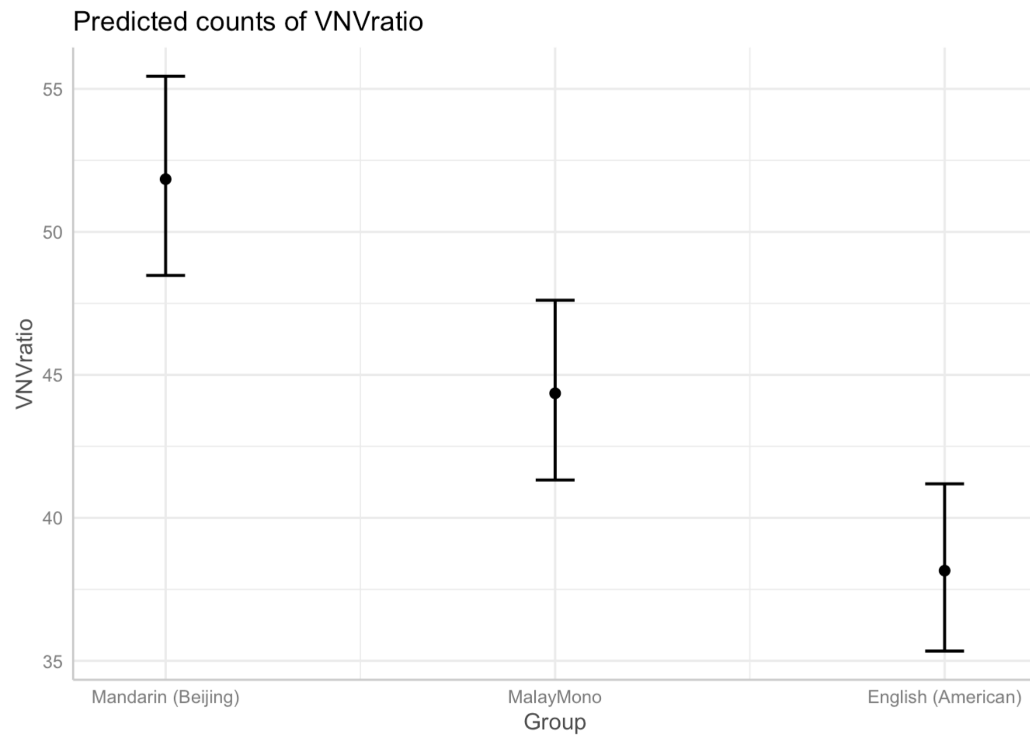


Figure 3. Predicted English verb-to-noun percentage ratios (and .95 confidence intervals) for English (American), Malay for Malay monolinguals and Mandarin for Mandarin (Beijing) children using zero-inflation negative binomial model, in production. The reporting of sample means were replaced with predicted means stemming from the model to account for the other factors in the model. Note that the verb-noun ratios are higher here due to the restricted set of common words used to allow direct comparisons with data from WordBank.

Received: 31 October 2020 | Accepted: 31 January 2021

DOI: 10.1111/infa.12401

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers

Chang Huan Lo[1] | **Audun Rosslund**[2] | **Jun Ho Chai**[1] | **Julien Mayor**[3] | **Natalia Kartushina**[2]

[1]School of Psychology, University of Nottingham Malaysia, Semenyih, Malaysia

[2]Department of Psychology & Center for Multilingualism in Society across the Lifespan (MultiLing), University of Oslo, Oslo, Norway

[3]Department of Psychology, University of Oslo, Oslo, Norway

Correspondence
Audun Rosslund, Center for Multilingualism in Society across the Lifespan, University of Oslo, PO Box 1102 Blindern, 0317 Oslo, Norway.
Email: audun.rosslund@iln.uio.no

Funding information
This work was supported by the following Malaysian grants awarded to J.M.: MOSTI-06-02-12-SF0351. The work of A.R. and N.K. was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223265.

Abstract
The present study explores the viability of using tablets in assessing early word comprehension by means of a two-alternative forced-choice task. Forty-nine 18–20-month-old Norwegian toddlers performed a touch-based word recognition task, in which they were prompted to identify the labeled target out of two displayed items on a touch-screen tablet. In each trial, the distractor item was either semantically related (e.g., dog–cat) or unrelated (e.g., dog–airplane) to the target. Our results show that toddlers as young as 18 months can engage meaningfully with a tablet-based assessment, with minimal verbal instruction and child–administrator interaction. Toddlers performed better in the semantically unrelated condition than in the related condition, suggesting that their word representations are still semantically coarse at this age. Furthermore, parental reports of comprehension, using the Norwegian version of the MacArthur–Bates Communicative Development Inventories, predicted toddlers' performance, with parent–child agreement stronger in the semantically unrelated condition, indicating that parents declare a word to be known by their child if it is understood at a coarse representational level. This study provides among the earliest evidence that

Chang Huan Lo and Audun Rosslund share first authorship.

remote data collection in 18-20 month-old toddlers is viable, as comparable results were observed from both in-laboratory and online administration of the touchscreen recognition task.

1 | INTRODUCTION

Historically, studies of early language development involved observations of children's spontaneous speech while interacting with their parents or an experimenter/clinician (Clark, 1974). Despite this method's undeniable appeal of ecological validity, the process of collecting, transcribing, and analyzing spontaneous language samples is labor-intensive and time-consuming.

To go beyond these limitations, researchers have turned to indirect assessment methods—parental reports—that provide insights into their child's communicative–linguistic development. Parental reports systematically utilize parents' extensive experience with their children, and thus allow for the collection of data that is not just more extensive than what can typically be collected during a brief laboratory or clinical session, but that might also be more representative of children's abilities (Fenson et al., 2000). Furthermore, the application of parental reports, such as the widely-used MacArthur–Bates Communicative Development Inventories (CDIs) in cross-linguistic studies, has provided invaluable insight into infants' and toddlers' early lexical development (Bleses et al., 2008; Braginsky et al., 2019; Frank et al., 2021), while other studies have evinced predictive relationships between early vocabulary and subsequent academic outcomes (e.g., Bleses et al., 2016; Duff et al., 2015; Morgan et al., 2015).

Yet, concerns have been raised regarding the sole use of parental reports, in particular when it comes to the assessment of comprehension (more than for production), since parents can at best infer comprehension based on infants' and toddlers' non-verbal responses to language (Feldman et al., 2000; Houston-Price et al., 2007; Tomasello & Mervis, 1994). For instance, while, on a "general level," previous studies have found moderate to strong correlations between average parental reports on the CDI and direct measures of infants' and toddlers' word knowledge (Fernald & Marchman, 2012; Fernald et al., 2006; Friend et al., 2012; Hurtado et al., 2008), on an "item-level," the evidence is mixed. For example, studies using indirect, eye-tracking measures revealed both underestimation (Houston-Price et al., 2007) and alignment (Styles & Plunkett, 2009; Syrnyk & Meints, 2017) between parental reports and child comprehension operationalized as visual gaze preference; studies using direct measures (i.e., child's overt answer, a touch response), on the other hand, for example, Friend et al. (2012) and Friend and Zesiger (2011), reported moderate item-level agreement.

Inconsistencies between parental reports and direct measures of child word comprehension might reflect immaturity of children's early lexical-semantic representations, which makes it challenging for parents to pin down whether a child knows a given word. Previous research has shown that early word representations are (semantically) coarse and infants and toddlers use a number of cues to disambiguate words, rather than a one-to-one word–object mapping. For instance, at 6 months of age, infants fail to disambiguate semantically/functionally related items (Bergelson & Aslin, 2017a), and at 8 months, they struggle to disambiguate items matched for frequency in child-directed speech (Kartushina & Mayor, 2019). Although word–object mappings undergo a progressive development through learning, and semantic specificity sharpens by 18–20 months of age (Bergelson & Aslin, 2017b), early word representations remain fragile by the end of the second year (Arias-Trejo & Plunkett, 2010). Arias-Trejo and Plunkett have shown that 18–24-month-olds failed to disambiguate items that were both perceptually and semantically related (e.g., an apple and an orange), as compared to semantically related but perceptually dissimilar items only (e.g., an apple and a banana), indicating that the presence of a perceptually similar distractor increases the burden of visual discrimination and feature overlap for semantically related objects.

