

Reliability and Associations of English and Mandarin Vocabulary Measures in Infants and Toddlers Growing Up in a Multilingual Society

Journal of Psychoeducational Assessment

2026, Vol. 0(0) 1–21

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/07342829261440789

journals.sagepub.com/home/jpa

Shaun Kok Yew Goh¹ , Vernessa Liew¹, Suzy Styles^{2,3}, He Sun¹, Lit Wee Sim⁴, Jambay Dorji⁵, Galih Kunarso⁵, Ngiap Chuan Tan^{5,6}, and Anne Rifkin-Graboi^{1,4}

Abstract

This paper aims to understand the reliability of English and Mandarin vocabulary measures, and their associations, among infants and toddlers growing up in a multilingual society. Reliability was examined via item-response Rasch models across English and Mandarin word categories. Parents reported on 137 children's (age = 5 to 27 months) English and Mandarin vocabulary with a newly developed measure, the developmental vocabulary checklist (DVC). High reliability (alpha .91 to .99) was observed across English and Mandarin DVC categories. Wright maps suggest a lack of sensitivity at the tail end of the language-ability distribution. One-way ANOVAs indicating a significant difference across expected categories of DVC words was found. Higher exposure associated with a higher composite score across DVC categories, but only in Mandarin and only among older children in the 1-to-2 year band. A threshold account is discussed as an explanation for differential association of exposure in Mandarin but not English.

Keywords

assessing vocabulary, reliability, age/maturation effects, bilingualism, psychometric assessments

¹National Institute of Education, Nanyang Technological University, Singapore

²Psychology, School of Social Sciences, Nanyang Technological University, Singapore

³Centre for Research and Development on Learning, Nanyang Technological University, Singapore

⁴Yong Yoo Lin School of Medical, National University of Singapore, Singapore

⁵Singhealth Polyclinics, Singapore

⁶Singhealth Duke-NUS, Singapore

Corresponding Authors:

Shaun Kok Yew Goh, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616.

Email: shaun.goh@nie.edu.sg

Anne Rifkin-Graboi, Yong Yoo Lin School of Medical, National University of Singapore, 1 Nanyang Walk, Singapore 637616.

Email: rifkina@nus.edu.sg

Introduction

Although much is known about early language development, relatively little is known within multilingual societies, and even less about how to measure this in a reliable and valid manner. Such information can be of value, bringing a developmental perspective to policies in multilingual societies which foster bilingualism and highlight the contribution of languages for cultural and economic connectivity. This manuscript aims to do so, by first introducing the multilingual context of Singapore and providing an overview of a growing body of research measuring language among bilingual infants and toddlers. We then leverage language data from an ongoing study (Rifkin-Graboi et al., 2019) in multilingual Singapore to quantitatively examine patterns of English and Mandarin vocabulary development. Here, we use a novel measure designed for measuring early language in multilingual Singapore, the Developmental Vocabulary Checklist (DVC).

Singapore's Multilingual Context

Singapore's multilingual landscape includes four official languages: Bahasa Melayu ('Malay'), Mandarin Chinese, Tamil and English. Moreover, the Singaporean government established the bilingual policy (Ministry of Education; MOE, 2010), requiring every student to study English as the main language of instruction in all content areas, and to also study a 'Mother Tongue' typically consisting of either Mandarin, Malay or Tamil. Apart from the official languages, a plethora of other languages enrich the nation's linguistic tapestry (Dixon, 2005). For instance, Malay ethnicities may also utilise Javanese, Bahasa Indonesian and Boyanese. Chinese dialects include Hokkien, Teochew, Cantonese and Hakka. Lastly, apart from the Dravidian language Tamil, which is spoken by most people of South Asian ethnicity, there is also Punjabi, Bengali, Telugu, Hindi, Malayalam and Gujarati. Furthermore, there are a number of guest workers, many of whom work as domestic workers caring for children. Among these domestic workers Tagalog and other Filipino languages, are common alongside other languages of the region. In general, the language context of Singapore is generally characterised as described as 'English knowing bilingualism' (Pakir, 1993).

Given this diversity in language exposure, it follows that infants and toddlers exhibit varying levels of proficiency in each language, Infants and toddlers must allocate their language-learning time between two languages, and some words are likely encountered exclusively in contexts where only one language is used (Bialystok et al., 2010). Moreover, infants and toddlers in Singapore are typically exposed to two or more languages (Goh et al., 2017; Woon, 2018). With such variability in exposure, usage and proficiency, characterising the typical range of language development among infants and toddlers growing up in this multilingual setting is a complex endeavour. Prior work in Singapore has attempted to quantify language with some success. However, as will be discussed below, these have predominantly focused on English and have limited information on reliability and validity of measurement.

Measuring Language Among Infants and Toddlers in Singapore: Reliability, Item Difficulty and Validity

Reliability. One body of work in Singapore has focused on measuring language to arrive at an indication of infants and toddlers who may require additional language support. Goh et al. (2017) examined language items from the Bayleys Scales of Infant and Toddler Development-III on 24-month old toddlers. Items were administered by an examiner in the child's dominant language, and in the child's non-dominant language if it was judged that the child had not understood. Statistical

analyses suggest that two-word utterances were harder for bilingual than monolinguals, while being asked to point to a 'shirt' was harder for English than non-English exposed children. Reliability was acceptable, but only when measuring language at the lower tail end of language proficiency. [Agarwal et al. \(2020\)](#) examined the Ages and Stages Questionnaire. This parent report measure in English, was translated into Mandarin, Malay and Tamil. Reliability for the communication subscale was less than adequate for 18-months toddlers, and adequate for 9 and 24 months toddlers.

These mixed findings of reliability therefore signal the need for a better understanding of how local language measures could be made more reliable. It may be that aforementioned measures of language which report good psychometric properties in their validation samples ([Bayley, 2006](#); [Squires et al., 2009](#)) were not designed for multilingual contexts, resulting in mixed findings of reliability when examined locally. For instance, one common feature is that when aforementioned measures are utilised locally ([Goh et al., 2017](#); [Agarwal et al., 2020](#)), different types of languages are informally aggregated into a single score, with no separate scores by types of language (i.e. English or a 'Mother Tongue' such as Mandarin, Malay or Tamil). This may be a source of unreliability, as local and international studies among multilingual populations commonly find different type of languages to be related but separable factors ([He et al., 2020](#); [Pratt et al., 2024](#)). It is therefore important to understand the extent to which scores can be found to be reliable when disaggregated into two languages.

Item Difficulty. Another body of work has focused on typical language development of infants and toddlers, utilising the MacArthur-Bates Communicative Development Inventory (MCDI). The MCDI is a widely used measure of vocabulary size in young children below 30 months of age ([Law & Roy, 2008](#); [Singh et al., 2022](#)). The MCDI is to be completed by caregivers or parents of young children, who must indicate on the form which words their children can already understand and/or produce. [Singh et al. \(2022\)](#) studied a local adaptation of the MCDI, the Singapore Communicative Development Inventories (SCDI), among a large sample of 1316 infants 17 to 21 months old growing up in Singapore. Adaptation involved the review of existing items with native speakers, adding or removing items as necessary, and backwards and forwards translations for Mandarin and Malay vocabulary lists. The SCDI's validity was supported, as vocabulary scores significantly associated with socio-economic status, age and exposure. However, in keeping with this study's aims, only English vocabulary scores were analysed, and reliability was not calculated.

[Woon et al. \(2020\)](#) analysed SCDI English and Mandarin data across 138 2 year old infants. In keeping with the study's aims, validity but not reliability was examined. If a child was reported to have said or understood a word 'cat' in English, it was credited as 1 mark for English vocabulary. If the child was reported to have said or understood the Mandarin equivalent word '貓 *māo*' for 'cat', they would be credited with 1 mark for Mandarin vocabulary, resulting in a total of 2 marks overall. This strategy allowed for vocabulary estimates in each language. This total word score which sums both Mandarin and English was found to still be lower than MCDI American English norms for children this age. The authors note one reason for this is that additional local words that are directed towards children living in Singapore were not a part of the SCDI.

