Studies in Second Language Acquisition **42** (2020), 279–325
doi:10.1017/S0272263119000597

Research Article

PREDICTING LANGUAGE PROFICIENCY IN BILINGUAL CHILDREN

Cécile De Cat*

University of Leeds

Abstract

Using advanced quantitative methods, this article demonstrates that cumulative exposure to the school language is the best language experience predictor of proficiency in that language (as indexed by sentence repetition, lexical semantic, and discourse semantic tasks) in a highly diverse group of 5- to 7-year-old bilingual children in monolingual education. An objective method is proposed to identify the amount of school language experience beyond which bilingual children are likely to perform within the monolingual range, and show that relative passivity in the home language does not translate into better school language proficiency. Socioeconomic status is shown to interact in complex ways with language exposure, such that it is only above a certain level of exposure to the school language that the benefits of a more privileged background have a tangible impact on school language proficiency. To tease apart the effect of environmental predictors from the effect of cognitive factors, memory and cognitive flexibility measures are included as covariates in all analyses.

INTRODUCTION

Research on language proficiency in bilingual children has revealed a significant impact of language exposure, socioeconomic status (SES), and memory, although the interplay of all these factors has seldom been explored. Many studies tend to focus on predicting vocabulary, and until recently most research was based on relatively homogeneous home-language groups (although see Floccia et al., 2018 and Paradis, Rusk, Duncan, & Govindarajan, 2017).

While the impact of language exposure on language proficiency is uncontroversial, many questions remain about the relationship between the two. What aspects of language

The experiment in this article earned an Open Data badge for transparent practices. The materials are available at https://osf.io/g3pdx/.

*Correspondence concerning this article should be addressed to Cécile De Cat, School of Languages, Cultures and Societies, University of Leeds, Leeds LS2 9JT, UK. E-mail: c.decat@leeds.ac.uk.

Copyright © Cambridge University Press 2019

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

experience need to be taken into account? Is exposure alone a sufficient indicator? Does language *use* have a significant impact? Does past language experience matter (e.g., an initial monolingual period)? Is it possible to predict when bilingual children's proficiency in the school language (SL) can be expected to fall within the monolingual range, assuming the absence of developmental disorder?

SES is also well established as an important environmental predictor, but how to interpret its effect (e.g., in relation to variations in quantity and/or quality of input) remains unclear.

This study offers an investigation of environmental predictors of language proficiency in bilingual children, informed by an in-depth review of the relevant literature and the use of advanced quantitative methods. As explained in the next section, the novelty of our approach lies in (a) the deliberate choice of a highly heterogeneous participant group, (b) the breadth of proficiency aspects measured, (c) the comparison of alternative estimates of language experience as predictors of language proficiency, (d) the investigation of interactions between environmental predictors, and (e) the investigation of language exposure thresholds to inform comparisons between bilingual and monolingual groups. As such, our focus is both methodological and descriptive, so as to inform future research with a more theoretical focus.

THE STATE OF THE ART

KEY ASPECTS OF BILINGUAL LANGUAGE EXPERIENCE

Language development requires language input, that is, direct evidence relevant to the acquisition of particular aspects of language. Language acquisition theories differ in what constitutes necessary input, and modeling the availability of usable input in the language environment is not straightforward (see Carroll, 2017 for discussion). It is, however, reasonable to assume that language exposure is a reliable proxy for language input. Indeed, variation in language exposure has been shown to affect many aspects of language development in bilingual children (see e.g., Grüter & Paradis, 2014; Hoff et al., 2012; Paradis, 2011; Paradis et al., 2017; Place & Hoff, 2011; Thordardottir, 2011; Thordardottir, Rothenberg, Rivard, & Naves, 2006; Unsworth, 2013b, 2017; Unsworth, Persson, Prins, & De Bot, 2014).

Bilingual children's language exposure is by its very nature extremely varied in terms of relative quantity and context (De Houwer, 2007). Some children are exposed to both languages at birth or thereafter, others are exposed to a second language later in childhood (De Houwer, 2011; Meisel, 2009). Relative exposure to the two languages may also vary over time, leading to different amounts of cumulative exposure to each over the course of their lives (Gutiérrez-Clellen & Kreiter, 2003; Thordardottir, 2011; Unsworth, 2013b). Given the heterogeneity of the bilingual language experience, it is now recognized that bilingualism should be conceptualized as a continuous variable rather than a categorical one (Kaushanskaya & Prior, 2015; Luk, 2015).

Studies that have investigated the relationship between language exposure and language proficiency in bilingual children have typically focused on a snapshot of the child's language experience at the time of assessment (De Houwer, 2009; Jia & Aaronson, 2003; Paradis, 2011). Recently, the notion of *cumulative* exposure has gained traction in the context of a more holistic approach to the child's language history

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

(Gutiérrez-Clellen & Kreiter, 2003; Unsworth, 2013b). Indeed, proficiency at time x, however it is measured, is the result of the child's language experience over time and not just at time x. This is particularly relevant in the case of sequential bilingual children who may experience rather abrupt changes in the relative exposure to their two languages when they start formal childcare placements or school.

The impact of bilingual children's language *use* on proficiency has hitherto received comparatively less attention (but see Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010; Ribot, Hoff, & Burridge, 2017; Unsworth, 2015). Research with adult bilinguals has shown that the opportunity to use the language is an important factor in determining the degree of bilingualism (Jia, Aaronson, & Wu, 2002). There is reason to believe that, for children too, the extent to which they use a language will be a determinant of the proficiency attained in that language. Some studies with child bilinguals explicitly include a measure of language use, based on children's experience at the time of testing (Bedore et al., 2012; Bohman et al., 2010; Goldstein, Bunta, Lange, Rodriguez, & Burrows, 2010; Gutiérrez-Clellen & Kreiter, 2003). Similarly to language exposure, language use is likely to change over time as children's circumstances and their opportunities to use their two languages vary.

This study investigates how to best predict language proficiency based on language experience, by comparing different ways of operationalizing language experience as a continuous measure. First, we ask if *a combined measure of cumulative language exposure and use* predicts (various aspects of) proficiency in the school/societal language. Then we compare that measure to alternative measures (cumulative exposure only, and current measures of exposure and use) to ascertain which is the most informative predictor in the population of interest.

COGNITIVE AND SOCIAL DETERMINANTS OF LANGUAGE PROFICIENCY

Memory skills, both short-term memory (STM) and working memory (WM) have been shown to be positively associated with language skills in monolingual children (Baddeley, Gathercole, & Papagno, 1998). STM is strongly associated with word learning skills (Avons, Wragg, Cupples Wragg, & Lovegrove, 1998; Majerus, Poncelet, Greffe, & Van der Linden, 2006), but the extent to which it is predictive of other language skills is less clear. Some studies have found a positive relationship between STM and the *comprehension* of syntactically complex sentences in adults (Lauro, Reis, Cohen, Cecchetto, & Papagno, 2010; Papagno, Cecchetto, Reati, & Bello, 2007), others have failed to find such a direct link in children, in spite of a strong effect of STM on sentence *repetition* (SRep) abilities (Willis & Gathercole, 2001). WM, however, is more clearly associated with metalinguistic abilities (McDonald, 2008) and sentence comprehension (Montgomery, 1995; Montgomery, Magimairaj, & O'Malley, 2008), particularly the comprehension of structurally complex sentences (see Kidd, 2013) for a recent review of the role of WM in children's sentence comprehension.

As for children who are exposed to more than one language, Thorn and Gathercole (1999) have shown a positive relationship between the ability to recall digits and nonwords in two STM tasks and the receptive and expressive vocabulary skills of English-French bilingual between the ages of 4 and 8 in their two languages. STM as measured by nonword repetition has been shown to be more weakly associated with

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

amount of exposure than vocabulary skills are (Thordardottir & Brandeker, 2013). To date very few studies have simultaneously investigated STM and WM in bilingual school-age children, and their relationship with lexical and grammatical skills. Engel de Abreu, Gathercole, and Martin (2011) explored this relationship in a group of Luxembourgish-German bilinguals between 5 and 6 years of age. STM was found to be predictive of vocabulary knowledge independently of WM, suggesting a highly specific association. WM and fluid intelligence (jointly interpreted as the capacity for controlled processing) were found to predict higher-order language processing (e.g., the understanding of syntactic contrasts). Compatible results were obtained by Verhagen and Leseman (2016), based on the comparison between a group of monolingual Dutch 5 year olds and an age-matched group of Turkish-Dutch bilingual children exposed to both languages naturalistically. The results of a confirmatory factor analysis showed that verbal STM predicted receptive vocabulary skills in Dutch, and that both verbal STM and verbal WM were predictive of grammar skills after controlling for vocabulary and SES.

While STM and WM can be considered child-internal determinants of language acquisition, a child-external factor that has been shown to be of great significance is the SES of the child, indirectly measured using parental information. SES has been variously conceptualized in the literature in terms of household income, parental occupation, or parental (often maternal) education. Given the typically high correlation between these variables, studies have often used them interchangeably.

The relative contribution of SES to bilingual children's language skills is not easy to identify. As highlighted by Hoff (2013), both bilingualism and SES have an impact on language development, and not enough research has yet considered the relative contribution of amount of language exposure and SES to disentangle these two factors. However, those studies that have included bilingual children from different SES backgrounds (see the following text) have typically found that SES can be an independent predictor of language proficiency.