Imprecision of parental reports may have implications when such instruments are used as measures in research or as a basis for decisions in clinical settings (Yoder et al., 1997). For these reasons, the use of supplemental measures to parental reports is encouraged (Dale et al., 2003; Fenson et al., 1993), and further assessment of their validity is needed.

A direct language measure can serve both as a convergent and a supplemental measure of parental reports. While many structured tests, such as the Peabody Picture Vocabulary Test (Dunn, 2018) and the Expressive Vocabulary Test (Williams, 2018), are available to assess young children's vocabulary knowledge, direct measures that are appropriate for assessing children below two years of age remain scarce, due to the inherent difficulty in maintaining infants' and toddlers' interest and attention (Friend & Keplinger, 2003) as well as behavioral non-compliance (Kaler & Kopp, 1990). Whereas looking-based measures, such as the Intermodal Preferential Looking Paradigm (Golinkoff et al., 1987; Hirsh-Pasek & Golinkoff, 1996) and the Looking-while-listening task (Fernald et al., 1998, 2006), have been successfully used with infants as young as 4 months old by eliminating the need for a volitional response (Golinkoff et al., 2013), the passive and repetitive nature of such measures may quickly lead to boredom among older toddlers, thus making an extensive assessment impracticable. The Computerized Comprehension Task (CCT; Friend & Keplinger, 2003), on the other hand, is a reliable and valid touchscreen-based measure designed specifically for assessing comprehension among toddlers between 16 and 24 months of age and has been shown to be effective in maintaining children's attention as well as improving compliance (Friend & Keplinger, 2003, 2008; Friend et al., 2012; Friend & Zesiger, 2011; Hendrickson et al., 2015; Poulin-Dubois et al., 2013).

Following the approach of the CCT—in providing an engaging direct language assessment—the present study explores the viability of tablets in assessing toddlers' word comprehension by means of a word recognition task, with the following three objectives. First, despite tablets and apps being increasingly commonplace among children of all ages, the use of tablet-based assessments has been primarily limited to adults and older children. Given that tablets are easy to operate even for the youngest children and additionally, given children's increasing proficiency with tablets (Abdul Aziz et al., 2014; Marsh et al., 2015), there is a need to examine how such devices can be used most effectively to collect child language data. Neumann et al. (2019), for instance, demonstrated that a tablet-based assessment could provide a valid and reliable measure of early literacy skills, at least among the older children ($n = 45$, $M_{age} = 4.65$) tested in their study. Twomey et al. (2018) further showed that children as young as 24 months old were able to complete a tablet-based assessment of early cognitive functions.

Second, compared to traditional paper-and-pencil tests, tablet-based assessments provide a testing situation that is more engaging and motivating. While the CCT offers the same advantage, the assessment is typically administered in laboratories, where screens are often mounted on a wall or placed on a desk and thus require full arm movements, which may in turn, lead to fatigue in longer sessions (Frank et al., 2016). In contrast, tablet-based assessments require only minimal motor movements and are much more portable due to the small form factor of tablets.

Third, there is a need to further evaluate the alignment between parental reports and children's word comprehension, and, in particular, to assess whether parental evaluations fit best their toddlers' word recognition in coarse (the semantically unrelated condition) or finer-grained contexts (the semantically related condition). Children vary in the strength of their word knowledge at the item-level and capturing this variability is important for a robust understanding of a child's lexical development.

In order to explore the viability of using a tablet-based measure in assessing early word comprehension and to examine the role of semantic relatedness in early word recognition, the present study employed a two-alternative forced-choice (2AFC) word recognition paradigm (similar to the CCT) with Norwegian toddlers aged between 18 and 20 months. As the CCT is only available in three languages (i.e., English, Spanish, and French), lexical items were selected from the Norwegian adaptation of the CDI–Words and Gestures (CDI–WG; Simonsen et al., 2014), with varying levels of difficulty (defined based on the normative data). Within each trial, toddlers saw on a screen two images: one representing the lexical target, and the other representing the distractor. In contrast to the CCT, in which only semantically related item pairs were used, the current design examined the role of semantic relatedness on toddlers' performance in the word recognition task, by pairing the lexical target with a distractor belonging to a different semantic category (e.g., a car and a cat) and with another distractor belonging to the same semantic category (e.g., a car and an airplane). It was expected that toddlers, in the current study, would be more accurate in semantically unrelated than related trials. Based on previous work using the CCT (Friend & Keplinger, 2003, 2008), accuracy was also expected to mirror the a priori difficulty levels, with accuracy decreasing with increasing difficulty. Finally, if parental reports are an accurate predictor of toddlers' word knowledge, a positive relationship between parent-reported comprehension and toddler's accuracy in word recognition was expected.

2 | METHOD

2.1 | Participants

Parents of 49 monolingual (>75% exposure) Norwegian toddlers (aged between 18 and 20 months) from the Greater Oslo Region, Norway, were contacted to participate in the current study through social media, leaflets distributed in a kindergarten, postal mailing lists, and email lists. After consenting to participate in the study, parents completed the Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) online within one week prior to the study so that the current estimates of their child's vocabulary size could be obtained. Parents' socioeconomic status (SES), indicated by mother's highest education level, ranged from 0 (primary school) to 5 (doctoral degree), with the mean score 3.57 ($SD = 0.82$).

All recruited toddlers were full-term at birth, had no hearing or visual impairments, and had Norwegian as their native language. Toddlers participated in the study in one of three settings: the BabyLing laboratory, a municipal kindergarten, and online (i.e., at toddlers' own homes).[1] In both the laboratory and the kindergarten settings, toddlers were tested by an experimenter, whereas online, toddlers were tested by their parents.[2] Thus, for simplicity, both the laboratory and kindergarten samples ($n = 21$; 16 females, 5 males) were categorized under the *lab*oratory setting, and the online sample ($n = 28$; 15 females, 13 males), the *online* setting. An additional 11 participants had to be excluded for failing to complete the task ($n = 7$; 2 laboratory and 5 online) and for attempting the task more than once ($n = 4$; all online). Mean age, age range, and standard deviation for each setting are detailed in Table 1.

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or a guardian for each child before any assessment or data collection. The study was approved by the ethics committee at the Department of Psychology, University of Oslo and by the Norwegian Centre for Research Data (NSD, ref. 807456).

[1]Data were initially collected in the laboratory and kindergarten. Due to the COVID-19 pandemic-related lockdown in Norway (Klesty & Fouche, 2020), data collection proceeded online.

[2]Parents consented to not to interfere with the task or influence their child's responses.

TABLE 1 Age mean, standard deviation, and range for laboratory and online settings

Setting	M_{age} (months)	SD_{age} (months)	$Range_{age}$ (months)
Laboratory	19.29	0.60	17.91–20.30
Online	19.63	0.63	18.60–20.60

2.2 | Design

The present study used a within-subjects design. Toddlers' comprehension of 24 lexical items of three levels of difficulty (easy, moderately difficult, and difficult; see Lexical Items section, below) was assessed using a tablet-based 2AFC word recognition task. Lexical targets were assessed under two conditions: semantically related (i.e., the lexical target was presented with a distractor from the same semantic category) and semantically unrelated (i.e., the lexical target was presented with a distractor from a different semantic category).