These findings from the SCDI therefore signal that a better understanding of the difficulty of items in measuring vocabulary in Singapore is needed. For instance, one explanation of aforementioned findings is that existing English or Mandarin items from the SCDI may be too difficult perhaps due to a lack of local words resulting in scores that are lower than expected when compared to the MCDI ([Woon et al., 2020](#)). This is in keeping with expectations that item difficulties of vocabulary will shift when transported across countries ([Dale, 2015](#)). However, it is not currently possible to directly verify this speculative explanation, as not study has examined the

difficulty of items in measuring vocabulary among Singapore's multilingual context remains unknown. Without such information, it is unclear if and where gaps may exist in the difficulty level of vocabulary items, leaving it unclear if English or Mandarin vocabulary in Singapore is being accurately measured.

Validity. Evidence suggestive of the validity of language measures among infants and toddlers in Singapore is accumulating. With regard to construct validity, age, socio-economic status and gender correlated with language scores from the Ages and Stages Questionnaire (Agarwal et al., 2020) and SCDI (Singh et al., 2022). This is in keeping with known findings from international samples, that higher language scores associate with higher socio-economic status, age and female gender (McAvinue, 2025). Less consistent is the role of exposure to a language, which was found to correlate with English in the SCDI (Singh et al., 2022) but not with language scores on the Bayleys-III (Goh et al., 2017). Understanding the conditions under which language exposure associates with language scores in Singapore may be especially important to advancing the local evidence base of construct validity. This is because language exposure is an extensively studied predictor of subsequent bilingual language development (Oller et al., 2007), with variation in the timing of exposure, amount and type of exposure, shown to have differential effects (Hoff, 2021). Language exposure may therefore serve as an important node in the nomological network of multilingual development in Singapore. Criterion validity has been reported by one local study (Agarwal et al., 2024), which found the ages and stages questionnaire (Squires et al., 2009) to have a sensitivity of 27.9% to 68.4% and a specificity of 76.5 to 97.4% when compared to the Bayleys-III (Bayley, 2006) child direct measure. Content validity is mentioned (Singh et al., 2022) but not examined. Taken together, evidence about validity of language measures among infants and toddlers is accumulating in Singapore.

Present Study

We improve upon this state-of-the-art by examining the reliability and concurrent associations of not only English, but also Mandarin, which are the top two most used languages at home, by 48.3% and 29.9% of Singaporeans, respectively (Singapore Department of Statistics, 2022). This also adds to the relatively sparse global literature on English and Mandarin vocabulary among infants and toddlers ($k = 6$; Weisleder et al., 2024). English and Mandarin subscales from a newly developed measure of vocabulary are utilised. The Developmental Vocabulary Checklist (DVC) was developed by Principal Investigator Dr Suzy Styles for cataloguing vocabulary across the languages of Singapore, as part of the Talk Together Study (Woon & Styles, 2021). The DVC is designed to study English, Mandarin, Tamil, Bahasa Melayu and words of unclear origin ('Red-dot') used by many different children across a wide age range of early years. For each word on the checklist, parents were asked to check off the boxes as to whether their child was able to 'understand' or 'understand and say' certain words. Vocabulary items were created and refined through an iterative process involving several native speakers with experience working with children to develop a checklist with vocabulary items suitable for young children in Singapore.

This study of the DVC is based on a distributed model of bilingualism (Oller et al., 2007). For instance, a bilingual child might know some concepts in English only, Mandarin only, or both English and Mandarin. This distribution of semantic knowledge across different types of language is in keeping with findings that bilingual children's vocabulary is lower than monolingual children's vocabulary in English (Oller et al., 2007) but is on par once vocabulary from both languages is considered (Hoff, 2021). Or, that bilingual children are by definition exposed to two languages while growing up, which predicts words and concepts distributed across both languages such as English and Mandarin (Pérez-Navarro & Lallier, 2025). In this way, this research does not

only measure English alone, or Mandarin alone, but rather examines both subscales to measures both. In-line with its newly developed status, this is the first time that the English and Mandarin subscales of the DVC's reliability and structure, item difficulty and construct validity will be studied, as follows.

Research Question 1. To what extent do the English and Mandarin DVC exhibit psychometric properties that allow for valid interpretation? Like other vocabulary batteries, DVC interpretation involves counting and summing the total number of words endorsed by parents, which comprises of systematic and random variability. If a large number of DVC words exhibit random variability, the sum of the words may not index children's overall vocabulary development in a reliable fashion. For example, some vocabulary words may be qualitatively different in usage across the population from other words in the same semantic group. For example, 'temple' within the category of 'places' could be better known among some religious families, leading to unexpected performance of this item within the 'places' group of words when administered to a diverse multi-denominational population. This is examined here through infit and outfit values for each DVC word, which corresponds to quantifying which DVC words exhibit unexpectedly high or low amounts of random variation. Specifically, via a form of item response theory, known as a Rasch model (Boone & Staver, 2020). Also, unidimensionality values are estimated as part of Rasch modelling, which inform how effectively DVC words can be summed into a coherent total. Such analyses can reveal words which skew the sum score in ways that would be more consistent with an alternative category of words (e.g. 'Car' and 'boat' are both vehicles and toys, and may cause a unidimensional category of 'toys' to become multidimensional), which indicate these words should be removed. Lastly, Cronbach's alpha will be calculated to determine the level of systematic and reliable variation at the sum of words scale level.

In this way, research question 1 aims to improve upon the local state-of-the-art of reliability. One common feature in aforementioned international measures is that when they are utilised locally (Goh et al., 2017; Agarwal et al., 2020), different types of languages are informally aggregated into a single score, with no separate scores by types of language (e.g. English or a 'Mother Tongue' such as Mandarin, Malay or Tamil). This may be a source of unreliability, as local and international studies among multilingual populations commonly find moderate to strong correlations between languages spoken (He et al., 2020; Pratt et al., 2024). Research question 1 contributes to this because it examines the extent to which scores can be found to be reliable when disaggregated into two languages.

Research Question 2. How do the observed patterns of English and Mandarin vocabulary development from the DVC correspond with the expected patterns of the six categories of words (e.g. 'family members', 'body parts', 'animals and creatures', 'clothing', 'vehicles' and 'food/drinks')? Children's interest in different natural categories is understood to shape their word learning (Ackermann et al., 2020). Children who display a higher level of interest in vehicles, for instance, are more likely to learn novel vehicle-related vocabulary compared to children who are more interested in animals and vice versa. The nouns in the MCDI were chosen to represent a broad catalogue of words found in early infant vocabularies and are organised according to categories found in infant vocabularies (Borovsky et al., 2016), and the DVC follows a similar structure. Categories identified from the MCDI served as an indicator of the word categories that children are learning most commonly at this age (e.g. six categories: 'animals', 'clothing', 'vehicles', 'body parts', 'fruits' and 'drinks'; Borovsky et al., 2016). Additionally, only words from categories of 'family members' and 'body parts' are mentioned in early local milestones around 12 to 18 months, per nationally normed checklists of language in Singapore (Lim et al., 1994). This suggests that children should be more familiar with words in categories such as family

members and body parts than the words in other categories such as ‘animals and creatures’, ‘clothing’, ‘vehicles’ and ‘food and drink’.

This research question will also advance the local state-of-the-art by providing data about whether vocabulary items are too difficult or not (Woon et al., 2020). Utilising the Rasch model, the relative difficulty of each item in measuring a category of vocabulary is estimated. For instance, how difficult are items such as ‘lorry’ or ‘van’ relative to other ‘vehicle’ type vocabulary items, for infants and toddler growing up in Singapore. In keeping with Woon et al., (2020), it may be that most items are found by the Rasch model to have higher item difficulties that capture stronger linguistic proficiency. These item difficulties will be estimated across six categories of English and Mandarin in the DVC, contributing to the current evidence base about item difficulties for measuring infant and toddler vocabulary. Such item level information sets the stage for clarifying how difficult items are, allowing for a more informed approach for future construction of new language items in terms of their difficulty (easier or difficulty), for which type of language (English or Mandarin) and for which of the six categories tested.