To date, four studies have simultaneously evaluated the effect of SES and bilingualism on children's different linguistic and cognitive domains. Calvo and Bialystok (2014) focused on receptive vocabulary, nonverbal intelligence, and executive function tasks in a linguistically heterogeneous group of 6 to 7 year olds in Canada with 26 different home languages (HLs). Gathercole, Kennedy, and Môn Thomas (2016) assessed vocabulary, grammar, and cognitive skills in Welsh-English bilinguals in Wales. Chiat and Polišenská (2016) included Turkish-English and Spanish-English bilinguals in England and included measures of nonword repetition and vocabulary. Most recently Meir and Armon-Lotem (2017) addressed the independent and combined effect of SES and bilingualism in Russian-Hebrew bilinguals in Israel on expressive vocabulary and verbal STM.

The results for vocabulary measures are consistent across the four studies showing significant effects of SES and bilingualism, with bilingual children and children from lower SES performing more poorly. A lack of interaction between the two factors indicates that SES affects monolingual and bilingual children similarly.

Performance on a test of nonverbal intelligence in the Calvo and Bialystok (2014) study was not affected by either SES or bilingualism, while executive functioning was adversely affected by low SES, but positively affected by bilingualism. In the study by Gathercole et al. (2016), SES, but not bilingualism, was a significant predictor of performance on cognitive measures.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

As for memory, Chiat and Polišenská (2016) report that neither SES nor bilingualism affected children's performance on their nonword repetition tasks. By contrast, Meir and Armon-Lotem (2017) found that performance on a forward digit task was only affected by SES but not by bilingualism, even after controlling for vocabulary. For a nonword repetition task neither SES nor bilingualism had an effect, with the exception of a negative effect of bilingualism on a subset of nonwordlike stimuli that disappeared after controlling for vocabulary. In a SRep task, there was a significant effect both of SES and bilingualism, but the latter was nonsignificant after controlling for vocabulary.

The risk conferred by low SES is not induced by poverty, but by poverty of the child's environment, which translates into lower quantity and quality of the language addressed to the child. Factors such as high lexical diversity, syntactic complexity, and the frequency of decontextualized language use are all features of good quality input that facilitate the acquisition process, and they tend to be less well represented in parents from lower SES backgrounds (Hoff, 2006; Huttenlocher, Waterfall, Vasilyeva, & Vevea, 2010; Rowe, 2012).

This study will probe the impact of SES on language proficiency in two ways. Assuming SES is a proxy for the richness of the language environment, we will investigate how it interacts with the amount of language exposure experienced by the child. Also, given that both SES and cognitive abilities are predictive of performance in language proficiency tasks, and that SES predicts many aspects of cognitive development, we will investigate the extent to which SES predicts language proficiency *over and above* the effect of cognitive predictors.

BILINGUALISM PROFILES

It is not uncommon to be faced with the question of "how bilingual" a child is. Parents of bilingual children who understand the home language but answer back in the societal language ask themselves if their child is "truly bilingual." Teachers wonder which of the bilingual children in their care may need additional support with the school language. Researchers seeking to assess the impact of bilingualism on other aspects of a child's development need to consider whether the children in their study are "bilingual enough" for an effect to be detectable—or may have as a primary question what the relevant threshold is.

Two questions are of particular interest. The first one relates to functional thresholds within the bilingualism continuum. The second one relates to language dominance more generally.

Functional Thresholds

From what amount of language experience in a "second" language should a child be considered bilingual? In their study of vocabulary development in bilingual children, Pearson, Fernández, Lewedeg, and Oller (1997) report that children with less than 20% exposure to one of their languages tended "not to produce utterances in that language willingly or spontaneously" (p. 56). While they were very careful to note that children below this exposure threshold might "become bilingual" nonetheless (and that some children with more exposure might not), their recommendation that the children in their sample "whose exposure was less balanced than 75:25 not be considered for future

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

bilingual studies" (ibid.) has been used as an exclusion criterion in many studies since (even if the use of that language was not under investigation, as in e.g., Laloi, de Jong, & Baker, 2017). The assumption is that, below a certain critical threshold of experience in more than one language, a child should be considered *functionally monolingual* (Bedore et al., 2012). What this means in practice is not clear. A functionally monolingual child may have a good level of comprehension in her "weak" language. Furthermore, while cross-linguistic interference during language comprehension (e.g., semantic priming) has been demonstrated in bilingual children as young as 30 months of age with at least 75:25 exposure balance (Jardak & Byers-Heinlein, 2018), we do not yet know from what amount of language exposure the effect of cross-linguistic interference is detectable.

A related question is that of how much experience a child needs to have of a language to be likely to perform within the monolingual range of their age group. Many rightly object to the evaluation of bilingual children's performance in terms of deficiency in relation to monolingual norms, as language processing has been shown to be fundamentally *different* in monolinguals and bilinguals (Cook, 1994; Grosjean, 1989; Romaine, 1989). However, assessing whether a bilingual child's performance falls within the range of typically developing monolinguals (which features a considerable amount of variation) remains useful for education purposes, as bilingual children's academic achievement is assessed only in the SL. For instance, poor academic performance might be due to a lag in language proficiency rather than a lack in academic abilities. In turn, a lag in language proficiency will in most cases be due to lower levels of exposure to the SL. As the academic demands are the same for bilingual and monolingual children, it is important to be able to take into account the likelihood of a disadvantage induced by reduced exposure to the SL.

Cattani et al. (2014) provide evidence suggesting that typically developing 2;6 year olds acquiring English and an additional language from birth will perform equivalently to their monolingual peers in terms of receptive and productive vocabulary if they are exposed to English 60% of the time.

What the threshold is for sequential bilinguals (who start out with a monolingual period) remains unknown. Also unknown is the threshold of language experience required for a child to perform within the monolingual range in other aspects of language proficiency. Our study will investigate these questions in 5 to 7 year olds, with respect to a range of proficiency measures.

Language Dominance

Variation in bilinguals' language experience can result in an imbalance between their two languages. Children with superior proficiency in one language have been argued to be dominant in that language (following e.g., Genesee, Nicoladis, & Paradis, 1995), in contrast with "balanced" bilinguals, who are expected to enjoy a similar level of proficiency in both languages. The notion of language dominance has prompted much controversy in the literature. Changes in language dominance over the life span (De Houwer, 2011) make it difficult to pin it down, and the distributed nature of bilingualism (where each language tends to be associated with different contexts—e.g., school vs. home) makes it difficult to assess language dominance across the board (Grosjean, 2016). There is also little agreement regarding how language dominance should be evaluated in the first place (see Silva-Corvalán & Treffers-Daller, 2015 for in-depth discussion).

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

However, there is a general consensus that language dominance is a useful construct, and that it can be estimated based on proficiency (Montrul, 2015) or based on language experience (Unsworth, 2015). A number of studies have also shown that language dominance predicts the directionality of cross-language influence in bilinguals: the dominant language has repeatedly been found to influence the nondominant one (see e.g., Bernardini & Schlyter, 2004; Kupisch, 2007; Lanza, 2004).

While generally derived from continuous measures of language experience or proficiency, language dominance tends to be conceptualized in terms of discrete categories based on critical thresholds (e.g., Bedore et al., 2012). Yet, as argued by Treffers-Daller (2015, p. 253), language dominance should be regarded as a continuum, given that bilingualism is not categorical.

Our study will compare cumulative language exposure and language dominance as measures to estimate whether a bilingual a child in monolingual education can be expected to perform within the monolingual range in terms of proficiency in the SL.

Summary

This review of the literature has brought to light a number of gaps that our study intends to address. These include (a) a systematic comparison of alternative measures of language experience (as predictor of a range of aspects of language proficiency) in bilinguals; (b) a better understanding of the impact of SES, in relation with the amount of language exposure experienced by the child, and, in relation with cognitive determinants of language proficiency; and (c) the objective identification of the amount of language exposure required for a bilingual child educated in monolingual settings to be expected to perform within the monolingual range in terms of language proficiency.

AIM AND QUESTIONS

Our overarching aim is to exploit advanced statistical methods to address the following three broad research questions, related to environmental predictors of language proficiency in bilingual children :

1. Does a gradient measure of bilingual language experience (combining exposure and use) and a gradient measure of SES (as a proxy for the richness of the child's language environment) significantly predict the following aspects of school language proficiency: comprehension and production of complex sentences, lexical semantics, and discourse-semantics?
2. How does a combined measure of cumulative language experience (combining exposure and use over time) compare with simpler measures (i.e., cumulative exposure or cumulative use, current exposure or current use), as predictors of school language proficiency?
3. What is the critical amount of school language experience required for bilingual children to perform within the monolingual range? In other words: At what point can one expect the gap between bilinguals and monolinguals to be "closed" in the school language?

We focus on children educated monolingually in the societal language (English), between the ages of 5 and 7 (i.e., in the first two years of formal education in England), with a broad range of exposure to a different language at home and diverse socioeconomic backgrounds. Gender, age, and cognitive factors (i.e., memory, cognitive

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

flexibility) will be included in the analyses as covariates, as they can be expected to have an impact but are not central to our research questions.