2.3 | Apparatus and materials

The study was conducted via a custom-based online experimental platform developed by Lo et al. (2021). In the laboratory setting, a Samsung Galaxy Tab S4 was used to run the study, whereas in the online setting, parents' own touchscreen devices were used. The Norwegian adaptation of the CDI–WG (Simonsen et al., 2014) was used as a measure of vocabulary size.

2.3.1 | Lexical items

Four highly familiar lexical items were selected for the familiarization phase: "ball" [ball], "hus" [house], "sko" [shoe], and "tre" [tree]. For the test phase, a total of 24 lexical items were selected. Each lexical target was assessed twice, by pairing its referent with semantically related and unrelated referents as distractors. Item pairs varied in difficulty (defined a priori on the basis of the Norwegian CDI–WG normative data for 20-month-olds; Frank et al., 2017; Simonsen et al., 2014) and were comprised of an equal number of easy (comprehended by more than 80% of the normative sample), moderately difficult (comprehended by 40%–80% of the normative sample), and difficult (comprehended by less than 40% of the normative sample) items. Within each level of difficulty, there was also an equal representation of animate and inanimate referents. The list of item pairs is provided in Table 2.

2.3.2 | Visual and auditory stimuli

To remove potential biases due to familiarity effects (from assessing the same item twice), visual stimuli for the test phase included 48 images of prototypical referents (as reported by 2 adults in a separate stimuli assessment) for the 24 lexical items assessed (i.e., two images for each item). The set of images used can be found in Appendix 1 (see also Appendix 2 for the images used in the familiarization phase). Within each item pair, the side (left or right) on which a referent appeared was counterbalanced. All auditory stimuli were recorded by a female native speaker of Norwegian in

TABLE 2 Item pairs

Difficulty level	Semantically related	Semantically unrelated
Easy	bil [car]—fly [airplane]	hest [horse]—banan [banana]
	eple [apple]—banan [banana]	hund [dog]—fly [airplane]
	hest [horse]—ku [cow]	katt [cat]—bil [car]
	hund [dog]—katt [cat]	ku [cow]—eple [apple]
Moderate	elefant [elephant]—tiger [tiger]	elefant [elephant]—saks [scissors]
	lastebil [truck]—tog [train]	løve [lion]—tog [train]
	saks [scissors]—blyant [pencil]	sjiraff [giraffe]—lastebil [truck]
	sjiraff [giraffe]—løve [lion]	tiger [tiger]—blyant [pencil]
Difficult	elg [moose]—pingvin [penguin]	elg [moose]—pasta [pasta]
	gås [goose]—ugle [owl]	gås [goose]—shorts [shorts]
	pasta [pasta]—sukkertøy [candy]	pingvin [penguin]—sukkertøy [candy]
	shorts [shorts]—glidelås [zipper]	ugle [owl]—glidelås [zipper]

child-directed speech and then processed in Praat (Boersma & Weenink, 2020) to remove noise and equalize intensity across the 24 prompts.

2.4 | Procedure

The study began with an introductory phase, followed by a familiarization phase and a test phase.

2.4.1 | Introductory phase

Before the familiarization phase began, a smiley face was presented at the center of the screen with an introductory audio "Hei! Har du lyst til å spille?" [Hi! Do you want to play?] to attract participants' attention. In order to proceed to the familiarization phase, the experimenter/parent had to tap on the "Next" button at the bottom-right corner of the screen.

2.4.2 | Familiarization phase

The familiarization phase consisted of four 2AFC trials to (a) ensure that participants understood the context of the task and (b) familiarize them with the tapping paradigm. In each trial, participants were presented with a pair of highly familiar objects (placed on the left and right sides of the screen respectively) and prompted to tap on the referent for the heard lexical target X embedded in the carrier phrase "Kan du trykke på X?" [Can you touch the X?] Tapping was disabled for the first 2000 ms from the onset of the trial to prevent impulsive responses during the audio prompt that lasted between 1500 and 2000 ms. The timeout was 8000 ms (comparable to Friend et al., 2012), to accommodate for considerable individual variation in response times (see Ackermann et al., 2020). As soon as a (touch) response was provided, the next trial was presented.

2.4.3 | Test phase

Before the test phase began, a smiley face was again presented at the center of the screen, accompanied by an audio with an encouraging phrase "Da fortsetter vi!" [Let's continue!] The experimenter/parent had to tap on the "Next" button to begin the test phase.

The test phase consisted of 48 2AFC trials, in which each lexical target was assessed twice (paired with either a semantically related distractor or a semantically unrelated distractor). In each trial (see Figure 1 for a screenshot), participants were presented with an item pair (see Table 2) and prompted to tap on the referent for the heard lexical target X (see carrier phrase from the familiarization phase). Each item pair was presented twice so that each item within the pair served as a target and a distractor in an equal number of trials. As with the familiar trials, tapping was disabled for the first 2000 ms of the trial (to prevent participants from providing responses during the audio prompt that lasted between 1500 and 2000 ms), after which participants were given 8000 ms to respond until the subsequent trial was presented. Trials were presented in a random order, with three breaks interspersed throughout the test phase. During each break, a smiley face was presented in the same manner as before, accompanied by one of the following encouraging phrases: (a) "Da fortsetter vi!" [Let's continue!], (b) "Nå går vi videre!" [Now, we move on!], (c) "Da har vi den neste!" [Then, we have the next (one)!], and (d) "Da er du nesten ferdig! Bra!" [You're almost done! Good!] In order to continue with the test, the experimenter/parent had to also tap on the "Next" button at the bottom-right corner of the screen. Upon completion of the test phase, the smiley face was once again presented, accompanied by an audio with the phrase "Nå er du ferdig! Kjempebra!" [Now you're done! Great!].

3 | RESULTS

The results are organized around three central questions. First, potential differences between data collected online and in-laboratory were considered. Second, the influence of semantic relatedness and difficulty of item pairs on toddlers' motivation to produce a response as well as on their performance in the word recognition task were examined. Finally, the convergent relation between toddlers' performance and parental report (CDI–WG) was assessed. In accordance with previous work using the CCT (Friend & Keplinger, 2003; Friend et al., 2012), missing responses (i.e., trials in which the child did not produce a response) were treated as errors of comprehension.[3]

3.1 | Attempted trials

The number of trials in which a tap response was produced, regardless of whether the response was correct (i.e., tap on target) or incorrect (i.e., tap on distractor), was used as a measure of toddlers' motivation to produce a response during the word recognition task. Results from a Welch's t-test indicated that toddlers who were tested online ($M = 44.286$, $SD = 6.359$) and those who were tested in the laboratory ($M = 40.810$, $SD = 7.061$) did not differ significantly in the number of attempted trials ; $t(40.601) = -1.779$, $p = .083$ (see Figure 2).

[3]The analysis of the number of attempted trials (see below) revealed that toddlers produced less tap responses for difficult trials than for easy trials, suggesting that un-answered trials are not random, but predominantly reveal errors of comprehension—in line with similar observations reported in Friend and Keplinger (2003) and Friend et al. (2012).