Research Question 3. To what extent is the amount of exposure to each language related to children’s English and Mandarin vocabulary development, while accounting for expected vocabulary increases with age? Higher Mandarin exposure and usage among kindergarten children is associated with higher Mandarin vocabulary scores ($r = .33$; Sun et al., 2018). Yet, there is emerging evidence that this relationship between exposure and vocabulary may be asymmetrical and dampened in English. Goh et al. (2017) reported non-significant associations between English exposure and composites of English receptive and expressive language ability among 24-month year old toddlers in Singapore. This dampened relationship in English is in keeping with international work (Thordardottir, 2011) suggestive of a threshold beyond which additional input becomes unnecessary for children’s receptive vocabulary. This may stem from the finite number of different words in the language environment of children (e.g. home and school). Moreover, children exposed to a certain language for a certain number of hours might acquire considerable amount of vocabulary but eventually plateau in their learning, suggesting that additional input may not significantly enhance their vocabulary acquisition (Thordardottir, 2011). This threshold account has also been shown in local work. Sun et al. (2018) posited that threshold effects would be evident in English but not Mandarin, as English usage is of a higher amount and frequency among Singaporean homes than Mandarin (Singapore Department of Statistics, 2022). In support of this, regression models on a sample of young children in Singapore, found language exposure to explain 7% of variability in English vocabulary, whilst explaining 21% of variability in Mandarin vocabulary. Given these findings, there is reason to predict that relationships between home language exposure and vocabulary size in each language may be stronger for Mandarin than for English.

This research question therefore furthers current local research into the construct validity of language measures among toddlers and infants. Socio-economic status, age and gender have been shown in local studies to correlate with language scores among infants and toddlers (Agarwal et al., 2020; Singh et al., 2022). However, the conditions under which language exposure associate with language scores remains unclear, with mixed local findings (Singh et al., 2022; Goh et al., 2017). This is particularly important to clarify, as a body of aforementioned research studies (Hoff, 2021; Pérez-Navarro & Lallier, 2025) would predict that language exposure to play an important role in the nomological network of the construct of bilingual language development. This research question seeks to clarify if a threshold effect is present (Thordardottir, 2011), with a stronger relationship for the less frequently utilised language in this society (Mandarin; Singapore Department of Statistics, 2022).

Methods

Participants

Participants are drawn from a larger ongoing study, the Beginning Early: SingaPore's Ongoing Study starting in Infancy of Twenty-first-century-skills, Individual differences and Variance in the Environment (BE POSITIVE; Rifkin-Graboi et al., 2019). In brief, BE POSITIVE started recruiting parents and babies in 2021 from public outpatient clinics in the community and is ongoing in Singapore. The BE POSITIVE study is cross-sequential in design, comprising of multiple longitudinal and cross-sectional cohorts with participants entering the study at different ages. Inclusion criteria for BE POSITIVE are that the offspring of participants are Singaporean citizens or permanent residents (PRs) or in the process of applying for PR. Children had to be between the ages of 0 and 40 months at the time of recruitment. They must be singletons at birth (i.e. not part of a set of twins or triplets) and have an adult (biological, adoptive or legal parent) participate in BE POSITIVE. Exclusion criteria included children with a known developmental, neurological or psychiatric condition or a history of severe head injuries.

For this current manuscript, data was extracted on March 2023 from children which entered BE POSITIVE at 4 months, 5 to 8 months, 18 months and 22 to 25 months. One child's data was not analysed because her father and mother only spoke to her in Indonesian, which is neither English nor Mandarin, the focus of this study. Although BE POSITIVE administered the DVC to multiple caregivers, only the mothers' reports for English and Mandarin DVC were utilised in this manuscript as there were insufficient reports from other caregivers for statistical analyses. As a result, 137 parental reports on 71 male children and 66 female children were included in the sample size of the data used for statistical analysis. Children were on average 14.9 months, ranging from 5.8 to 26.8 months. The majority (80.3%) of mothers' had a University degree or higher as their highest form of education. While higher than the national average, it is inline with the fact that the majority of Singaporean women of childbearing age have university degrees (Singapore Department of Statistics, 2022). Mothers reported that children understood or said an average of 23.0% English words ($SD = 26.93$) and 12.3% Mandarin words ($SD = 21.50$), based on composites of 5 DVC categories as detailed in research questions 2 and 3. This sample size number ($n = 137$) is smaller from the total available stated in preregistration ($n = 189$), as inspection of the data revealed that both father and mother had filled in DVCs for 51 participants; for these participants, for maximum consistency among the sample, we utilised only the mother's report. Nonetheless, the sample size is still larger than that needed for adequate power according to preregistration power calculations.

Measures

Developmental Vocabulary Checklist (DVC). The DVC consists of a diverse collection of English, Mandarin, Tamil, Bahasa Melayu and words of unclear origin ('Red-dot') used by many different children across a wide age range of early years. For each word on the checklist, parents were asked to check off the boxes as to whether their child was able to 'understand' or 'understand and say' certain words. The DVC used for the BE POSITIVE study was of the 3rd and 4th iterations. In the 4th edition, an additional word 'digger' was added to the category 'vehicles'. Also, 'I've read all the words on this page and did not find one that my child understands or says yet' (i.e. for the category 'Small items at home') was also added in the 4th iteration. To facilitate filling out the form, the words were grouped into different categories (e.g. Animals and Creatures, Vehicles, and Food/drinks). If the child did not understand or say the word, parents or caregivers were to leave the boxes blank. These were scored as '0'. If the child understood but did not say the word or

understood and also said the word, they were to mark the boxes accordingly. These were scored as '1'. Here, 'understanding' of the word was taken as an appropriate threshold of word knowledge in the current study, as word recognition studies using eye tracking have established that parents are indeed able to identify which words their children understand the meanings of (Styles & Plunkett, 2009).

Home Language Exposure. Caregivers were asked to report the percentage of each language the mother used with their child. They were administered a questionnaire with the following question; For each language, please estimate what percentage the mother uses with the child. Answers could range from 0 to 100% and were utilised to calculate English home language exposure and Mandarin home language exposure.

Demographics. Age was calculated by subtracting the date that participants completed the DVC from the reported date of birth.

Procedure

Relevant to the current study, BE POSITIVE mothers were asked to complete the DVC electronically in English and Red Dot, followed by a choice to select a mother tongue DVC for Mandarin, Malay or Tamil. DVC questionnaires were initially administered using PDF forms. After May 2022, questionnaires were administered using REDCap, a participant management website that hosts digital tasks, including the DVC. Mothers were advised to complete the DVC within a week. They did not have to complete it in a single sitting. A separate parent caregiver questionnaire, containing demographic and language environment questions were also administered to BE POSITIVE mothers. Mothers were asked to complete this questionnaire in one setting which took approximately 15 to 30 minutes to complete.

Analytic Plan

All hypotheses and analytic plans were preregistered on the Open Science Framework along with a detailed analysis plan (https://osf.io/zbnjd/?view_only=b958f6c016ae41e9a61c2874ff771327). The English and Mandarin DVC data was cleaned before being analysed. In keeping with the aforementioned change from iteration 3 to 4 of the DVC, one word 'digger' from the English DVC category vehicles was dropped from analysis as not all participants were presented with this word. Participants with missing data were deleted. To prevent clustering effects which can bias analyses due to multiple copies of DVC per participant, only the latest time point was used for participants who filled up the DVC at two or more time points ($n = 38$ for English, $n = 35$ for Mandarin).

For research question 1 which asks about psychometric properties of the DVC, the internal consistency was measured by Cronbach's Alpha. Rasch models, a form of item response theory (Boone & Staver, 2020) that estimates the difficulty of each item was utilised here. Specifically, infit and outfit values for Rasch models of the DVC are calculated using Winsteps. DVC items were screened and those with outfit and infit mean-square values larger than 2 were removed. This is less stringent than the 0.5 to 1.5 range in the preregistration, as only values larger than 2 degrade the Rasch model and require removal (Linacre, 2002). Unidimensionality of DVC was assessed by running a principal component analysis (PCA) of residuals. We follow guidelines (Mallinson et al., 2022) which note that >50% of the variance is ideally explained by the Rasch model and that the 1st candidate for an additional dimension/groupings explains <10% of the variance.