METHODOLOGY

PARTICIPANTS

We aimed to study a representative sample of the bilingual population in a geographical area that is characterized by a high degree of linguistic and cultural heterogeneity. We recruited 174 children (including 87 monolinguals) between the ages of 5 and 7 from schools throughout the North of England. Ethical approval was obtained from the University of Leeds (Ref. PVAR 12-007), and parental consent was obtained prior to data collection.

The SL was exclusively English for all the children. The bilingual children were also exposed to another language (henceforth the *home language*, or HL) in varying degrees (see the following text). There was a total of 28 HLs in our sample:[1] Arabic (9%); Bengali, Cantonese, Catalan, Dutch, Farsi, and French (8%); Greek, Hindi, Italian, Kurdish, Mandarin, Marathi, Mirpuri, Nepalese, Pashto, Polish, Portuguese, and Punjabi (21%); Shona, Somali, and Spanish (6%); and Swedish, Tamil, Telugu, Thai, Tigrinya, and Urdu (17%). Bilingual and monolingual children were recruited from the same schools for maximum comparability. None of the children were excluded from the study.

Table 1 summarizes the distribution of the two groups in gender and age. For ease of presentation, children with any amount of exposure to a language other than English are referred to as "bilinguals"; children who had no exposure to a language other than English are referred to as "monolinguals." All children were reported by the school to be developing typically and did not have any known hearing deficit.

Socioeconomic Profiles

The schools targeted were in areas of varying degrees of affluence, so as to recruit children from as broad as possible a socioeconomic spectrum. The SES of the children's families was estimated on the basis of information gathered through a parental questionnaire (see Supplementary Materials). Two measures were obtained: one for parental level of education and one for parental level of current occupation. The highest level was chosen in each case (on the assumption that the status of the household was determined by the best educated parent and the highest occupation among the parents).

TABLE 1. Participant distribution in gender and age (in years;months)

	Gender	Min.	Max	Mean	St. Dev.
Bilinguals	F (n = 44)	5;1	6;9	5;10	0;5
(n = 87)	M (n = 43)	5;1	7;0	5;10	0;6
Monolinguals	F (n = 52)	5;0	7;0	6;0	0;7
(n = 87)	M (n = 35)	5;0	7;0	6;0	0;7

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

(1) a. None
b. Primary school
c. Secondary school
d. Further education
e. University

The occupational data was scored using the reduced method of the National Statistics Socio-economic Classification (simplified NS-SEC, which is based on the Goldthorpe Scheme of sociological classification [Goldthorpe, 1980]).[2] The score obtained was reversed for ease of interpretability (i.e., a positive correlation with educational level).

The two measures are significantly associated (χ^2 (4, N = 174) = 83.57, $p < .0001$), as shown in Figure 1. There is a weak but significant negative correlation between the cumulative amount of HL exposure and SES (based on the occupational classification): $r = -.24$, $p = .0014$.

Language Experience

Estimates of the amount of exposure and use for each child in English and in their HL were calculated on the basis of information gathered using parental questionnaires (see Online Supplement).

The information sheet, consent form, and questionnaires were translated into the three most common HL with a high risk of low English proficiency in the parents (Bengali, Punjabi, Urdu). Because some of the parents speaking these languages turned out to be illiterate, help was offered from the research assistants, the teachers, or other employees at the schools, or in some cases from other parents who were able to translate.

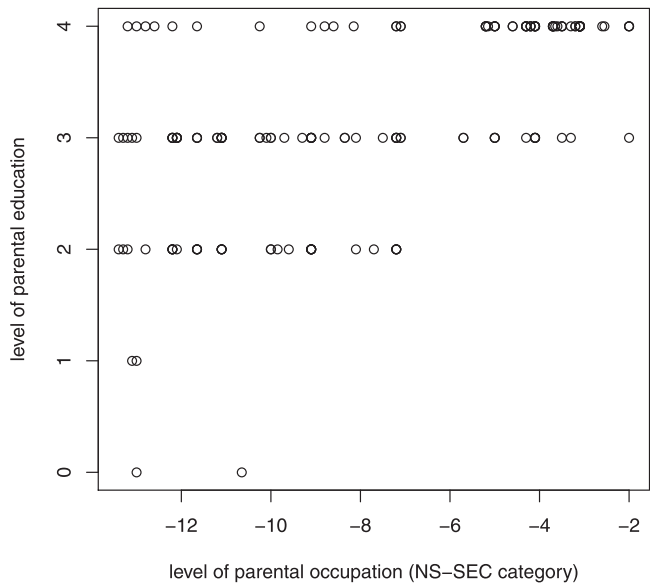


FIGURE 1. SES by education and occupation.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

The language questionnaire asked parents about their own proficiency in English and any other languages,[3] and their child's exposure and use of English and their other language in a range of contexts inside and outside the home. It was based on a simplified version of the BiLEC (Unsworth, 2013b), to make them understandable by parents with limited levels of literacy, without one-to-one support for filling the questionnaire (as this was not feasible for our sample size, given our means and time frame). Parents were asked to list their child's interlocutors at various periods during weekdays and during weekends in school term, as well as during holidays. The periods captured all the child's waking hours. School hours corresponded to a single block and the rest of school days was broken down by hours. Weekends and holidays were broken down in time slots of two hours. Type of interlocutor was constrained as a choice between mother, father, siblings, school, or other. For each interlocutor (except "school"), we asked which language was used to address the child, and which language the child used with that interlocutor.[4] The language of interaction had to be specified as either *English*, *home language*, *other home language* (i.e., other than the main one identified at the outset of the questionnaire), *2 home languages*, or *both English and the home language*. For each interlocutor (including the child), the proportion of interaction in English or the HL had to be specified (as *always*, *usually*, *half of the time*, *rarely*, or *never*, which we converted into a 5-point scale ranging from 100% to 0%). For mother, father, and any other significant carer in the home, their proficiency in English was rated by the parent filling the questionnaire (as speaking the language *very well*, *quite well*, *not well*, or *not at all*). Parents also reported the age of onset of exposure to English, the context of that first exposure (*home*, *play group*, *nursery*, *school*, or *other*) and whether their child was born in the United Kingdom. The total number of weeks spent in the HL country was also reported. Finally, parents were asked to indicate the types of activities their children engaged in, in each language (*reading with an adult*, *using a computer*, *watching television*, *sports/club*, *playing with friends*).

Current measures of language exposure and use were calculated as follows. The total number of hours of interaction with the child per year was calculated for each of the child's interlocutors. The proportion of total exposure received by the child from each interlocutor in the HL was calculated as the total number of hours of interaction with that interlocutor, multiplied by the proportion of the time the HL was used with that interlocutor. By adding these proportions from all the child's interlocutors and dividing the sum by the total number of hours of interaction, we obtained a measure of current exposure to the HL for that child, expressed as a proportion of their total interaction time (assumed to equate to waking hours). Current exposure to English was calculated in the same way. The sum of current exposure to the HL and in English came to 100%. We repeated the same procedure for language use.

Cumulative measures of exposure and use of the HL were calculated as the sum of (a) the number of months of monolingual exposure (i.e., prior to onset of exposure to English—this amounted to 0 for simultaneous bilingual children and for monolingual children) and (b) the number of months of bilingual exposure multiplied by the proportion of current exposure to (or use of) the HL. The cumulative measures thus correspond to the total number of months equivalent to full-time exposure to the HL.

The cumulative measures do not make a distinction between stages of linguistic development. One might object that, during the earliest months of life, the amount of

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

exposure to each language might be of little consequence, given that the child is not yet able to produce use of a particular language. However, it has recently emerged that the first year of life constitutes a critical period for the development of syntax (Friedmann & Rusou, 2015). We therefore take the default assumption to be that exposure to a language has an impact from the earliest stages of exposure.

A limitation of our estimation methods is that the cumulative measures cannot take into account variability over time, except that induced by age of onset of exposure to the societal/school language.

Our sample ranges across an evenly distributed continuum of bilingual language experience. At the lower end, some children had a very limited experience in a language other than English. At the higher end, some children were late bilinguals, having only experienced the HL until their first significant exposure to English at primary school.

As shown in Figure 2, in the bilinguals, current exposure to the HL ranged from 9% to 89% of estimated waking hours (left panel) while cumulative exposure to the HL ranged from 6 to 69 full months equivalent (right panel). At the time of the study, 31 (36%) children used their HL to the same extent as they were exposed to it (as shown by the top diagonal line in the plot). The rest of the children were relatively more "passive" in their HL, with nine children totally "passive" over their lifetime (i.e., always answering in English when addressed in the SL).

It is worth noting that simultaneous bilinguals, who were exposed to the two languages from birth, are not necessarily less "passive" in the HL: out of the 41 simultaneous bilinguals in our sample, only 11 used their HL to the same extent that they were exposed to it at the time of the study. Furthermore, even if a child started out as monolingual in the HL, this is no guarantee that they will remain "active" in that language when they start school in English. For instance, out of the 18 children who had had very little exposure to English before the age of 3, 8 of them had become relatively "passive" in the HL at the time of the study.