FIGURE 1 Screenshot of a trial in the test phase

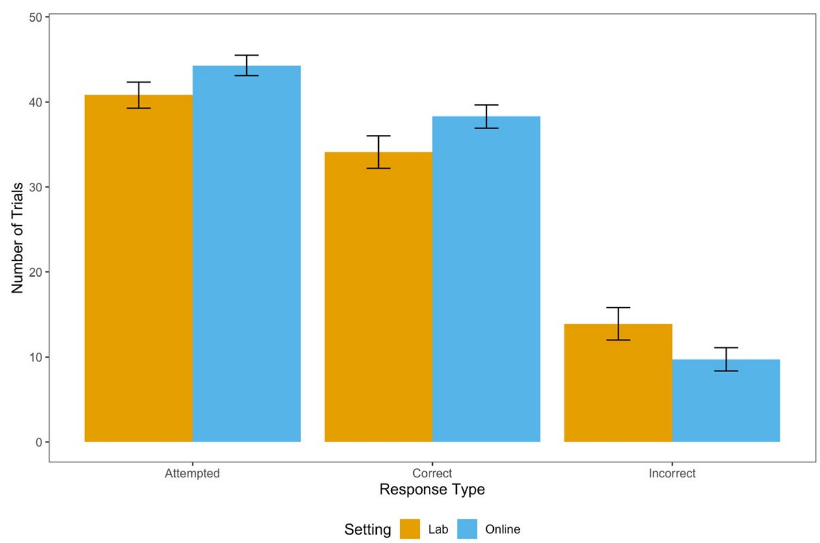


FIGURE 2 Attempted, correct, and incorrect trials across different settings

To assess whether toddlers' motivation (as indexed by whether an attempt to provide a tap response was made) differed across semantic relatedness and difficulty of the trials, a binomial generalised linear mixed-effects model (GLMM) with a logit link function was fitted and analyzed using the mixed() function from the *afex* package (Singmann et al., 2020), which relies on the *lme4* package (Bates et al., 2015) for model fitting. The model included semantic relatedness (related, unrelated), difficulty (easy, moderately difficult, and difficult), toddlers' age (in months), and the interaction between semantic relatedness and difficulty as fixed effects, as well as participant and selected object as random intercepts.[4] Both semantic relatedness (−1: unrelated; 1: related) and difficulty (−1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age was centered on the mean. To determine a model with a parsimonious random effect structure (Matuschek et al., 2017), the forward "best-path" approach, with $\alpha = .20$ as the inclusion criterion, was used to test random slopes for inclusion (Barr

[4]The inclusion of setting (i.e., online vs. lab) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

TABLE 3 GLMM results for attempted trials

	Model summary			Model comparison		
	β	*SE*	*z*	χ^2	*df*	*p*
Intercept	3.080	0.281	10.956	103.539	1	<.001***
Relatedness	−0.087	0.075	−1.163	1.355	1	.244
Difficulty				8.516	2	.014*
Moderate	−0.057	0.105	−0.542			
Difficult	−0.249	0.103	−2.432			
Age	0.949	0.395	2.402	5.686	1	.017*
Relatedness:Difficulty				1.618	2	.445
Relatedness:Moderate	−0.106	0.105	−1.006			
Relatedness:Difficult	0.116	0.102	1.136			

**p* < .05.; ***p* < .01.; ****p* < .001.

et al., 2013). As none of the random slopes fell below the inclusion criterion, the random-intercepts-only model was retained:

$$\text{Attempted} \sim \text{Relatedness}*\text{Difficulty} + \text{Age} + (1\,|\,\text{Participant}) + (1\,|\,\text{Object})$$

The results are detailed in Table 3, with chi-square statistics and *p*-values obtained using likelihood ratio tests. Follow-up pairwise comparisons, with *p*-values adjusted using the Tukey method, were conducted using the pairs() function in the *emmeans* package (Lenth, 2020).

As shown in Table 3, there were significant main effects of trial difficulty and age, with the number of attempted trials increasing with age. No significant main effect of semantic relatedness was found; neither did semantic relatedness interact with difficulty. Results from the follow-up tests indicated that toddlers attempted significantly more easy than difficult trials ($\beta = 0.556$, $SE = 0.186$, $z = 2.995$, $p = .008$), while no such difference was found between easy and moderately difficult trials ($\beta = 0.363$, $SE = 0.189$, $z = 1.917$, $p = .134$) as well as moderately difficult and difficult trials ($\beta = 0.193$, $SE = 0.176$, $z = 1.096$, $p = .517$; see also Figure 3).

3.2 | Correct trials

Results from a Welch's *t*-test indicated that there was no statistically significant difference between toddlers who were tested online ($M = 38.286$, $SD = 7.262$) and those who were tested in the laboratory ($M = 34.095$, $SD = 8.717$) in terms of the number of trials in which they correctly identified the target referent; $t(38.508) = -1.787$, $p = .082$ (see Figure 2). To assess whether toddlers' accuracy differed across semantic relatedness and difficulty of the trials, a binomial GLMM with a logit link function was again fitted and analyzed. The model included the same fixed effects as the previous model (i.e., semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty) and the same random intercepts (i.e., participant and selected object), with by-participant adjustments to the slope of difficulty:[5]

[5]The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

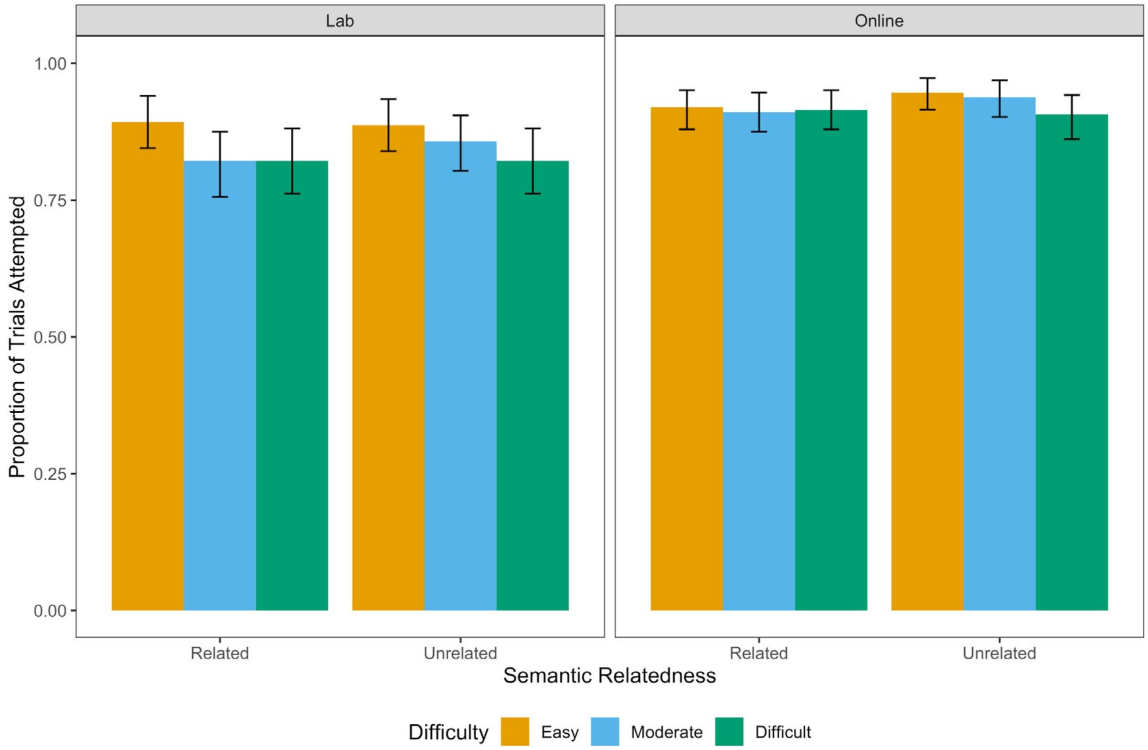


FIGURE 3 Proportion of attempted trials across settings by semantic relatedness and difficulty

$$\text{Accuracy} \sim \text{Relatedness} * \text{Difficulty} + \text{Age} + (1 + \text{Difficulty}|\text{Participant}) + (1\,|\,\text{Object})$$

The results are detailed in Table 4, with chi-square statistics and *p*-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with *p*-values adjusted using the Tukey method.