For research question 2 which asks about patterns of vocabulary development, Winsteps was used to plot estimated person abilities with estimated item difficulties. The distribution of the

participants' estimated ability in DVC vocabulary is shown on the left side of the map, while item difficulties are plotted on the right side of the map. To test whether items in some categories (e.g. family members) have lower difficulty than items in the other categories (e.g. animals, creatures, clothing, vehicles, and food/drink), DVC scores will be compared across categories in a univariate one-way ANOVA for each language. Post hoc tests will establish whether there are significantly lower scores for some categories than the other categories. Note that we deviate from preregistration, as the category of 'family members' is not analysed, as it is highly variable in the Singaporean context, and remains under future investigation. For instance, [Choo and Styles \(2024\)](#) recorded Chinese, Malay and Tamil parent's responses about the terms they used to refer to family members, when speaking to their children and other family members. One of the most utilised terms was the names or nicknames of family members, which necessarily vary from family to family and may overwhelm a checklist if every possible name or nickname is included. Moreover, the frequency of terms also varied, with increased usage of relational terms (e.g. mummy, daddy) when referring to other adults, implying an additional source of variability beyond word knowledge that depends on the child's family composition. Taken together, such sources of variability require further examination on how this category might be sufficiently queried.

For research question 3 which examines the association between exposure and vocabulary in English and Mandarin, two multiple regression analyses were employed to examine the relationship between home language exposure and DVCs, while also considering age of the child. Model 1 specified age, home language exposure and age* home language exposure as predictors of English DVC. Model 2 specified age, home language exposure and age* home language exposure as predictors of Mandarin DVC. The interaction term is included to account for the possibility that the contribution of language exposure varies with age. These two models were run with multiple imputation with 20 imputation sets to mitigate bias arising from missing home language exposure data. Multiple imputation is a modern data technique which is less biased than complete case analysis ([Graham, 2009](#)).

Results

Research Question 1: Does the DVC Exhibit Psychometric Properties for Further Interpretation?

As shown in [Table 1](#), English and Mandarin DVC categories exhibited Cronbach alphas ranging from .91 to .98, suggestive of an 'excellent' level of reliability. Rasch modelling was utilised to further understand the properties of words in each DVC category. For Mandarin and English DVCs, 0 to 8 items per category were flagged as having outfit values >2 , which are values suggestive of items which distort or degrade the category score ([Linacre, 2002](#)). For instance, 'butt' and 'butt bum' were flagged in English and Mandarin DVC by the Rasch models, perhaps because these are 'impolite' words that parents may hesitate to endorse items even if their children understood/says these words, leading to higher amounts of unwanted variability.

Mandarin DVC categories of 'animals', 'vehicles', 'body parts', 'games & routines' and 'actions' showed support for unidimensionality. The remaining 9 Mandarin DVC categories had Rasch models which explained $<50\%$ of the variance, of which 'People' and 'Time' also had 1st contrasts explaining $>10\%$ of the residual variance. Further examination of the top three loadings of residual contrasts appear to indicate that 'People' could be grouped as 'Job-people' (Police, Teacher, Fireman) and 'Generic-People' (Man, Woman, Girl). Similar further examination for 'Time' did not suggest systematic groupings.

Table 1. Reliability and Rasch Results for Developmental Vocabulary Checklist Mandarin and English Items and Categories

	Before removal of items		After removal of items		Items removed (Outfit >2, Infit >2)	Unidimensionality	
	N	Cronbach α	N	Cronbach α		Variance explained by Rasch model (%)	Variance explained by 1 st residual Contrast (%)
Mandarin DVC categories							
动物dòng wù 'animals'	41	.98	38	.98	Ant, Lizard, Dolphin	55.2%	4.0%
交通工具jiāo tōng gōng jù 'vehicles'	15	.94	13	.95	Car, Bus	59.4%	6.4%
食物shí wù 'food/drink'	58	.98	53	.98	Cheese, Bean, Coffee, Tea, Water	44.3%	6.4%
衣物yī wù 'clothing'	18	.94	17	.94	Slipper	46.6%	9.6%
身体shēn tǐ 'body parts'	28	.96	26	.96	Butt, Nail	59.1%	7.9%
家里xiǎo wù jīn 'small items at home'	59	.98	58	.98	Light	46.8%	5.2%
房间fáng jiān/diàn 'appliances, furniture, rooms'	31	.97	25	.97	Aircon, High chair, Washing machine, Crib, Stroller ^e , Stove	46.2%	5.3%
游戏/rì cháng shēng huó xí guān 'games & routine'	29	.96	28	.96	Goodbye	51.0%	6.4%
户外wài wài 'outdoors'	52	.98	50	.98	Lightning, Mosque	46.1%	9.0%
周围的人zhōu wéi de rén 'people'	16	.94	16	.94	-	46.0%	12.4 ^a
时间shí jiān 'time'	12	.90	11	.91	First	39.5%	14.0 ^b
动作dòng zuò 'actions'	85	.99	77	.99	Bring, Fall, Love, Pretend, Say, Stop, Swim, Want	51.4%	3.5%
形容 xíng róng 'descriptions'	55	.98	49	.98	Fast, Naughty, Scared, Spoil, Well-behaved, Small	48.4%	6.1%
虚词xū cí 'function words'	40	.97	37	.97	With, One more, Some	47.6%	5.0%
English DVC categories							
Sounds	10	.92	10	.92	-	32.7%	11.8 ^c
Animals and creatures	49	.98	45	.98	Wolf, Monster, Hamster, Spider	58.2%	3.1%

(continued)

Table 1. (continued)

	Before removal of items		After removal of items		Items removed (Outfit >2, Infit >2)	Unidimensionality	
	N	Cronbach α	N	Cronbach α		Variance explained by Rasch model (%)	Variance explained by 1 st residual Contrast (%)
Vehicles	17	.95	16	.95	Scooter	59.3%	5.5%
Food/drinks	64	.98	61	.98	Drink (noun), Yakult, Ribena	48.6%	3.2%
Clothing	27	.95	26	.95	Leggings	54.0%	4.4%
Body parts	28	.98	25	.97	Head, Backside,butt,bum, Belly,tummy,stomach	58.6%	3.7%
Small items at home	70	.98	68	.98	Youtube, Pencil	51.0%	3.7%
Appliances, Furniture, Rooms	39	.97	34	.97	Fridge, Oven, Cot, Playpen, Pram,Stroller	54.3%	4.1%
Games & routine	31	.96	30	.96	Boo, peek-a-boo	55.9%	4.2%
Outdoors	55	.98	48	.98	Church, House, Leaf, Outside, Supermarket, Temple, Swing	55.3%	6.3%
People	18	.93	16	.93	Helper, Maid, Teacher	67.4%	5.8%
Words about time	13	.94	11	.94	Later, Yesterday	49.3%	10.9 ^d
Actions	93	.99	87	.99	Build, Carry, Cook, Lick, Touch, Work	52.3%	2.9%
Descriptions	55	.99	47	.99	Angry, Careful, Dark, Gentle, Happy, Purple, Smelly, Spoil	55.2%	5.1%
Function words	52	.98	48	.98	I, My, Where, In	48.5%	5.0%

Note: ^aThree most positive loadings are Man, Woman, Girl and three most negative loadings are Police, Teacher, Fireman.

^bThree most positive loadings are Yesterday, Time, Today and three most negative loadings are Later, Night, Afternoon.

^cThree most positive loadings are Yum yum, Oh Oh, Choo Choo and three most negative loadings are Woof, Meow, Quack.

^dThree most positive loadings are Night, Morning, Afternoon and three most negative loadings are Now, Today, First/Eat First.

^eThe only item that had Infit >2.