Children who are bilingual from birth did not necessarily come from "one-parent one-language" households (see Table 2): only 51% (21/41) of simultaneous bilinguals did so in this study. English exposure started outside the home (in daycare) for 17% (7/41) of

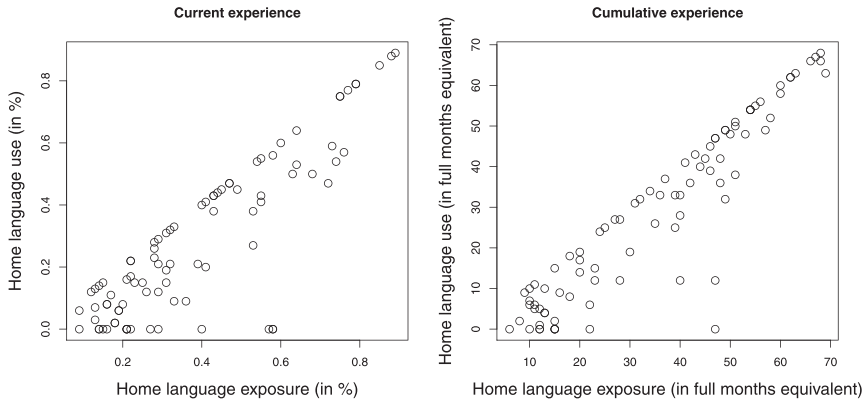


FIGURE 2. Bilingual children's current and cumulative experience in their HL.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 2. Distribution of children according to onset of bilingual exposure (in years), by home-language policy categories

	0	1	2	3	4	5
HL used with all	13	7	5	8	1	0
HL used with parents only	7	3	6	8	4	1
One-parent one-language	21	0	1	2	0	0

the simultaneous bilinguals. Given this variability, neither the language policy of the household nor the age of onset of bilingual exposure are sufficient to give the full picture of children's patterns of language exposure over time.

PROFICIENCY MEASURES

Several measures of English-language proficiency were collected from a battery of tests aiming to tap into a broad range of aspects of language competence. The tests included (a) a SRep task (the short version of the LITMUS Sentence Repetition test—Marinis & Armon-Lotem, 2015; Marinis, Chiat, Armon-Lotem, Gibbons, & Gipps, 2010), (b) the sentence structure subtest of the Clinical Evaluation of Language Fundamentals (CELF-4-UK—Semel, Wiig, & Secord, 2006), and (c) four subtests of the Diagnostic Evaluation of Language Variation: the verb and preposition contrasts, real verb mapping, novel verb mapping, and articles subtests (DELV—Seymour, Roeper, & de Villiers, 2005). The aspects of proficiency probed by each test are outlined in the following text.

The SRep task was designed specifically for bilingual populations.[5] It taps into language processing at all levels of representation (phonological, morpho-syntactic, and semantic) in both comprehension and production, as successful comprehension is necessary for correct repetition (Marinis & Armon-Lotem, 2015). We used the short version, which comprises 30 sentences with three levels of difficulty described in Table 3.

The Sentence Structure component of the CELF aims to "measure the acquisition of grammatical (structural) rules at the sentence level" (CELF manual, p. 88). In each trial of this comprehension test, the child is presented with four pictures and asked to point at the one that matches the prompt sentence. The test is designed for 5- to 8-year-old children.

The DELV is a dialect-neutral assessment for 4 to 9 year olds, aiming to limit the effects of language exposure differences in bicultural populations. We administered the following tasks to evaluate children's semantic competence, and in particular lexical semantics, argument structure, and discourse semantics:

(2) a. Verb contrast and preposition contrast
  b. Real verb and novel verb mapping
  c. Articles

The contrast tasks tap into the organization of the child's lexicon into contrastive words and levels of meaning. These tasks focus on action and location words, which

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 3. Sentence structures by level of difficulty in the SRep test

Level	Structure	Example
1	Declarative with one auxiliary	The boy must sweep the floor in the kitchen.
	Short actional passive	The children were taken to the office.
	Object wh-question with *what/who*	What did the princess buy last month?
2	Declarative with two auxiliaries	The kitten could have hit the ball down the stairs.
	Long actional passive	She was seen by the doctor in the morning.
	Object wh-question with *which*	Which picture did he paint at home yesterday?
	Temporal clause	He will feed the cow before he waters the plants.
3	Conditional clause	If the kids behave we will go into the garden.
	Object relative clause	The bee that the man swallowed had hurt him.

relate to common aspects of children's experience across cultural/social groups. They require the child to provide a word that contrasts appropriately with the one negated in a prompt sentence, based on the information provided by the accompanying picture. For instance, in the preposition contrast task, the child might be presented with a picture showing two children pulling angrily at the same train and with the prompt in (3). Adequate answers are illustrated in (a), and inadequate ones in (b).

(3) They're not sharing the train, they're…
 a. E.g., fighting over it, arguing over it (ADEQUATE RESPONSES)
 b. E.g., fighting, breaking it (INADEQUATE RESPONSES)

The preposition contrast task comprises 6 items, and the verb contrast task comprises 10 items.

In the mapping tasks (2-b), the child is presented with a series of pictures representing an event, and a second series of side pictures representing the participants in the event (see Figure 3). The examiner describes the event (as in [4]) while pointing at the pictures in a continuous, flowing motion, and then asks the child a series of questions (4-a)–(4-d).[6] The child answers by choosing one of the side pictures.

(4) The boy is pouring the juice.
 a. Which one was the pourer? (boy)
 b. Which one got poured? (pitcher of juice)
 c. Which one did he pour the juice into? (glass)
 d. Which one was pourable? (pitcher of juice)

The novel verb mapping task follows the same protocol, but with a nonce verb. The aim of this task is to assess children's abilities to map new meanings from the linguistic context onto verb frames.

(5) The man is lelling the clown.
 a. Which one was the leller? (man)
 b. Which ones was lellable? (clown)

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

c. Which one was the clown holding on to? (DISTRACTOR)
d. Which one got lelled? (clown)
e. Which one was lelling? (man)

The real verb and the novel verb mapping tasks comprise three verbs/scenarios each, about which the child is asked 10 and 15 questions, respectively.

Finally, the articles test of the DELV (2-c) aims to assess children's understanding (and implementation) of the discourse-semantic differences in the English article system. Children were asked a series of 30 questions (each preceded by a minimal verbal context) requiring a noun phrase as their answer. Based on their assessment of what the interlocutor knows, the child has to choose between a definite (*the*) or an indefinite (*a*) article for that noun phrase (Schafer & de Villiers, 2000). This depends on how the object is defined and what is known or assumed about the object by the listener. The test was preceded by practice trials until it was clear that the child understood the task.

(6) George bought a shirt and a bag. He's wearing one of them. Guess which.
a. The + appropriate noun (e.g., the bag or the shirt) (CORRECT)
b. A + appropriate noun (INCORRECT)
c. Bare noun; inappropriate noun; no response (INCORRECT)

COGNITIVE MEASURES

Language proficiency tests require children to process and remember language prompts but also, when based on pictorial stimuli, to compute inferences (e.g., to choose the correct picture from a set). We therefore collected several cognitive measures to use as covariates in the analyses.

Measures of STM and WM were obtained from the Digit Span tasks (Wechsler Intelligence Scale for Children III—Wechsler, 1991). The use of numerical memoranda has been shown to be relatively independent of test language and cultural status (Engel de Abreu, Baldassi, Puglisi, & Befi-Lopes, 2013). In those tasks, the examiner verbally presents digits that the child has to repeat in the same order (in the Forward Digit Recall task) or in reversed order (in the Backward Digit Recall task). The number of digits increases by one until the child consecutively fails two trials of the same digit span length. There is a maximum of four trials per digit span length.

The Forward Digit Span measure was used as a proxy for children's STM capacity, a key component in Baddeley's (2000) Multicomponent Working Memory Model, which has been argued to represent a constraint in language processing (Boyle, Lindell, & Kidd, 2013). The Backward Digit Span measure was used as a proxy for children's WM capacity, which has been shown to correlate with spoken sentence comprehension (Magimairaj & Montgomery, 2012; Montgomery et al., 2008). The raw results on the memory tests are summarized in Tables 4 and 5.[7]

To assess children's cognitive flexibility, we administered the Dimensional Change Card Sort task (henceforth DCCS). The protocol was as described in Zelazo (2006). The child was presented in each trial with a picture representing either a rabbit or a boat, that was either blue or red, and asked to place the card in one of two boxes according to a sorting rule (by shape or by color). The boxes were identified by either