As shown in Table 4, there were significant main effects of semantic relatedness, difficulty, and age. Specifically, toddlers responded with higher accuracy in semantically unrelated than related trials. Toddlers' accuracy also increased significantly with age. No interaction was found between semantic relatedness and difficulty, however. Results from the follow-up tests indicated that toddlers were significantly more accurate in easy trials relative to both moderately difficult ($\beta = 0.523$, $SE = 0.183$, $z = 2.861$, $p = .012$) and difficult trials ($\beta = 1.113$, $SE = 0.164$, $z = 6.799$, $p < .001$). Toddlers were also significantly more accurate in moderately difficult than difficult trials ($\beta = 0.590$, $SE = 0.150$, $z = 3.924$, $p < .001$; see also Figure 4).[6]

3.3 | Convergent validity

At the general level, toddlers' receptive vocabulary size, as measured by the CDI–WG, and their overall accuracy in the word recognition task significantly correlated in both unrelated, $r_{(47)} = .631$, $p < .001$ and related trials, $r_{(47)} = .603$, $p < .001$. Partialling out the effect of age further revealed that

[6]A Spearman correlation between toddlers' overall word recognition accuracy and SES revealed no relationship, rho = 0.1, $p = .46$.

TABLE 4 GLMM results for accuracy

	Model summary			Model comparison		
	β	*SE*	*z*	χ^2	*df*	*p*
Intercept	1.438	0.143	10.038	56.979	1	<.001***
Relatedness	−0.141	0.054	−2.624	6.782	1	.009**
Difficulty				36.405	2	<.001***
Moderate	0.022	0.097	0.229			
Difficult	−0.568	0.085	−6.660			
Age	0.537	0.193	2.779	7.233	1	.007**
Relatedness:Difficulty				3.887	2	.143
Relatedness:Moderate	−0.114	0.076	−1.511			
Relatedness:Difficult	0.127	0.071	1.785			

*p < .05.; **p < .01.; ***p < .001.

toddlers' receptive vocabulary size accounted for a significant proportion of unique variance in their recognition accuracy, beyond that accounted for by their age in both unrelated, $r_{(46)} = .593$, $p < .001$, $R^2 = .352$ and related trials, $r_{(46)} = .538$, $p < .001$, $R^2 = .289$.

To explore the consistency between toddlers responses and parent-reported comprehension on the test items (i.e., parent–child agreement), item-level agreement was calculated (see Table 5) and a binomial GLMM with a logit link function was fitted. The model included semantic relatedness, difficulty, age, and the interaction between semantic relatedness and difficulty as fixed effects. Both semantic relatedness (−1: unrelated; 1: related) and difficulty (−1: easy; 1: moderately difficult, difficult) were sum-coded, whereas age was centered on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slopes of semantic relatedness, difficulty, and their interaction:[7]

$$\text{Accuracy} \sim \text{Relatedness*Difficulty} + \text{Age} + (1 + \text{Relatedness*Difficulty}|\text{Participant}) + (1|\text{Object})$$

The GLMM results are detailed in Table 6, with chi-square statistics and *p*-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with *p*-values adjusted using the Tukey method.

Overall, as shown in Table 6, there was good item-level agreement between parental reports and toddlers' responses, although this attenuated with increasing item difficulty. Results from the GLMM indicated that semantic relatedness, difficulty, as well as the interaction between semantic relatedness and difficulty (but not age) significantly predicted parent–child agreement (see also Figure 5). The follow-up tests revealed that parent–child agreement was significantly higher in semantically unrelated than related easy trials ($\beta = 0.795$, $SE = 0.299$, $z = 2.662$, $p = .008$), but no significant differences were found across the different semantic conditions in the moderately difficult ($\beta = 0.253$, $SE = 0.169$, $z = 1.495$, $p = .135$) and difficult trials ($\beta = -0.166$, $SE = 0.164$, $z = -1.014$, $p = .311$).

To further examine whether item-pair comprehension status (i.e., whether the target or the distractor label was known or not known by the toddler as indicated by parental responses on the

[7]The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

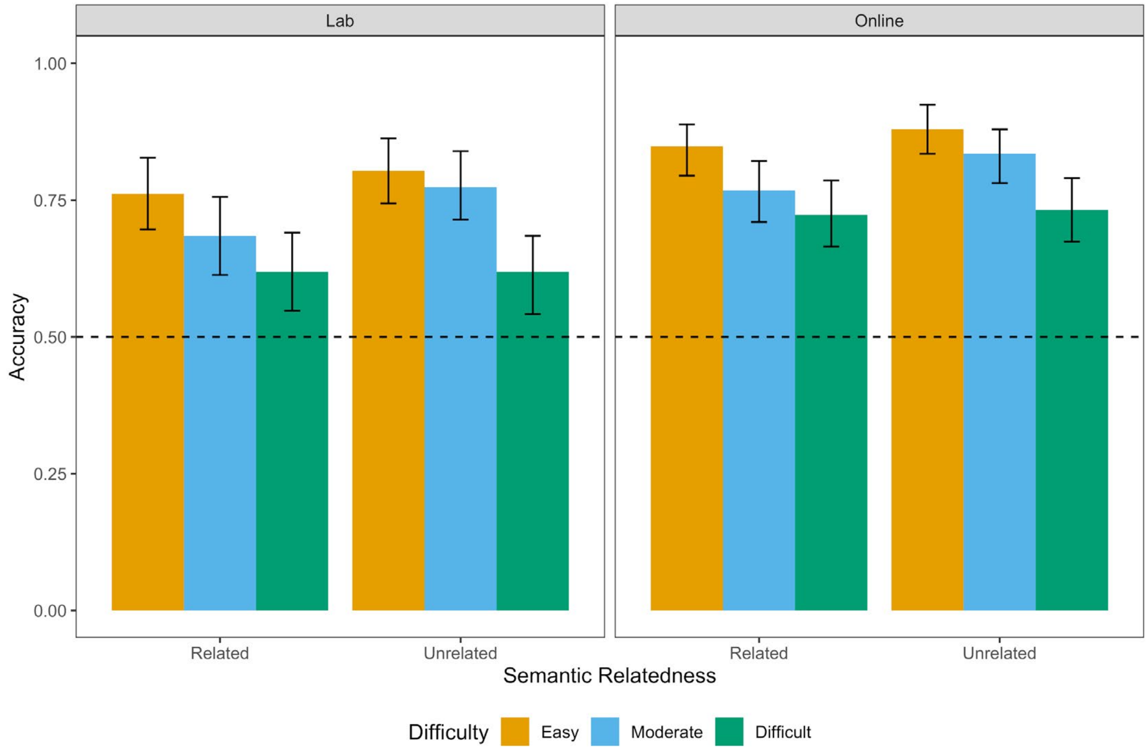


FIGURE 4 Accuracy by semantic relatedness and difficulty across different settings. *Note*. Dashed line represents chance (.50)

TABLE 5 Item-level agreement between parental report and toddler performance

Difficulty level	Semantically related	Semantically unrelated	Overall
Easy	.781	.827	.804
Moderate	.615	.661	.638
Difficult	.564	.538	.551
Overall	.653	.675	.664

CDI–WG) was an accurate predictor of toddlers' performance in the word recognition task, another binomial GLMM with a logit link function was fitted, with semantic relatedness, difficulty, item-pair comprehension status, age, and the interaction between semantic relatedness and difficulty as fixed effects. Semantic relatedness (−1: unrelated; 1: related), difficulty (−1: easy; 1: moderately difficult, difficult), and item-pair comprehension status (−1: both unknown; 1: both known, target known only, distractor known only) were sum-coded, whereas age was centered on the mean. Random intercepts included participant and selected object, with by-participant adjustments to the slope of difficulty:[8]

$$\text{Accuracy} \sim \text{Relatedness*Difficulty} + \text{Pair Comprehension} + \text{Age} + (1 + \text{Difficulty}|\text{Participant}) + (1|\text{Object})$$

[8]The inclusion of setting (i.e., online vs. laboratory) and sex as fixed effects in the model did not change the conclusions and were thus omitted.