For English DVC, categories of ‘Animals and Creatures’, ‘Vehicles’, ‘Clothing’, ‘Body Parts’, ‘Small items at Home’, ‘Appliances Furniture and Rooms’, ‘Games and Routines’, ‘Outdoors’, ‘People’, ‘Actions’ and ‘Descriptions’ showed support for unidimensionality. The remaining 4 English DVC categories had Rasch models which explained <50% of the variance, of which ‘Sounds’ and ‘Words about Time’ also had 1st contrasts explaining >10% of the residual variance. Further examination of the top three loadings of residual contrasts appear to indicate that ‘Sounds’ could be grouped as ‘Sounds-animals’ (‘Woof’, ‘Meow’, ‘Quack’) and ‘Sounds’ (‘Yum yum’, ‘Oh Oh’, ‘Choo choo’). Similar further examination for ‘Word about Time’ appeared to indicate that it could be grouped as ‘Time-Day-Night’ (‘Morning’, ‘Afternoon and Night’) and ‘Time’ (‘Now’, ‘Today’, ‘First/Eat First’).

Research Question 2: Patterns of English and Mandarin Vocabulary Development from the DVC

Figures 1(A) and (B) display the difficulties of items for DVC English categories of vehicles and food and drink. See [Online Supplement S1](#) for plots of all 5 DVC English categories as per research question 2, with only the smallest and largest DVC categories shown here due to space constraints. Visual inspection of the ‘Vehicles’ category indicates ‘crane’ as the most difficult item and ‘car’ as the easiest. Within ‘Food and Drink’, ‘soft drink’ is the most difficult and ‘banana’ the easiest. DVC items of similar levels of difficulty were also present, such as ‘fire/engine/helicopter/taxi/van’ for ‘Vehicles’ and ‘butter/jam/nuts/potatochip/sandwich’ for ‘Food and Drink’. DVC English items appear to largely follow the distribution of the estimated person abilities. However, there are fewer items for children estimated to have the highest and lowest ability in saying/understanding ‘Food and Drink’ or ‘Vehicle’ vocabulary. This appears more pronounced for the younger band of 0 to 1 years, where the person ability distribution appears to largely consist of a single bar with little variation. Hence, items may not be sensitive to variation at the lowest levels of language ability at 0 to 1 years. This is in line with the general field of language development, in which infants under the age of one have not yet begun to demonstrate the ‘vocabulary explosion’ which tends to occur in second half of the second year ([Gomez Diaz et al., 2024](#)).

Figures 1(C) and (D) display the estimated difficulties of items for DVC Mandarin categories of Vehicles and Food and Drink. See [Online Supplement S1](#) for plots of 5 DVC Mandarin categories as per research question 2. Visual inspection indicates ‘lorry’ or ‘van’ as the most difficult and ‘airplane’ as the easiest for vehicles. There were multiple items that were in the most difficult band of ‘desert/jam/jelly/peanut butter/sauce/softdrink/sugarcane’ whilst ‘apple’ or ‘bread’ was the easiest for ‘Food and Drink’. DVC Mandarin items appear to be in keeping with children of higher Mandarin vocabulary ability, but not so for children with lower estimates of Mandarin vocabulary ability. This is most pronounced for 0 to 1 year band of children and also noticeable for 1 to 2 year band, with most of the items in the Mandarin DVC categories above the majority of children’s estimated Mandarin vocabulary ability.

Taken together, this is in keeping with the local profile of bilingualism to be dominant in English with relatively less usage and exposure to Mother Tongue such as Mandarin. All figures were estimated on $n = 136$ participants, as 1 child’s date of DVC attempt could not be estimated and was dropped. A further $n = 13$ participants did not have Mandarin date estimates, and English date estimates were used in place, as these did not differ statistically (English DVC $M = 14.837$ Months, $SD = 4.87$, Mandarin DVC $M = 14.867$ Months, $SD = 4.87$, $n = 122$, one-sample $t = -1.603$, $p = .112$).

Figures 2(A) and (B) display one-way repeated ANOVAs for English and Mandarin DVC categories, respectively. A main effect was detected, which continued to hold after a correction of Greenhouse-Geiser was applied for violation of sphericity, indicating a statistically significant

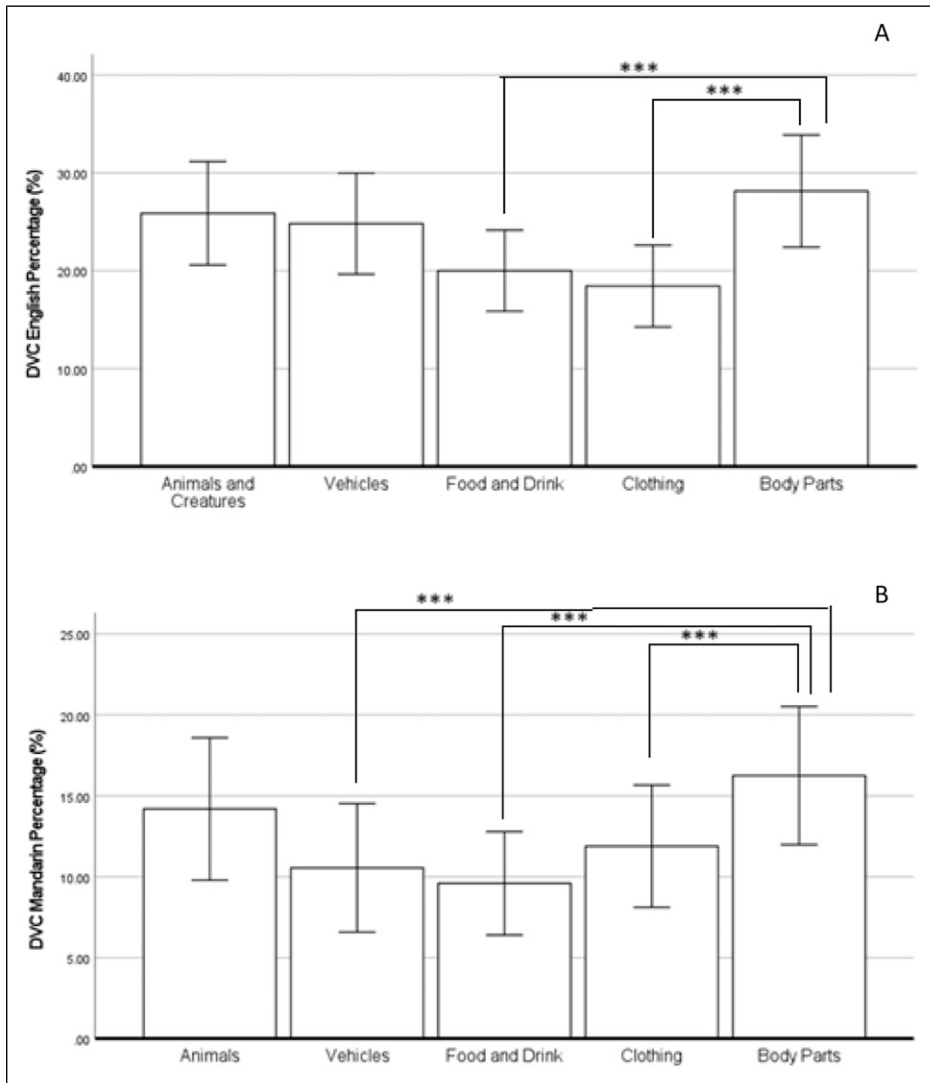


Figure 2. One-way repeated ANOVA across (A) English DVC and (B) Mandarin DVC

was found, with older age predicting DVC English scores ($b = 3.33$, $SE = 0.94$, $p < .001$) and interacting with exposure at home to predict DVC Mandarin scores ($b = 0.03$, $SE = 0.01$, $p = .022$).