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

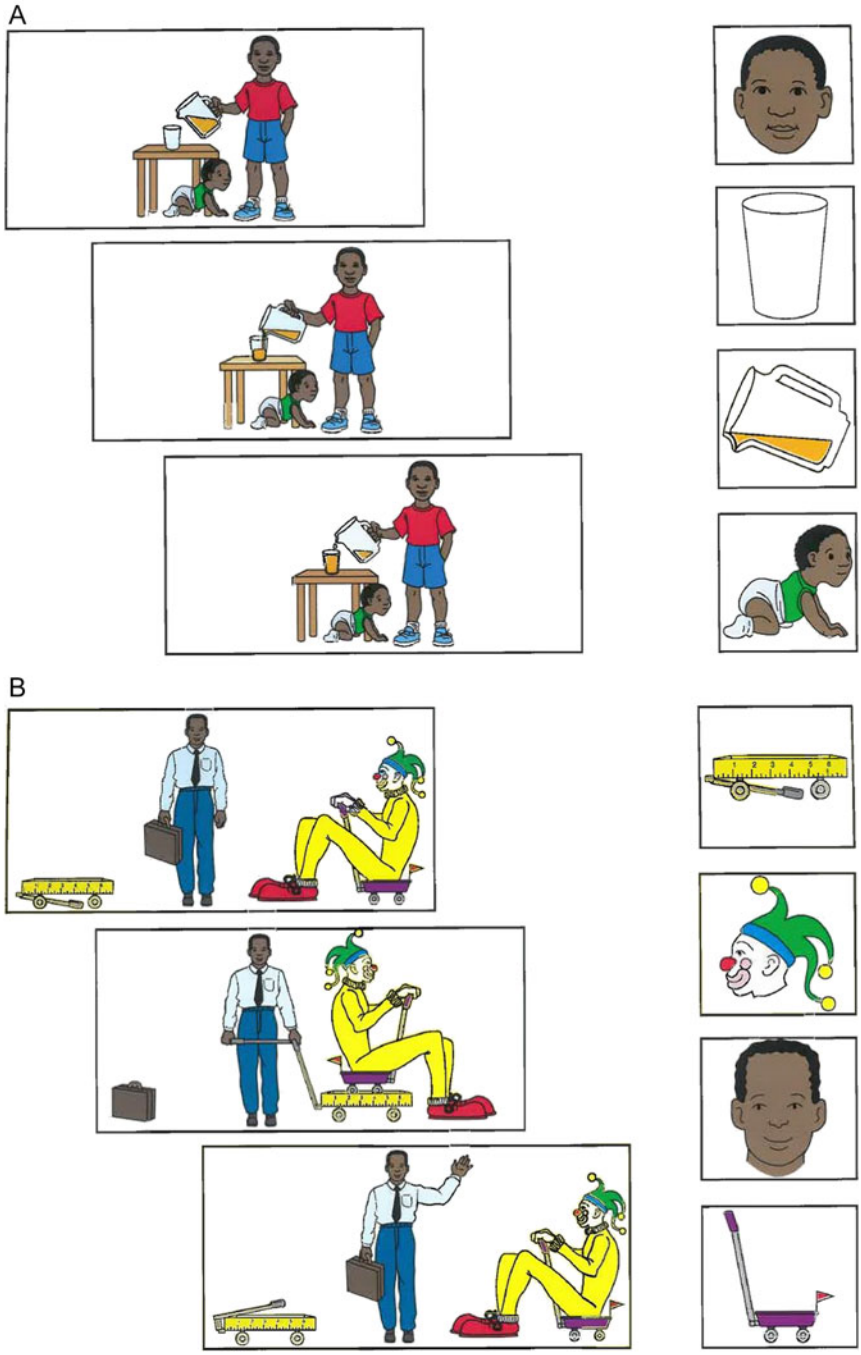


FIGURE 3. Real verb mapping (top) and novel verb mapping (bottom).

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 4. Forward Digit Recall (number of digits correctly recalled)

	3	4	5	6	7
Monolinguals	9	34	35	8	1
Bilinguals	13	34	37	2	1

a blue rabbit or a red boat. The first block trials (N = 6) required using the shape criterion, and the second block (N = 6) required using the color criterion. As the children in our sample were older than 5, we also administered a more advanced block of trials (N = 12) in which a star appeared on some of the cards. Cards without a star had to be sorted according to the shape dimension, and cards with a star had to be sorted according to the color dimension. The repetition of instructions on every trial (in all blocks) ruled out the possibility that difficulty could be attributed to hypothesis testing or memory of the relevant rules. The test session was preceded by a demonstration and two practice trials. The test trials were presented in two counterbalanced orders.

We adopted the method recommended by Zelazo (2006) to score the DCCS data. This is based on a pass-fail criterion on each trial block. Passing the first two blocks requires sorting at least five out of six cards correctly (on each block). Passing the third block requires sorting at least 9 out of 12 cards correctly. A child is assigned a score of 0 if they fail the preswitch block, a score of 1 if they pass the preswitch block only, a score of 2 if they also pass the postswitch block, and a score of 3 if they pass all three blocks. Little variability is observed, as shown in Table 6, as many children performed at ceiling. Only one (bilingual) child did not pass the first block.

DESCRIPTIVE RESULTS FOR THE PROFICIENCY TESTS

Results for each proficiency test are first reported descriptively, showing the range of performance of bilinguals compared with that of monolinguals. Unsurprisingly, as a group, the bilinguals' performance is below that of monolinguals in each test.

SENTENCE REPETITION

The SRep test can be analyzed in different ways, depending on the criterion against which the children's utterances are evaluated (Marinis & Armon-Lotem, 2015):

TABLE 5. Backward Digit Recall (number of digits correctly reversed)

	0	2	3	4
Monolinguals	0	32	52	3
Bilinguals	3	34	46	4

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 6. Distribution of overall DCCS scores (based on block pass-fail)

DCCS score	0	1	2	3
Monolinguals	0	2	26	59
	0	1%	15%	34%
Bilinguals	1	12	39	35
	1%	7%	22%	20%

(7) a. Exact repetition of the target sentence (Full Accuracy)
b. Grammatical production of the targeted structure (Target Accuracy)
c. Grammaticality of the sentence produced (whether or not it fully matches the target) (Grammaticality)
d. Types of mismatches (i.e., omissions, substitutions or additions) compared with the target sentence) in the following categories:
(i) Lexical
(ii) Inflectional
(iii) Functional

The distribution of raw scores according to the first three types of analysis is plotted in Figure 4. Utterances containing unintelligible material were excluded (total: 208, i.e., 4% of the data).

SENTENCE COMPREHENSION

Although not designed for bilingual populations, the CELF is frequently used to evaluate the proficiency of bilingual children (e.g., Barac & Bialystok, 2012; Paradis, Crago, Genesee, & Rice, 2003).

The distribution of mean scores in the CELF Sentence Structure task is shown in Figure 5.

LEXICAL-SEMANTICS

We analyze the lexical-semantic tests of the DELV separately from the discourse-semantic test, as they tap into different domains of language competence: the lexical-semantic tests tap into the organization of the lexicon, and the discourse-semantic tests taps into the child's ability to express contrasts in information status through the appropriate use of determiners.

The distribution of raw scores on the lexical-semantic tests is plotted in Figure 6.

DISCOURSE-SEMANTICS

The distribution of the mean scores for the articles test of the DELV are plotted in Figure 7.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

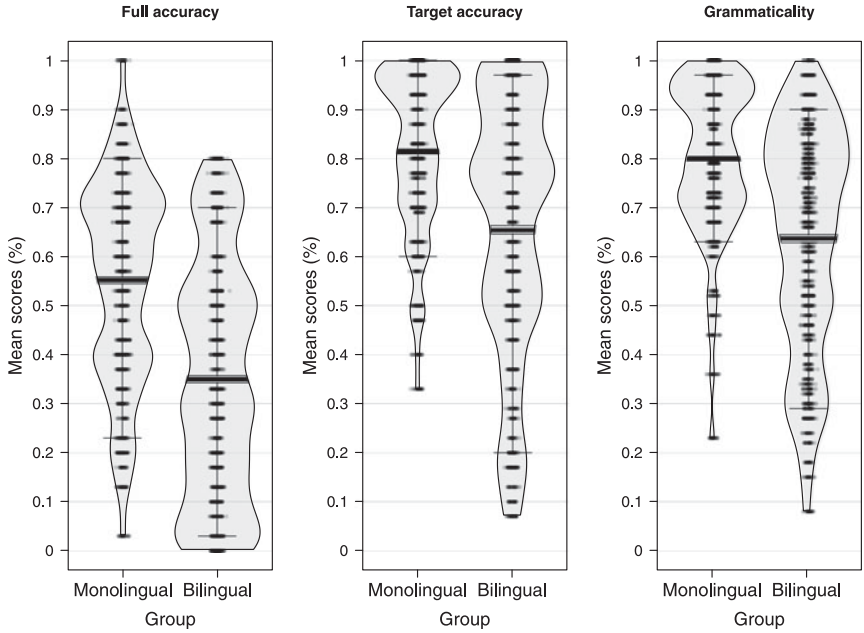


FIGURE 4. Pirate plots showing the distribution of the raw SRep test scores (according to three different scoring methods). Each plot shows group mean (thick line), confidence intervals (lighter area around the mean), 10% and 90% quantiles (whiskers), and individual scores.

METHODS OF ANALYSIS

To model the relationship between English proficiency (as indexed by the response variables from each test) and the potential predictors of proficiency we aimed to investigate, we performed linear regression analyses. For each predictor, the model effectively fits a single line through the scatterplot of individual scores, showing how the mean response changes with the predictor variable. The model can also account for the interaction of two (or more) predictor variables. An advantage of regression analyses is that they can model variance due to continuous as well as categorical predictors.

Linear models assume that the observations (of the response variable) are independent. However, in each proficiency test, multiple observations were collected from each individual, resulting in a nested data structure (rather than truly independent observations). To account for this, and to control for sampling effects (due for instance to lexical properties of the test items), we included Participant and Test Item as random effects in each model, alongside the fixed effects (the predictor variables). The resulting analysis is a *mixed-effect* model, as it combines random and fixed effects in a single analysis (Baayen, 2008).