TABLE 6 GLMM results for parent–child agreement

	Model summary			Model comparison		
	B	*SE*	*z*	χ^2	*df*	*p*
Intercept	0.921	0.163	5.663	68.207	1	<.001***
Relatedness	−0.147	0.066	−2.237	5.436	1	.020*
Difficulty				21.564	2	<.001***
Moderate	−0.240	0.168	−1.423			
Difficult	−0.752	0.182	−4.134			
Age	0.074	0.153	0.486	0.218	1	.641
Relatedness:Difficulty				9.994	2	.007**
Relatedness:Moderate	0.020	0.082	0.249			
Relatedness:Difficult	0.230	0.076	3.030			

**p* < .05.; ***p* < .01.; ****p* < .001.

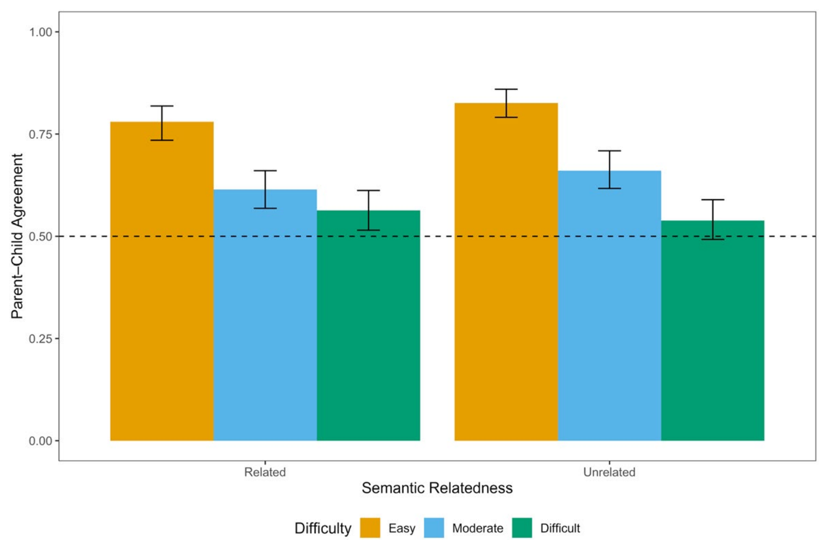


FIGURE 5 Parent–child agreement by semantic relatedness and difficulty. *Note*. Dashed line represents chance (.50)

The results are detailed in Table 7, with chi-square statistics and *p*-values obtained using likelihood ratio tests. Follow-up pairwise comparisons were conducted with *p*-values adjusted using the Tukey method.

As shown in Table 7, parent-reported item-pair comprehension was a significant predictor of toddlers' performance, along with semantic relatedness, difficulty, and age. No significant interaction effect between semantic relatedness and difficulty was found. Results from the follow-up tests indicated that toddlers were significantly less accurate when both target and distractor were reported as unknown compared to when both were known ($\beta = -0.628$, $SE = 0.190$, $z = -3.300$, $p = .005$) and when only the target was known ($\beta = -0.769$, $SE = 0.196$, $z = -3.923$, $p < .001$). No significant differences were found in other cases: (a) both known and target known only ($\beta = -0.141$, $SE = 0.195$, $z = -0.725$, $p = .887$); (b) both known and distractor known only ($\beta = -0.284$, $SE = 0.184$, $z = 1.539$, $p = .414$); (c) target known only and distractor known only ($\beta = 0.425$, $SE = 0.205$, $z = 2.070$, $p = .163$); (d) distractor known only and both unknown ($\beta = -0.344$, $SE = 0.186$, $z = 1.846$, $p = .252$; see also Figure 6).

TABLE 7 GLMM results for accuracy (with parent-reported item-pair comprehension as predictor)

	Model summary			Model comparison		
	β	*SE*	*z*	χ^2	*df*	*p*
Intercept	1.402	0.144	9.749	58.245	1	<.001***
Relatedness	−0.139	0.054	−2.588	6.586	1	.010*
Difficulty				14.702	2	<.001***
Moderate	0.007	0.098	0.068			
Difficult	−0.403	0.107	−3.776			
Pair comprehension				18.108	1	<.001***
Both known	0.193	0.114	1.685			
Target known	0.334	0.125	2.667			
Distractor known	−0.091	0.117	−0.778			
Age	0.511	0.181	2.817	7.428	1	.006**
Relatedness:Difficulty				4.141	2	.126
Relatedness:Moderate	−0.120	0.076	−1.581			
Relatedness:Difficult	0.132	0.072	1.832			

*$p < .05$.; **$p < .01$.; ***$p < .001$.

4 | DISCUSSION

In the interest of developing a performance-based measure of comprehension during the second year of life that addresses the need for a convergent and supplemental measure of parental reports, while taking into account young children's non-compliance and limited attention capabilities (as in Friend & Keplinger, 2003), the present study explored the viability of using a tablet-based 2AFC word recognition task in assessing early word comprehension.

Toddlers aged between 18 and 20 months were tested—either in the laboratory setting by an experimenter or online (i.e., at home) by their parents—on their comprehension of 24 lexical items selected from the Norwegian CDI–WG (Simonsen et al., 2014). During the task, toddlers were asked to identify the referent for the lexical target presented alongside a distractor. Target–distractor pairs were manipulated such that each lexical target was paired once with a semantically related distractor and once with a semantically unrelated distractor. Item pairs also varied in three levels of difficulty (defined based on the Norwegian CDI–WG normative data for age-matched children).

Both the analyses on the number of attempted trials (regardless of whether the response was correct or incorrect) as well as the number of trials in which toddlers provided a correct response revealed no significant differences between the online and laboratory samples, suggesting that toddlers were equally motivated to produce a response in the task and that neither setting led to better or poorer performance. This demonstrates that remote infant data collection with fully automatized tasks can be as efficient and reliable as in situ laboratory assessments. High-quality data through remote administration are not only an important enabler during this time of the global COVID-19 pandemic, but also provide a promising avenue for data collection associated with developmental research, with increased speed, lowered cost, and the potential to an improved sample diversity by reaching to a wider sociodemographic background than traditional laboratory-based research (Sheskin et al., 2020).

Overall, in line with Friend and Keplinger (2008), toddlers attempted significantly more easy than difficult trials. Older toddlers also attempted significantly more trials than younger toddlers. Together,

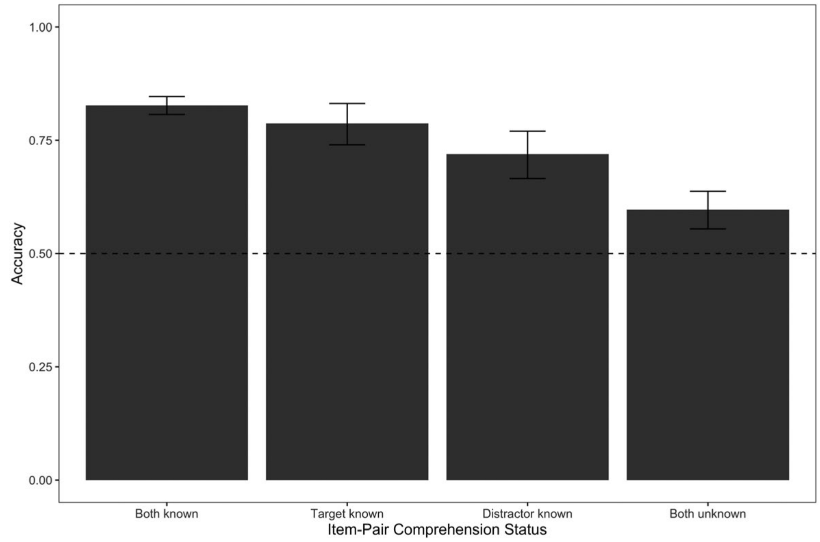


FIGURE 6 Accuracy by parent-reported item-pair comprehension status. *Note*. Dashed line represents chance (.50)

these findings suggest that toddlers were responding non-randomly and bolster the support for the notion that non-responses represent toddlers' true inability to map the lexical target to its referent, rather than their non-compliance or the lack of motivation, while incorrect responses might be taken as evidence of partial word knowledge, and correct responses—robust word knowledge (Hendrickson et al., 2015).