To further examine the nature of this significant interaction, linear regression models 3 and 4 were run. Among younger aged participants 0 to 1 years, exposure to mandarin did not predict higher percentage of DVC Mandarin items ($b = -0.02$, $SE = 0.14$, $p = .86$). Among older aged participants 1 to 2 years, exposure to mandarin did predict higher percentage of DVC Mandarin items ($b = 0.44$, $SE = 0.12$, $p < .001$). As shown in Table 2, compared to complete case analyses of models 3 and 4, a similar pattern of findings were found where the effects of exposure to mandarin are amplified among older 1 to 2 years rather than younger 0 to 1 year participants. Further sensitivity analyses in Online Supplement S2 were conducted with socio-economic status as a covariate in these models. In brief, the same pattern of findings was found, with exposure to mandarin predicting higher percentage of DVC mandarin, only among older 1 to 2 years ($b = 0.46$,

Table 2. Multiple Regression of Age and Exposure Predicting DVC English and Mandarin

	Multiple imputation	Complete case	B
	b (SE)	b (SE)	
Model 1: English DVC ^a			
Age	3.60 (1.15)**	3.45 (0.96)***	0.72
English Exposure	0.00 (0.24)	-0.06 (0.25)	-0.06
Age * English exposure	0.00 (0.02)	0.00 (0.02)	-0.01
Model 2: Mandarin DVC ^b			
Age	0.60 (0.59)	0.30 (0.65)	0.07
Mandarin Exposure	-0.17 (0.23)	-0.29 (0.20)	-0.35
Age * Mandarin exposure	0.03 (0.01)*	0.04 (0.01)***	0.91
Decomposition of significant interaction			
Model 3: Mandarin DVC for 0 to 1 yr ^c			
Mandarin Exposure	-0.02 (0.14)	0.01 (0.01)*	0.50
Model 4: Mandarin DVC for 1 to 2 yr ^d			
Mandarin Exposure	0.44 (0.12)***	0.52 (0.12)***	0.53

Note: ^aSample size is 137 for Multiple Imputation and 68 for Complete Case.

^bSample size is 137 for Multiple Imputation and 65 for Complete Case.

^cSample size is 42 for Multiple Imputation and 17 for Complete Case.

^dSample size is 94 for Multiple Imputation and 50 for Complete Case.

SE = 0.11, $p < .001$) and not younger 0 to 1 years old participants ($b = 0.02$, SE = 0.13, $p = .86$). [Figure 3](#) provides a visualisation of this amplified effect, with a more sloped line present for 1 to 2 years over 0 to 1 years.

Discussion

We examined the reliability and patterns of English and Mandarin vocabulary development among a sample of 137 children ranging from 5-27 months growing up in Singapore. We first asked if the Developmental Vocabulary Checklist (DVC) showed reliability and could be understood as a type of psychometric scale known as an item response theory Rasch model (Boone & Staver, 2020). We then considered if the DVC English and Mandarin vocabulary responses were in keeping with expectable patterns of vocabulary development, such as ‘Bodyparts’ words being relatively easier as they appear early on at 12 to 18 months, in local milestones of language (Lim et al., 1994). Lastly, the association of home language exposure with DVC English and Mandarin was examined, to understand if this expectable association was also present here.

First, the DVC showed high levels of reliability ($\alpha = .91$ to $.99$) across 15 categories of English and 14 categories of Mandarin words. This is in keeping with reports of high reliability among parental vocabulary checklists in English and non-English languages from international studies (Holm et al., 2023; Kachergis et al., 2022). Nonetheless, future refinement of the DVC could consider the following as candidates for further examination. Residual analyses indicated that the English DVC word category of ‘Sounds’ could be further separated into groupings of ‘Human and Non-human Sounds’, and ‘Words about Time’ could be separable into ‘24-hour and Time’. Likewise, the Mandarin DVC word category of ‘People’ could be further separated into groupings of ‘Jobs and People’. Poor-fitting words have been found when parental vocabulary checklists are fit to item response theory models (Kachergis et al., 2022). Similarly, 0 to 8 individual words for every DVC category showed high outfit values that was not expected by a

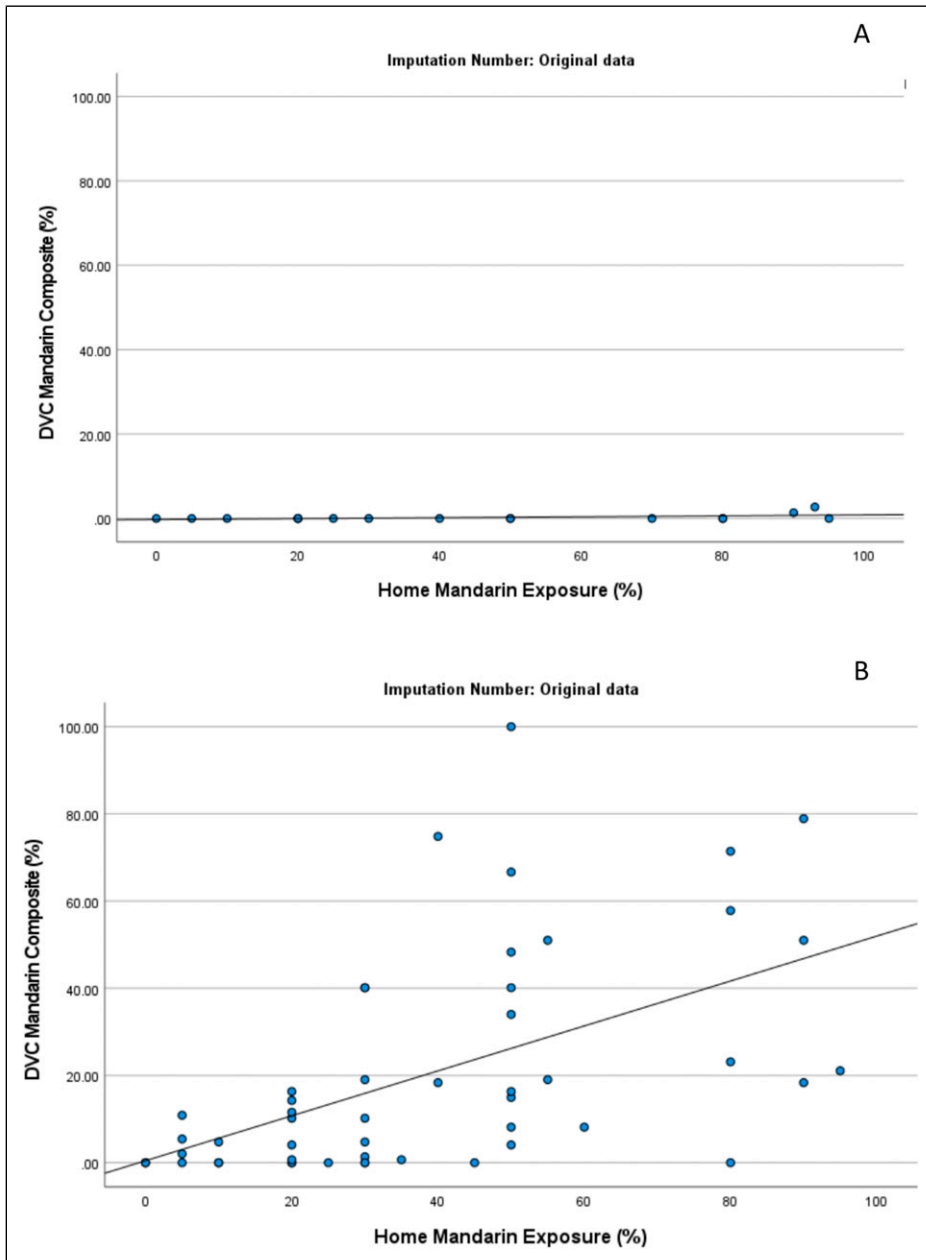


Figure 3. Decomposition of exposure on mandarin DVC split by age of (A) 0 to 1 years [$b = -0.02$, $SE = 0.14$, $p = .86$] and (B) 1 to 2 years [$b = 0.44$, $SE = 0.12$, $p < .001$]

Rasch model. These words appear to be specific exemplars of the category which require additional factors beyond that of the DVC category to explain. For example, ‘Mosque’, ‘Church’ and ‘Temple’ in the category ‘Outdoors’ could be influenced by religious status of families, whilst ‘Yakult’ and ‘Ribena’ in the category ‘Food/Drink’ are brand names and could be influenced by brand preference of families. ‘Butt’ and ‘Backside/bum’ could be construed as impolite words to

report in a checklist and could be influenced by social desirability. ‘Nail’ could be an ambiguous item as it refers to a bodypart as well as a tool. Taken together, these findings constitute evidence-based considerations for constructing reliable vocabulary batteries, if validated in future studies. Specifically, consideration of words with multiple or specialised meanings (‘Nail’ or ‘Yakult’) or that may inadvertently depend on language-related processes when read (e.g. social desirability and religiosity).