We fitted linear mixed models to the data from each proficiency test, using the "lme4" package (version 1.1.21) in R (version 3.6.0). The models were built by adding predictors incrementally, starting from a null hypothesis model including only random effects for Participant and Item. A predictor was retained only if it improved the fit of the model, yielding a significant reduction in Akaike Information Criterion (AIC)[8] and a significant R-squared value for the model.[9] Random slopes lead to model nonconvergence and were

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

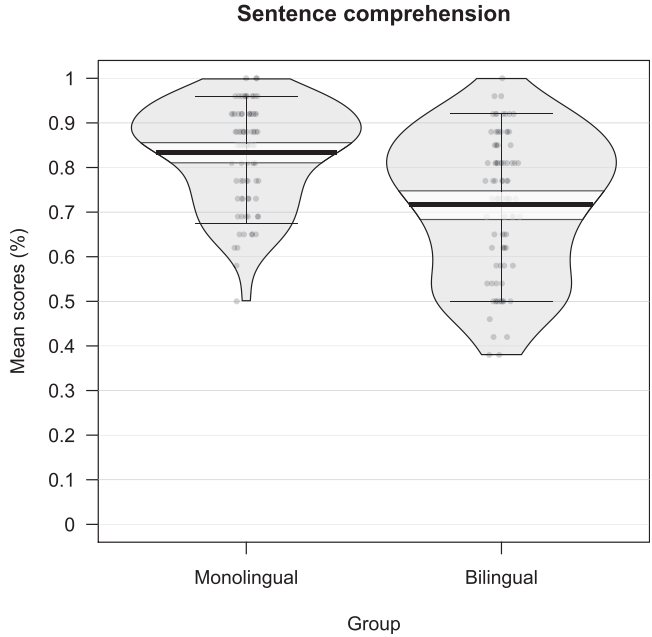


FIGURE 5. Pirate plot showing individual mean CELF scores, showing group mean (thick line), confidence intervals (lighter area around the mean), and 10% and 90% quantiles (whiskers).

therefore not included in the final models. In each case the final model was checked against the null hypothesis model, and its residuals (i.e., the observations not accounted for by the model) were checked for normality of distribution, as a further test of the quality of the model.

In all analyses, we tested whether the following variables were significant predictors: age, HL experience, SES, STM, WM, and gender (treating the latter three as covariates). All continuous variables were standardized so that the impact of the predictors can be compared within each model.[10]

For each analysis, we report the optimal model in a table. The statistics for non-significant factors will be given in the text where relevant. In the first instance, we present models identifying the predictors of proficiency in bilingual children only. Subsequently, we include monolingual children in the models and identify the threshold of HL experience below which the performance of "bilinguals" does not differ significantly from that of monolinguals in the language of schooling.

QUESTION 1: PREDICTING LANGUAGE PROFICIENCY FROM BILINGUAL CHILDREN'S LANGUAGE EXPERIENCE AND SES

Before reporting on the analyses carried out to address this question, we explain how language experience was apprehended as a variable.

We saw that children's experience of their HL changed over time, and that many tended to become more "passive" in their usage (even if they spoke only the HL before

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

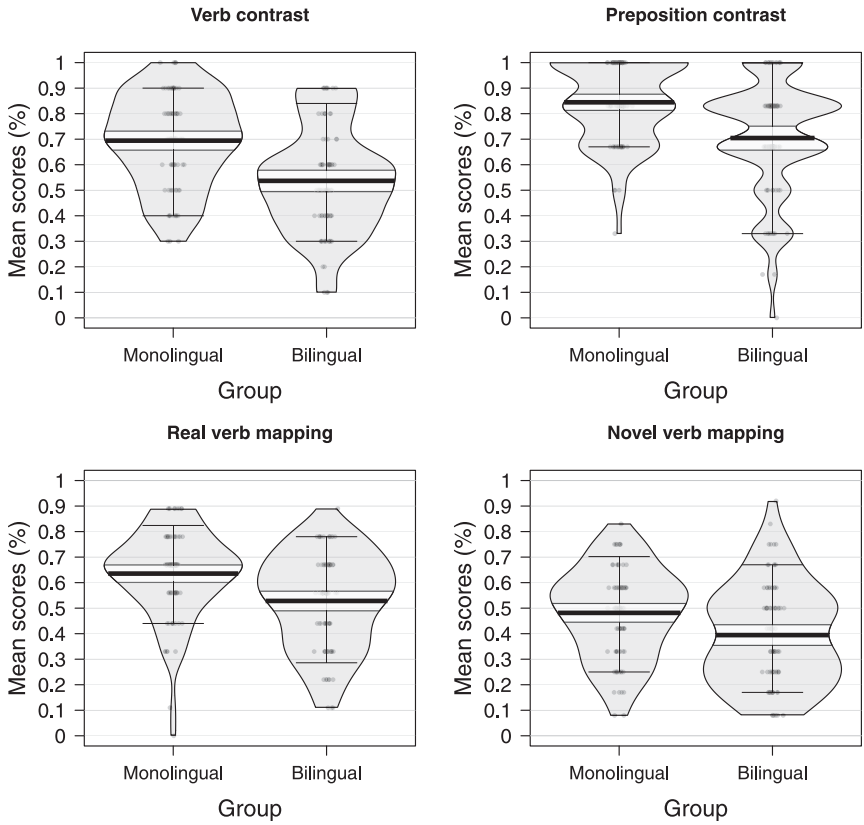


FIGURE 6. Distribution of the scores on the DELV subtests, in bilingual and monolingual children (showing mean, 95% CI, and full range).

the age of 3 or 4). Cumulative measures of language experience are therefore likely to be more accurate than current measures, as they can reflect these changes over time. Consequently, we based the first set of analyses on cumulative measures (and then subsequently compared them with analyses based on current measures, to see which fared better).

As all the children in our sample were exposed to two languages only, the amount of experience obtained in one language was inversely correlated with that obtained in the other language (e.g., cumulative exposure to HL vs. English: $r = -.955$, $p < .0001$). In terms of statistical modeling, it should therefore not make a difference whether exposure to one or the other language is used as a predictor. For practical reasons, we chose to base our language experience measure on the HL measure, so as to make it more easily interpretable: monolingual children would naturally get a bilingual score of 0 whatever their age, as they had no exposure to a HL other than English; for bilingual children, the higher the score, the greater the child's experience in their HL (and the lesser their experience in English). Another advantage of basing our combined measure on the HL is that, given the substantial variation in HL experience in the group under investigation,

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

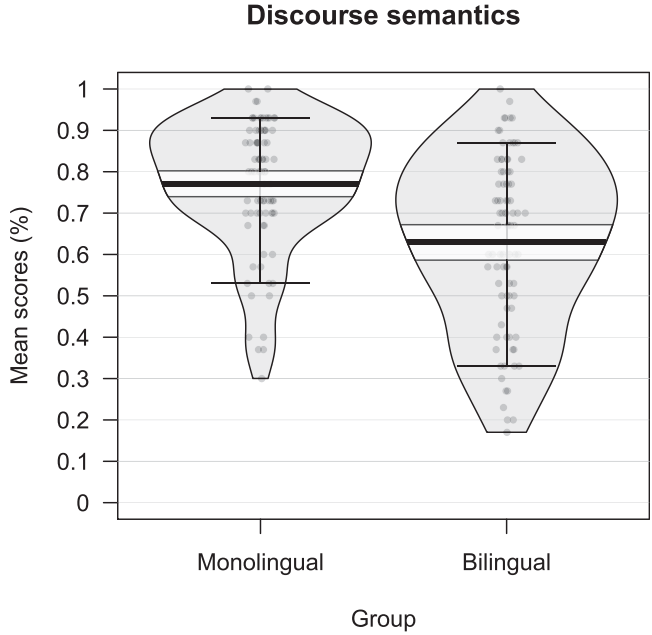


FIGURE 7. Pirate plots for individual mean scores in the articles test, showing mean (thick line), confidence intervals (lighter area around the mean), and 10% and 90% quantiles (whiskers).

the resulting measure is not correlated with age. This makes it possible to tease apart the respective effect of language exposure and age within a single model.

To take "passivity" into account (on the assumption that a bilingual who is relatively passive in a language is likely to be more proficient in their other language), language use has to be taken into account in relation with language exposure. Regression analysis can deal with the effect of continuous predictor variables, but is not able to handle severe collinearity between predictors—as is the case with our language exposure and use measures.[11] We therefore combined the cumulative measures of HL exposure and use into a single measure, as explained below.

To address the first part of our first research question (i.e., whether a gradient measure of language experience [combining exposure and use] significantly predicts SL proficiency), we created a new language experience measure combining the cumulative estimates of language exposure and use. Because English is the language of schooling, we assumed it was actively used by all children. Greater variability was expected in the exposure to and use of the HL, which, for many children, was (or would become) the weaker language. We therefore used HL estimates to derive our combined language experience measure.

As will be shown in the following text, this combined measure aligns very closely to the amount of cumulative HL exposure, with a "correction" if the HL is productively used by the child to a lesser extent than they are exposed to it (i.e., if cumulative use is lower than cumulative exposure). It can therefore be conceived as a measure of cumulative HL exposure, adjusted for relative "passivity" in that language. We will refer to this combined measure as the Bilingualism Profile Index (BPI).