With regard to the accuracy measure, toddlers demonstrated above-chance performance throughout the task. Congruent with previous work (Friend & Keplinger, 2003, 2008), toddlers' performance was consistent with the a priori "cohort-level" difficulty categorization, as their best performance was obtained for easy trials and their worst performance for difficult trials. As would be expected, older toddlers also performed with greater accuracy relative to younger toddlers.

Examining the role of semantic relatedness, it was found that toddlers displayed more robust recognition in semantically unrelated than related trials, suggesting that, and similar to research in younger infants (Bergelson & Aslin, 2017a), semantical relatedness between the target and the distractor triggered competition effects in referent selection. Although there is evidence that early word representations are semantically more specified by 18–20-months of age (Bergelson & Aslin, 2017b), they still might be lacking representational specificity (Arias-Trejo & Plunkett, 2010). In the current study, in addition to semantic relatedness, lower recognition on some related trials could be attributed to the increased burden of visual discrimination and feature overlap (e.g., both goose and owl are birds and have wings, feather, and a beak), as shown with 18–24-month-olds in Arias-Trejo and Plunkett (2010). It is likely that toddlers, upon hearing the lexical target, co-activated related (and thus, competing) word referents, which subsequently interfered with their lexical decision about the target. Such interference has been reported even among older children, between 3 and 9 years of age, as they took longer to provide a correct response in a visual search task when a related distractor was present than when an unrelated distractor was present (Vales & Fisher, 2019).

Comparing between toddlers' recognition accuracy and their receptive vocabulary size as measured by the CDI–WG, a significant and moderate correlation (comparable to that achieved with the CCT; Friend & Keplinger, 2008) was found, evincing acceptable convergent validity of the word recognition task employed in the present study, and also supporting the feasibility of the CDI–WG, as a general proxy of receptive vocabulary.

Consistent with the CCT (Friend et al., 2012; Friend & Zesiger, 2011), there was also good (albeit not perfect) item-level agreement between toddlers' responses and parental reports across both semantic conditions, with easy items having the highest agreement and difficult items having the lowest agreement. The results further indicated that parent–child agreement was significantly higher in semantically unrelated than related trials, although this was only limited to easy items. This discrepancy suggests that parents' inference on their child's word comprehension is not solely based on evidence of their child's true ability to comprehend the word, but rather on the confluence of both evidence of robust word knowledge (i.e., their child's true ability to comprehend the word) and evidence of partial word knowledge (i.e., their child's ability to respond appropriately when cued by the rich context in which the word is heard, or upon recognizing the sound of the word; Friend et al., 2018; Houston-Price et al., 2007; Tomasello & Mervis, 1994). Restating the finding that toddlers were less accurate in semantically related than unrelated trials, a performance-based measure that uses semantically related target–distractor pairs can potentially tap children's strong, rather than weak, word knowledge to supplement parental reports. Nevertheless, parent-reported item-pair comprehension (i.e., whether the target or distractor label was known or not known by the child) was found to be a significant predictor of toddlers recognition accuracy. Specifically, compared to trials where both the target and distractor were reported by parents as "not understood" on the CDI–WG, toddlers were more likely to respond correctly in trials where either the target or both the target and distractor were reported as "understood," indicating that parents are adequate informants of their child's language abilities.

It is important to note that the while the CCT uses a set of carefully selected test items consisting of an equal representation of nouns, verbs, and adjectives, the present study focused on nouns only. Nevertheless, the good item-level agreement between parental reports and their child's performance provides encouraging results. Supplemented with a principled selection of test items (Chai et al., 2020; Makransky et al., 2016) and with statistical methods to allow for an estimation of full CDI scores (Mayor & Mani, 2019) and total vocabulary sizes (Mayor & Plunkett, 2011), tablet-based word recognition tasks may provide a useful measure of receptive vocabulary skills in the second year of life—and potentially serve as a supplemental and convergent measure of parental reports.

ACKNOWLEDGEMENTS

We are very thankful to Ane Theimann for her help with participant recruitment and data collection. We are thankful to all toddlers and their parents who took part in the study.

ORCID

Chang Huan Lo https://orcid.org/0000-0002-8262-7258
Audun Rosslund https://orcid.org/0000-0002-2646-8053
Jun Ho Chai https://orcid.org/0000-0003-4316-9407
Julien Mayor https://orcid.org/0000-0001-9827-5421
Natalia Kartushina https://orcid.org/0000-0003-4650-5832

REFERENCES

Abdul Aziz, N. A., Mat Sin, N. S., Batmaz, F., Stone, R., & Chung, P. W. H. (2014). Selection of touch gestures for children's applications: Repeated experiment to increase reliability. *International Journal of Advanced Computer Science and Applications*, *5*(4), 97–102. https://doi.org/10.14569/IJACSA.2014.050415

Ackermann, L., Lo, C. H., Mani, N., & Mayor, J. (2020). Word learning from a tablet app: Toddlers perform better in a passive context. *PLoS One*, *15*(12), e0240519. https://doi.org/10.1371/journal.pone.0240519

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1–2), 63–80. https://doi.org/10.1016/j.jecp.2009.10.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bergelson, E., & Aslin, R. N. (2017a). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916. https://doi.org/10.1073/pnas.1712966114

Bergelson, E., & Aslin, R. N. (2017b). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, *13*(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476. https://doi.org/10.1017/S0142716416000060

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*(3), 619–650. https://doi.org/10.1017/S0305000908008714

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer* [Computer software]. http://www.praat.org/

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Chai, J. H., Lo, C. H., & Mayor, J. (2020). A Bayesian-inspired item response theory-based framework to produce very short versions of MacArthur–Bates communicative development inventories. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3488–3500. https://doi.org/10.1044/2020_JSLHR-20-00361

Clark, R. (1974). Performing without competence. *Journal of Child Language*, *1*(1), 1–10. https://doi.org/10.1017/S0305000900000040

Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*(3), 544–560. https://doi.org/10.1044/1092-4388(2003/044)

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes*? Journal of Child Psychology and Psychiatry*, *56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Dunn, D. M. (2018). *Peabody picture vocabulary test-fifth edition*. Pearson Assessment.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, *71*(2), 310–322. https://doi.org/10.1111/1467-8624.00146

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. Singular.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*(1), 95–116. https://doi.org/10.1017/S0142716400001053

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Frank, M. C., Braginsky, M., Marchman, V. A., & Yurovsky, D.(2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17. https://doi.org/10.1080/15248372.2015.1061528

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 302–309. https://doi.org/10.3758/bf03202556

Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): Data from American English and Mexican Spanish infants. *Journal of Child Language*, *35*(1), 77–98. https://doi.org/10.1017/s0305000907008264

Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the computerized comprehension task: Comprehension predicts production. *Developmental Psychology*, *48*(1), 136–148. https://doi.org/10.1037/a0025511

Friend, M., Smolak, E., Liu, Y., Poulin-Dubois, D., & Zesiger, P. (2018). A cross-language study of decontextualized vocabulary comprehension in toddlerhood and kindergarten readiness. *Developmental Psychology*, *54*(7), 1317. https://doi.org/10.1037/dev0000514

Friend, M., & Zesiger, P. (2011). Une réplication systématique des propriétés psychométriques du Computerized Comprehension Task dans trois langues [A systematic replication of the psychometric properties of the CCT in three languages]. *Enfance*, *63*(3), 329–344. https://doi.org/10.4074/S0013754511003053

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. https://doi.org/10.1017/s030500090001271x