Second, a higher percentage of ‘Bodyparts’ words were endorsed as compared to other categories of Clothing and Food and drink, across both the English and Mandarin DVC. This supports the validity of the DVC as a measure of vocabulary, as this is in line with nationally normed local milestones of language which report ‘Bodyparts’ as one of the earliest categories of words to be known (Lim et al., 1994). Face validity of the DVC was also observed across 10 English and Mandarin categories (Online Supplement S1). For example, ‘alien’ was estimated as the most difficult and ‘dog’ as the easiest word in English DVC animals, or ‘housefly or mosquito’ estimated as the most difficult and ‘dog’ as the easiest word in Mandarin DVC animals. English and Mandarin DVC plots (Figure 1) suggests words are targeted towards measuring variation in language proficiency among 1 to 2 year olds rather than 0 to 1 year olds, perhaps because babbling or vocalisations are uttered at this very young age in place of words. A visible pile-up of responses signifying a floor effect among 1 to 2 year olds for DVC English, and perhaps even more pronounced for DVC Mandarin, was present. This suggests that easier English and Mandarin vocabulary items, if they can be identified, are needed to quantify vocabulary at this very young age. This is in keeping with relatively low rates of sensitivity and specificity among screeners of language difficulties at 2-3 years in Singapore (Agarwal et al., 2024), suggesting further research is needed to more accurately measure low language proficiency at early ages.

Lastly, higher exposure to Mandarin at home amongst the older aged group of children associated with a higher percentage of Mandarin words known/said across 5 DVC categories. However, English home exposure did not significantly predict English words known/said. This is similar to local studies which have found associations with Mandarin but not in English (He et al., 2020; Goh et al., 2020; Wu et al., 2020). One possible explanation is a threshold account (Thordardottir, 2011) which suggests that exceeding a certain amount of exposure is no longer meaningful for greater vocabulary proficiency. Owing to the English-dominant nature of bilingualism in Singapore (Singapore Department of Statistics, 2022), the threshold for English but not Mandarin exposure could have been met. We observed a significant interaction of age*Mandarin exposure, indicating that the effect of Mandarin exposure was only for the older age band of 1 to 2 years. This could be due to a floor effect at 0 to 1 year old where Mandarin words may not be in use, and thus showed little variation on the Mandarin DVC.

For professionals and researchers involved in measuring language in multilingual societies, they may need to consider if additional items of lower difficulty are needed for languages which are less frequently used in their society. Without which, this could lead to an unwanted floor effect in quantification of less dominant languages of the society, such as in our results for Mandarin vocabulary across all ages, as only a few items were at the tail ends of the language distribution. This can be addressed by considering more exemplars at a lower level of difficulty (e.g. more items such as ‘dog’ as compared to ‘mosquito’ for Mandarin). Words used in less formal settings such as those that do not originate in an official language, may also be more widely used and therefore of lower difficulty. For example, simple onomatopoeic expressions such as ‘ng-ng’ (‘poop’) and ‘shee-shee’ (‘wee’), which appear in the specialized ‘Red Dot’ sections of the DVC, may be relevant and examined in future studies.

Findings have limited generalisability as sampling was not designed to be nationally representative and continues to add more participants to its enrolment. Nonetheless this sample size is similar to other published work (Holm et al., 2023) and is adequate for an initial analysis, in

keeping with the goals of the current study. Approximately 50% (68 out of 137) of caregiver questionnaires with home language use and exposure data were completed, limiting our ability to provide descriptives of the sample. Also, a single question was utilised to measure home language exposure and was only available for a subsample of those who completed the DVC, limiting the reliability of findings with it. Multiple imputation was utilised to minimise bias arising from missing data on language exposure.

To allow for conclusive findings beyond this initial analysis, larger samples in each age band with increased diversity across socio-economic status and ethnic background need to be conducted. Such a study would set the stage for more reliable and valid measurement of multilingual lexical growth among infants and toddlers, which appears to have mixed and limited findings both locally and internationally (Weisleder et al., 2024; Woon et al., 2020). Direct measures of language such as audio recordings of parent-child interactions (Pérez-Navarro & Lallier, 2025) or assessments of child language (Goh et al., 2017) can be used to mitigate self-report biases, strengthening construct validity if found to correlate with parent-report type measures such as the DVC. Comprehensive measures of language exposure would also be desirable, which could include adaptation of internationally validated questionnaires of language exposure or time-based diaries (Kascelan et al., 2022).

One type of future extension to consider repeatedly administers the DVC to track longitudinal growth in vocabulary would also be informative, as has been shown among English-Spanish bilingual samples (Hoff, 2021). A second extension type to consider examines the DVC for use among lower proficiency samples, to understand how child directed or simpler onomatopoeic expressions such as ‘ng-ng’ (‘poop’) and ‘shee-shee’ (‘wee’), which appear in the specialized ‘Red Dot’ sections of the DVC could be of lower difficulty and therefore suitable at lower proficiencies. A third extension type to consider iteratively refines the DVC items, reporting on the criteria, informant background, development and iteration processes, which would improve transparency and content validity, which are not currently fully available for the DVC and other local measures in Singapore. A fourth type of extension would be to administer different types of measures to the same sample of infants and toddlers. This may provide more information on which type of measure (i.e. screening tools vs vocabulary checklists, examiner-administrated vs parent report and translated/adapted tools vs locally developed) yields the most ideal set of psychometric properties.

In sum, a novel parental checklist called the DVC shows evidence of high reliability and associations with exposure and showing anticipated patterns of vocabulary development in both English and Mandarin. Nonetheless, further refinements to the DVC can consider examination of additional groupings found in ‘Time’, ‘People’ and ‘Sounds’, as well as flagged poor-fitting items which may require additional factors for explanation. Findings of higher exposure associating with higher vocabulary in Mandarin but not in English are in keeping with the English-dominant nature of bilingualism in Singapore, where a threshold amount of English may have been reached but not in Mandarin. Finally, floor effects and a lack of item correspondence at the lower tail end of estimated language proficiency suggest that easier vocabulary items are needed, if one is interested in quantifying English and Mandarin among the very young and those with low proficiency.

ORCID iD

Shaun Kok Yew Goh  <https://orcid.org/0000-0002-0432-8439>

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research/project is supported by the Office of Education Research’s Centre for

Research in Child Development (grant number CRCD 01/19 AR), Beginning Early: Singapore's Ongoing Study starting in Infancy of Twenty-first-century-skills, Individual differences, and Variance in the Environment. This research/project is supported by the National Research Foundation, Singapore, and the Agency for Science Technology and Research (A*STAR), Singapore, under its Prenatal/Early Childhood Grant (Grant No. H22P0M0006). Tools developed for this project were created with the support of Singapore's National Research Foundation under the Science of Learning grant, 'How do language mixes contribute to effective bilingualism and effective biliteracy in Singapore' (NRF2016-SOL002-011) and Nanyang Technological University under CRADLE@NTU grant (JHU IO 90071537). This research/project was also funded by the Singapore Ministry of Education (MOE) under the Strategic Research Fund (SRF 05/24 WCY). The funds were administered by the National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE or NIE.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