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

The BPI was derived by a standard measure of dimensionality reduction: a Principal Component Analysis (PCA) of the two scores.[12] The choice of this method to combine the two scores is justified as follows. The apparently simpler alternative, that is, linear combination of exposure and use measures, would require arbitrary manipulation, as a particular mathematical operation (e.g., multiplication, division) is required. The choice of operation should be justifiable on conceptual grounds, such that the resulting measure is clearly interpretable. Multiplying exposure and use or dividing one by the other would not result in scores that could be meaningfully interpreted. By contrast, dimension reduction (here by PCA) does not encounter these conceptual difficulties: the combination of the two measures is entirely determined by their distributional properties and exempt of (arbitrary) intervention from the researcher. The resulting score simply combines the information contained in the original two scores. The BPI can thus confidently be interpreted as a score for HL exposure "corrected" downward in the case of comparatively lower use of the HL (see the following text for illustrations).

The PCA of cumulative HL exposure and use yields two principal components, the first of which accounted for most (98%) of the variability in our dataset (due to the strength of the correlation between the two cumulative measures). The first component thus contains almost all the information from the two original measures. The BPI scores correspond to the reversed loadings of that first component.[13] The monolingual children obtain a BPI score of 0, as their score was 0 for both cumulative measures of HL experience. The bilinguals' BPI scores ranged from 4 to 96.

Figure 8 shows the correlation of the BPI with current and cumulative measures of exposure and use of the HL.[14] The dispersion from the linear relationship between the BPI and each of the cumulative measures shows that it is not reducible to any of these measures.[15]

The BPI is very strongly associated with the age of onset of exposure to English ($\chi^2 = 6876.01$ $p < .0001$), but simultaneous bilinguals do not necessarily have less HL experience than sequential bilinguals (who experienced an initial monolingual period in that language). This is shown in the left pane of Figure 9.

The BPI is also associated with the type of HL environment, which we operationalize as follows:

(8) Home language policy categories:
HS ("home split"): one-parent one-language
HOL ("home only, low"): both parents speak the home language; siblings tend to speak English
HOH ("home only, high"): the whole household speaks the home language

While Bilingualism Onset and Home Environment are significantly correlated or associated with the BPI score, neither of them is able to capture the variability in HL experience as precisely as the BPI. For instance, Figure 9 shows that there was a substantial amount of variability in the amount of HL experience obtained by children who were bilingual from birth.

The BPI has two methodological advantages. As a *single* measure capturing cumulative exposure and use, the BPI makes it possible to use the combined impact of these two highly correlated measures as a predictor in regression analyses.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

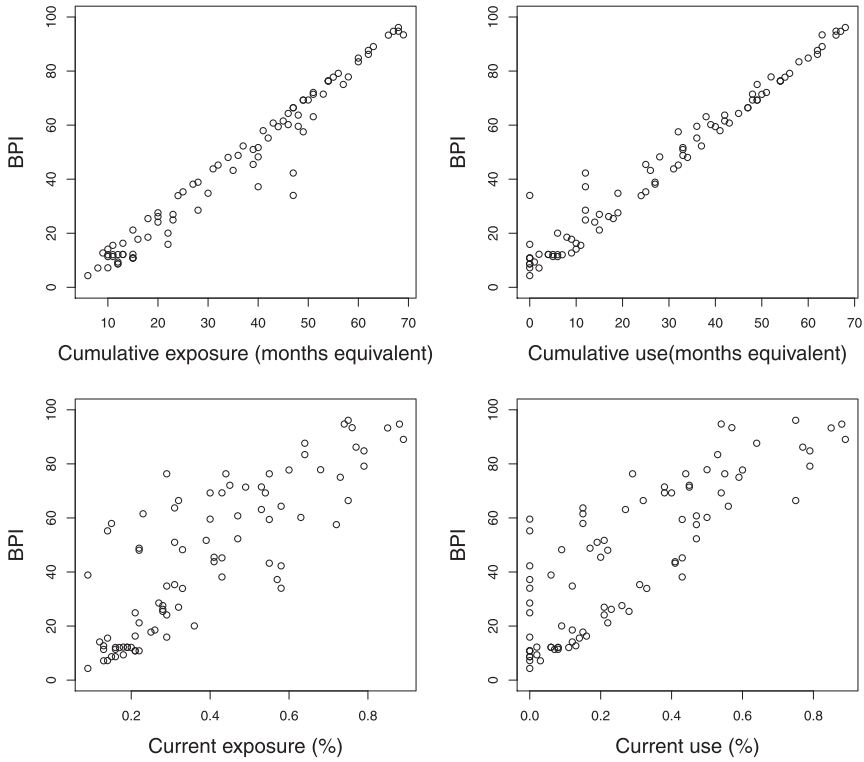


FIGURE 8. The BPI and its relationship with exposure and use measures in the HL. Each dot represents one bilingual child. Monolingual children would score 0 on both axes.

To answer to our first research question, we investigate the impact of the BPI and SES as predictors of children's proficiency in English (the SL) in the age group under consideration. The proficiency measures we consider in turn are SRep, sentence comprehension, lexical semantics, and discourse-semantics.

SENTENCE REPETITION

Using the methodology described in the preceding text, we fitted a mixed-effect model to bilingual children's accuracy data. The response variable was the correct repetition of the target structure on each it the SRep test items (see [7-b]). It indexes children's ability to produce structures of increasing grammatical complexity (listed in Table 3) independently of their ability to remember exactly all the words in the sentence.[16]

Table 7 summarizes the effect of the significant predictors of accuracy of the production of the target structure by the bilingual children (expressed in our model as the probability of reproducing that structure correctly), as per the optimal model. The estimates in the model summary report the likelihood of a correct response[17] for each predictor, compared with the likelihood observed at the intercept. As all the continuous

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

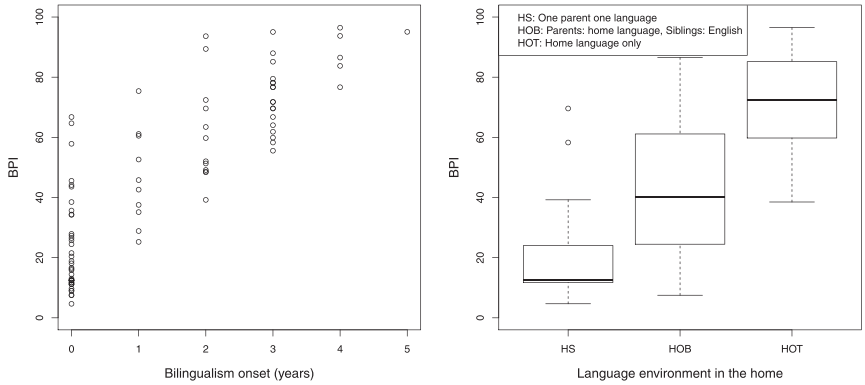


FIGURE 9. The relationship between the BPI and coarser estimates of bilingual experience: Age of Onset and Language Environment in the Home.

predictors were transformed into z-scores, the intercept is based on the combination of the mean values: the average score for SES, memory, age, and BPI characterizing the bilingual group. In terms of categorical variables, the intercept is calibrated for SRep Level 1 and Gender: Female.

As seen in the model summary, the BPI is the strongest child-related predictor ($\beta = -.7026$, $p = .0008$). Children's ability to reproduce the target structure is negatively correlated with the amount of experience in their HL: a higher BPI score decreases the probability of a correct response significantly. Children's performance was also predicted by their SES ($\beta = .5842$, $p = .002$) and their STM ($\beta = .5556$, $p = .006$). The effect of age (over and above that of STM) also approached significance ($\beta = .4082$, $p = .069$). Structural complexity (SRep Level) had a significant impact on performance. The more complex the targeted structure, the poorer the performance: even Level 2 yielded significantly poorer results than Level 1 ($\beta = -.9938$, $p = .041$).

There was no interaction between structural complexity level and children's amount of bilingual experience ($\beta = .13$, $p = .45$). WM had no significant effect $\beta = -.05$, $p = .77$). The random effects for participant and item had a variance of 2.55 and 1.12, respectively.

The same predictors were found to be significant in the other overall accuracy analyses of the bilingual children's performance in the SRep test (not reported here). The proportion of variance captured by STM compared with the other predictors is greater in the Full Accuracy analysis. In the Grammaticality analysis, the level of difficulty (SRep Level) is a lower-ranking predictor (which is to be expected as this analysis does not penalize avoidance strategies), with no significant difference between complexity levels 1 and 2. The relative ranking of predictors by decreasing order of importance in each model are listed in (9) (where SRep Level stands for the items' level of difficulty):

(9) a. Full Accuracy (7-a):
STM > SRep Level > SES > BPI
b. Grammaticality (7-c):
STM > SES > BPI > SRep Level

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 7. Fixed effects of the optimal Generalized Linear Mixed Model of accurate repetition of target structures in the SRep test by bilingual children. Random effects: participant, item

	Coefficient	Std. Error	Z	p
Intercept	3.1653	0.4683	6.7599	0.0000
SRep Level 2	−0.9938	0.4851	−2.0484	0.0405
SRep Level 3	−2.1150	0.5122	−4.1296	0.0000
BPI (scaled)	−0.7026	0.2103	−3.3405	0.0008
SES (scaled)	0.5842	0.1917	3.0469	0.0023
STM (scaled)	0.5556	0.2020	2.7510	0.0059
Age (scaled)	0.4082	0.2242	1.8207	0.0687
Gender: M	−0.3864	0.3701	−1.0439	0.2965

Error analysis was performed according to the domain affected (functional, inflectional, lexical). Errors are illustrated in the following text for the functional domain (10), the inflectional domain (11), and the lexical domain (12), respectively.