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science*, *18*(5), 723–735. https://doi.org/10.1111/desc.12250

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Language, speech, and communication. Methods for assessing children's syntax* (pp. 105–124). MIT Press.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language*, *34*(4), 701–724. https://doi.org/10.1017/s0305000907008124

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Kaler, S. R., & Kopp, C. B. (1990). Compliance and comprehension in very young toddlers. *Child Development*, *61*(6), 1997–2003. https://doi.org/10.2307/1130853

Kartushina, N., & Mayor, J. (2019). Word knowledge in six- to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science*, *6*(9), 180711. https://doi.org/10.1098/rsos.180711

Klesty, V., & Fouche, G. (2020, March 24). Norway extends coronavirus curbs until April 13. *Reuters*, https://www.reuters.com/article/us-health-coronavirus-norway-restriction/norway-extends-coronavirus-curbs-until-april-13-idUSKBN21B2ED

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means (Version 1.4.5)* [R package]. https://CRAN.R-project.org/package=emmeans

Lo, C. H., Mani, N., Kartushina, N., Mayor, J., & Hermes, J. (2021). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *PsyArXiv*. Under review. https://doi.org/10.31234/osf.io/u73sy

Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory–based, computerized adaptive testing version of the MacArthur–Bates Communicative Development Inventory: Words & Sentences (CDI:WS). *Journal of Speech, Language, and Hearing Research*, *59*(2), 281–289. https://doi.org/10.1044/2015.JSLHR-L-15-0202

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J., Lahmar, J., Scott, F., Davenport, A., Davis, S., French, K., Piras, M., Robinson, P., Thornhill, S., & Winter, P. (2015). *Exploring play and creativity in pre-schooler's use of apps: Final project report*. http://www.techandplay.org/reports/TAPFinalReport.pdf

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

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769–785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, *86*(5), 1351–1370. https://doi.org/10.1111/cdev.12398

Neumann, M. M., Worrall, S., & Neumann, D. L. (2019). Validation of an expressive and receptive tablet assessment of early literacy. *Journal of Research on Technology in Education*, *51*(4), 326–341. https://doi.org/10.1080/15391523.2019.1637800

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, *17*(1), 57–70. https://doi.org/10.1177/1367006911431198

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L. I., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences*, *24*(9), 675–678. https://doi.org/10.1016/j.tics.2020.06.004

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*(1), 3–23. https://doi.org/10.1177/0142723713510997

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). *afex: Analysis of factorial experiments (Version 0.27-2)* [R package]. https://CRAN.R-project.org/package=afex

Styles, S., & Plunkett, K. (2009). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*, *36*(4), 895–908. https://doi.org/10.1017/S0305000908009264

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202–217. https://doi.org/10.1111/bjdp.12157

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development*, *59*(5), 174–179. https://doi.org/10.1111/j.1540-5834.1994.tb00186.x

Twomey, D. M., Wrigley, C., Ahearne, C., Murphy, R., De Haan, M., Marlow, N., & Murray, D. M. (2018). Feasibility of using touch screen technology for early cognitive assessment in children. *Archives of Disease in Childhood*, *103*(9), 853–858. https://doi.org/10.1136/archdischild-2017-314010

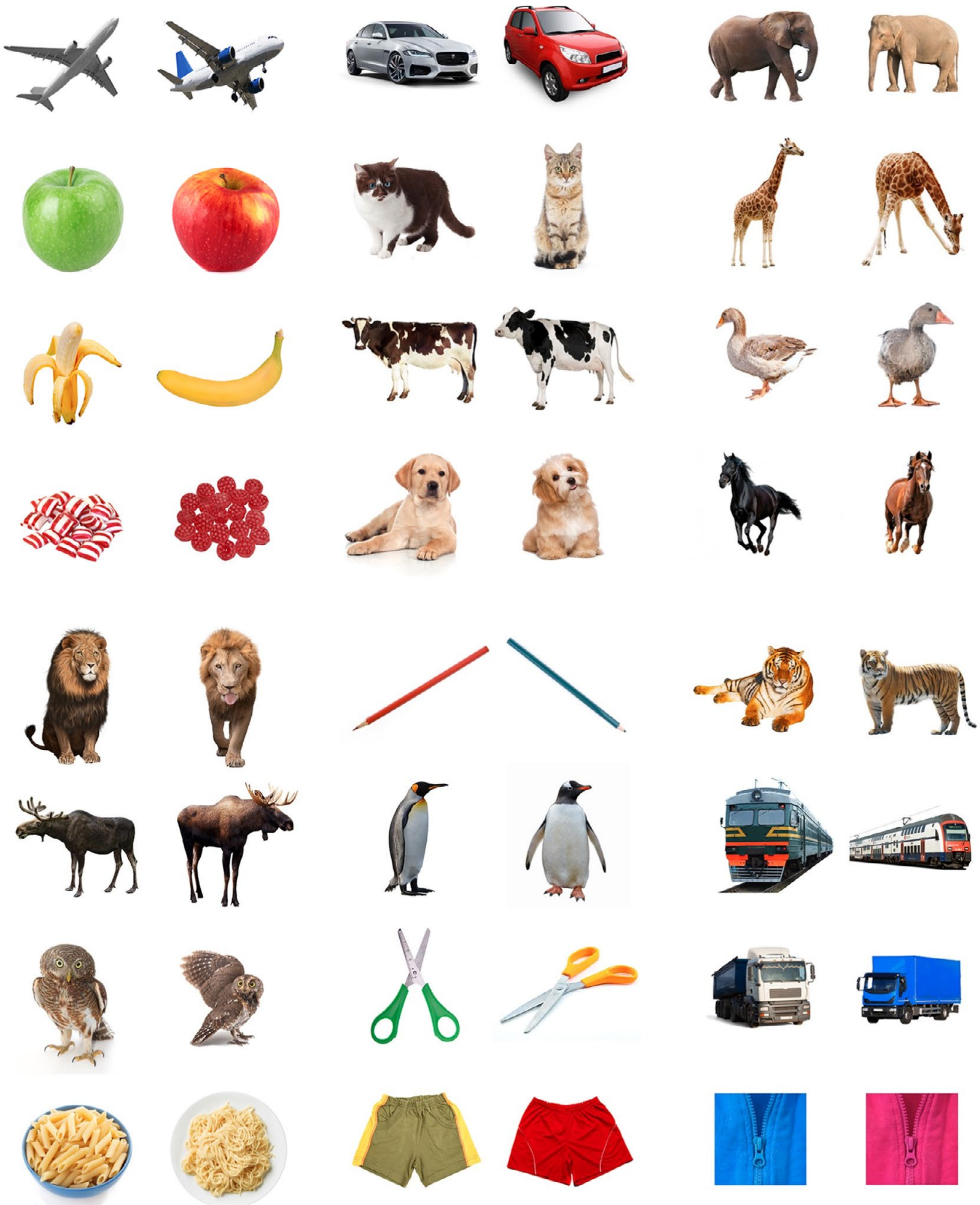
Vales, C., & Fisher, A. V. (2019). When stronger knowledge slows you down: Semantic relatedness predicts children's co-activation of related items in a visual search paradigm. *Cognitive Science*, *43*(6), e12746. https://doi.org/10.1111/cogs.12746

Williams, K. T. (2018). *Expressive vocabulary test-third edition*. Pearson Assessment.

Yoder, P. J., Warren, S. F., & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology*, *6*(1), 59–64. https://doi.org/10.1044/1058-0360.0601.59

How to cite this article: Lo CH, Rosslund A, Chai JH, Mayor J, Kartushina N. Tablet assessment of word comprehension reveals coarse word representations in 18–20-month-old toddlers. *Infancy*. 2021;00:1–21. https://doi.org/10.1111/infa.12401

APPENDIX 1



APPENDIX 2