References

- Ackermann, L., Hepach, R., & Mani, N. (2020). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, 23(3), e12915. <https://doi.org/10.1111/desc.12915>
- Agarwal, P. K., Xie, H., Rema, A. S. S., Meaney, M. J., Godfrey, K. M., Rajadurai, V. S., & Daniel, L. M. (2024). Concurrent validity of the ages and stages questionnaires with bayley scales of infant Development-III at 2 years–Singapore cohort study. *Pediatrics & Neonatology*, 65(1), 48–54. <https://doi.org/10.1016/j.pedneo.2023.03.013>
- Agarwal, P. K., Xie, H., Rema, A. S. S., Rajadurai, V. S., Lim, S. B., Meaney, M., & Daniel, L. M. (2020). Evaluation of the ages and stages questionnaire (ASQ 3) as a developmental screener at 9, 18, and 24 months. *Early Human Development*, 147, 105081. <https://doi.org/10.1016/j.earlhumdev.2020.105081>
- Bayley, N. (2006). *Bayley scales of infant and toddler Development– third edition: Administration manual*. Harcourt Assessment.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531. <https://doi.org/10.1017/S1366728909990423>
- Boone, W. J., & Staver, J. R. (2020). *Advances in rasch analyses in the human sciences* (pp. 287–302): Springer.
- Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, 19(6), 918–932. <https://doi.org/10.1111/desc.12343>
- Choo, R. Q., & Styles, S. J. (2024). Kinship terms of address and reference among families in Singapore [Poster]. *Open Science Framework*. <https://doi.org/10.17605/OSF.IO/4FR52>
- Dale, P. (2015). *Adaptations, not translations*. MacArthur–Bates Communicative Development Inventories. <https://mb-cdi.stanford.edu/documents/AdaptationsNotTranslations2015.pdf>
- Dixon, L. Q. (2005). Bilingual education policy in Singapore: An analysis of its sociohistorical roots and current academic outcomes. *International Journal of Bilingual Education and Bilingualism*, 8(1), 25–47. <https://doi.org/10.1080/jbeb.v8.i1.pg25>

- Goh, S. K., Tham, E. K., Magiati, I., Sim, L., Sanmugam, S., Qiu, A., & ... Rifkin-Graboi, A. (2017). Analysis of item-level bias in the Bayley-III language subscales: the validity and utility of standardized language assessment in a multilingual setting. *Journal of Speech, Language, and Hearing Research, 60*(9), 2663–2671.
- Goh, S. K., Yang, H., Tsotsi, S., Qiu, A., Chong, Y. S., Tan, K. H., & ... Rifkin-Graboi, A. (2020). Mitigation of a prospective association between early language delay at toddlerhood and ADHD among bilingual preschoolers: evidence from the GUSTO cohort. *Journal of Abnormal Child Psychology, 48*(4), 511–523.
- Gómez Díaz, M., Fibla, L., Tsui, R. K. Y., & Byers-Heinlein, K. (2024). Testing theories of the vocabulary spurt with monolingual and bilingual infants. *Developmental Psychology, 60*(8), 1357–1371. <https://doi.org/10.1037/dev0001777>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- He, S., Ng, S. C., O'Brien, B. A., & Fritzsche, T. (2020). Child, family, and school factors in bilingual preschoolers' vocabulary development in heritage languages. *Journal of Child Language, 47*(4), 817–843. <https://doi.org/10.1017/S0305000919000904>
- Hoff, E. (2021). Why bilingual development is not easy. *Advances in Child Development and Behavior, 61*, 129–167. <https://doi.org/10.1016/bs.acdb.2021.03.002>
- Holm, E., Hansen, P. B., Romøren, A. S. H., & Garmann, N. G. (2023). The Norwegian CDI-III as an assessment tool for lexical and grammatical development in preschoolers. *Frontiers in Psychology, 14*, 1175658. <https://doi.org/10.3389/fpsyg.2023.1175658>
- Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., & Frank, M. C. (2022). Online computerized adaptive tests of children's vocabulary development in English and Mexican Spanish. *Journal of Speech, Language, and Hearing Research, 65*(6), 2288–2308. https://doi.org/10.1044/2022_JSLHR-21-00372
- Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S., & De Cat, C. (2022). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition, 25*(1), 29–41. <https://doi.org/10.1017/s1366728921000390>
- 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>
- Lim, H., Chan, T., & Yoong, T. (1994). Denver developmental screening test (DDST). *Singapore Medical Journal, 35*(2), 156–160.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions, 16*(2), 878.
- Mallinson, T., Kozlowski, A. J., Johnston, M. V., Weaver, J., Terhorst, L., Grampurohit, N., Ehrlich-Jones, L., Heinemann, A. W., Melvin, J., Sood, P., & Van de Winckel, A. (2022). Rasch reporting guideline for rehabilitation research (RULER): The RULER statement. *Archives of Physical Medicine and Rehabilitation, 103*(7), 1477–1486. <https://doi.org/10.1016/j.apmr.2022.03.013>
- McAvinue, L. P. (2025). The association between socioeconomic status and oral vocabulary in young children: A quantitative meta-analysis. *First Language, 01427237251397678*. <https://doi.org/10.1177/01427237251397678>
- Ministry of Education. (2010). Nurturing active learners and proficient users. In: *Mother tongue languages review committee report*. Singapore: Ministry of Education.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191–230. <https://doi.org/10.1017/S0142716407070117>
- Pakir, A. (1993). Two tongue tied: Bilingualism in Singapore. *Journal of Multilingual & Multicultural Development, 14*(1-2), 73–90. <https://doi.org/10.1080/01434632.1993.9994521>

- Pérez-Navarro, J., & Lallier, M. (2025). The contribution of the amount of linguistic exposure to bilingual language development: Longitudinal evidence from preschool years. *Child Development, 96*(1), 176–191. <https://doi.org/10.1111/cdev.14164>
- Pratt, A. S., Durant, K., Peña, E. D., & Bedore, L. M. (2024). Modeling dimensionality of bilingual kindergarteners' language knowledge in Spanish and English. *Journal of Speech, Language, and Hearing Research, 67*(7), 2244–2268. https://doi.org/10.1044/2024_JSLHR-22-00140
- Rifkin-Graboi, A., Khng, K. H., Cheung, P., Tsotsi, S., Sun, H., Kwok, F., Xie, H., Yang, Y., Chen, M., Ng, C. C., Hu, P. L., & Tan, N. C. (2019). Will the future BE POSITIVE? Early life experience as a signal to the developing brain pre school entry. *Learning: Research and Practice, 5*(2), 99–125. <https://doi.org/10.1080/23735082.2019.1674907>
- Singapore Department of Statistics. (2022). *Population trends 2022*. <https://www.singstat.gov.sg/-/media/files/publications/population/population2022.pdf>
- Singh, L., Cheng, Q., Tan, S. H., Tan, A., & Low, Y. L. (2022). Language acquisition in a multilingual society: English vocabulary norms and predictors in Singaporean children. *Child Development, 93*(1), 288–305. <https://doi.org/10.1111/cdev.13676>
- Squires, J., Twombly, E., Bricker, D., & Potter, L. (2009). *ASQ-3 user's guide* (3rd ed.). Paul H. Brookes Publishing Co.
- Styles, S. J., & 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>
- Sun, H., Yin, B., Amsah, N. F. B. B., & O'Brien, B. A. (2018). Differential effects of internal and external factors in early bilingual vocabulary learning: The case of Singapore. *Applied Psycholinguistics, 39*(2), 383–411. <https://doi.org/10.1017/s014271641700039x>
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism, 15*(4), 426–445. <https://doi.org/10.1177/1367006911403202>
- Weisleder, A., Friend, M., Tsui, A. S. M., & Marchman, V. A. (2024). Using parent report to measure vocabulary in young bilingual children: A scoping review. *Language Learning, 74*(2), 468–505. <https://doi.org/10.1111/lang.12617>
- Woon, F. T. (2018). *Linguistic sound symbolism and reading development: Sound-shape matching and predictors of reading in multilingual Singapore*: Nanyang Technological University. Master's thesis.
- Woon, F. T., Cai, S., & Styles, S. J. (2020). Understanding the puzzle of vocabulary development in multilingual Singapore and its missing pieces. In: *International congress of infant studies (ICIS 2020)*.
- Woon, F. T., & Styles, S. J. (2021). *Red-dot baby-talk quiz*. DR-NTU.
- Wu, C. Y., O'Brien, B. A., Styles, S. J., & Chen, S. H. A. (2020). The impact of bilingualism on skills development and education. In: *Transforming teaching and learning in higher education: A chronicle of research and development in a Singaporean context* (pp. 47–69). Springer Nature.