(10) Target: He will feed the cow before he waters the plants.
He will feed the cow <u>when</u> he waters the plants. (Functional error)

(11) Target: The children were taken to the office.
The children take<u>d</u> to the office. (Inflectional error)

(12) Target: The mum bakes the meal that the children are eating.
The mum <u>makes food</u> for the children. (Lexical errors)

Three accuracy scores were obtained for each sentence by dividing the number of errors of each type from the total number of relevant words (Marinis et al., 2010).

We tested the significance of the same predictors as in the preceding analyses. The full models are reported in the Online Supplement. The relative importance of predictors (and their statistical significance) differs across error types, as shown in (13) (nonsignificant predictors are omitted).

(13) Functional: SRep Level > BPI > STM > SES
Inflectional: SRep Level > STM
Lexical: SRep Level = STM > SES > BPI

Neither the BPI nor SES turn out to be significant predictors of inflectional errors. The relative importance of the BPI is highest in the model predicting functional errors. The relative importance of STM is highest in the model predicting lexical errors.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

SENTENCE COMPREHENSION

A linear regression analysis of the CELF Sentence Structure data reveals that bilingual children's scaled scores were *not* predicted by children's BPI ($\beta = -.18$, p = .58), nor their SES ($\beta = .51$, p = .09), or their STM ($\beta = .14$, p = .67). Instead, the strongest predictors were WM ($\beta = .58$, p = .03) and cognitive flexibility, ($\beta = 1.36$, p = .00001).

In De Cat (2019), I argue that the CELF Sentence Structure test does not probe children's ability to deal with linguistic aspects of structural complexity, as these are confounded with cognitive complexity in the task. The lack of significance of English exposure (as the counterpart of HL exposure) as a predictor is most likely due to that confound. It is unclear whether the CELF Sentence Structure scores can be interpreted as a reliable index of language proficiency. We will therefore not consider them any further in this article.

LEXICAL-SEMANTICS

Given the different distribution of scores across tests in the bilingual group (Welch's F (3, 190.795) = 31.0695, $p < .0001$), we did not use a composite score in the first instance. We carried out a MANOVA to identify the significant predictors of performance in the four DELV subtests (in one go) without inflating the Type 1 error rate.

Using Pillai's trace, a statistically significant MANOVA effect on the performance of bilingual children was observed for the BPI ($V = .2$, F(4, 77) = 4.7, $p = .002$), SES ($V = .18$, F(4, 77) = 4.23, $p = .004$), STM ($V = .18$, F(4, 77) = 4.36, $p = .003$), and cognitive flexibility ($V = .23$, F(4, 77) = 5.77, $p = .0004$). Age approached significance ($V = .11$, F(4, 77) = 2.38, $p = .0586$), and WM was not significant ($V = .063$, F(4, 76) = 1.28, $p = .28$).

After the MANOVA had identified the significant predictors, linear regression models were fitted to predict the accuracy score on each of the DELV subtests. The BPI was a significant predictor of poorer performance in the Preposition Contrast task (t (82) = -2.63, p = .01) and the Real Verb Mapping task (t (82) = -2.39, p = .02), and approached significance in the Verb Contrast task (t (82) = -1.80, p = .07). It was not significant in the Novel Verb Mapping task (t (82) = -0.81, p = .42). SES significantly predicted performance in the Verb Contrast task (t (82) = 2.57, p = .01), and approached significance in the Real Verb Mapping task (t (82) = 1.96, p = .05). It was not significant in the Preposition Contrast task (t (82) = 1.47, p = .14) nor in the Novel Verb Mapping task (t (82) = 1.37, p = .17). STM significantly predicted performance in the Real Verb Mapping task (t (82) = 4.26, $p <$ 0.0001) and the Novel Verb Mapping task (t (82) = 3.32, p = .001). It approached significance in the Verb Contrast task (t (82) = 2.56, p = .1), and was not a significant predictor in the Preposition Contrast task (t (82) = 1.47, p = .14). This is summarized in Table 8.

If the four scores are combined into a composite score, the following factors are found to predict bilingual children's performance (in order of importance), as summarized in Table 9: STM, BPI, and WM. SES and Age are not significant.

DISCOURSE-SEMANTICS

Analysis was performed by fitting a linear regression model to the children's overall score. Performance was negatively correlated with the BPI ($\beta = -.0485$, p = .012). This was more than compensated by SES, which was a stronger predictor of performance in

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 8. Predictor significance in the DELV lexical-semantic tests (* indicates significance; . indicates a trend)

	BPI	SES	STM
Preposition contrast	*		
Verb contrast	.	*	.
Real verb mapping	*	.	*
Novel verb mapping		.	*

the bilingual children (β = .0549, p = .0074). The following factors proved nonsignificant predictors of the bilingual children's performance: age (β = .0306, p = .12), STM (β = .0125, p = .53), and WM (β = .029, p = .14). Table 10 reports the summary of the optimal model.

SUMMARY

In answer to our first research question, we have demonstrated that the BPI, as a gradient measure of language experience combining language exposure and use, is a significant predictor of several aspects of SL proficiency in the population under study. The BPI is the strongest child-related predictor in the SRep test (if scored based on the correct repetition of the target structure rather than full accuracy). It is also a strong predictor in the lexical semantic tests (the strongest predictor after cognitive flexibility) and in the discourse semantic test. Although the BPI indexes cumulative HL exposure and use in our models, the interpretation of its effect should not be that HL experience has a detrimental effect. The BPI was used as an indirect estimate of the amount of experience in the SL (as it is the inverse picture of HL experience in bilinguals). Therefore, our findings are in line with much previous research showing a positive association between amount of language input and use to performance on tests in the SL.

Furthermore, bilingual children's SES is a significant predictor of performance in the SRep task and in the Discourse-Semantics tasks. It was found not to be a robust predictor in the Lexical Semantics tasks.

TABLE 9. Coefficients of the optimal linear regression model fitted to bilingual children composite score in the DELV lexical-semantic tests

	Coefficient	Std. Error	t-value	p
Intercept	0.5249	0.0164	32.0553	0.0000
BPI (scaled)	–0.0349	0.0118	–2.9493	0.0042
Short-term memory (scaled)	0.0312	0.0131	2.3740	0.0200
Working memory (scaled)	0.0286	0.0125	2.2903	0.0246
Cognitive flexibility (scaled)	0.0566	0.0127	4.4560	0.0000
Gender: M	–0.0195	0.0235	–0.8272	0.4106

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 10. Coefficients of the optimal Linear Regression Model predicting bilingual children's accuracy score in the discourse semantics task

	Coefficient	Std. Error	Z	p
Intercept	0.6384	0.0255	24.9904	0.0000
BPI (scaled)	–0.0485	0.0189	–2.5628	0.0122
SES (scaled)	0.0549	0.0200	2.7476	0.0074
Cognitive flexibility (scaled)	0.0717	0.0197	3.6399	0.0005
Gender: M	–0.0170	0.0366	–0.4646	0.6434

Children's cognitive abilities were included as a covariates, and found to have an impact in all tasks (memory in most tests, and cognitive flexibility in those tests relying on inferencing abilities). Performance in the sentence comprehension task (the Sentence Structure task from the CELF) was correlated with cognitive predictors only, which we interpret as a problematic confound (see De Cat, 2019).

QUESTION 2: COMPARING ALTERNATIVE MEASURES OF LANGUAGE EXPERIENCE

Having demonstrated that a gradient, composite measure of language experience significantly predicts several aspects of bilinguals' proficiency in the SL, we now turn to our second research question, which seeks to identify the optimal way of modelling language experience as a predictor of SL proficiency. We break it down into the following two questions, which we address in turn:

1. How does a combined measure of language experience (combining exposure and use) compare with simpler measures (i.e., exposure *or* use), as predictors of school language proficiency? Are *cumulative* measures more accurate than current measures, in predicting school language proficiency?
2. What is the specific impact of language *use* as a predictor of school language proficiency?

QUESTION 2.1: COST-BENEFIT ANALYSIS: WHICH LANGUAGE EXPERIENCE MEASURE TO USE AS A PREDICTOR?

The BPI indexes the cumulative amount of exposure and use of the HL, and it is strongly correlated with alternative measures of language exposure (see Figure 8). Our rationale in choosing the BPI as predictor over these alternative measures was based on a number of considerations: (a) calibrating language experience on the HL makes it possible to include monolinguals in the continuum "naturally" (as they have had zero experience of a HL different to English); (b) encompassing language use (as well as exposure) yields a more precise measure, which in turn should increase the power of our analyses; and (c) focusing on cumulative experience further increases the accuracy of the measure, as it reflects the quantitative effect of sequential bilingualism.[18] It is, however, possible that the higher level of precision of the BPI does not afford a substantial benefit over a simpler

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

measure. In this section, we compare the BPI with simpler, alternative measures of language experience: cumulative exposure, cumulative use, current exposure, and current use.

To determine the optimal language experience measure predicting SL proficiency, we exploited the method described in Burnham and Anderson (2003): an information-theoretic approach to inform model selection and multimodel inference. Choosing between alternative statistical models requires estimating which of them best approximates the "true" process underlying the phenomenon under study. If several models compete closely, the "true" process should be inferred by combining the information from these models. In this approach, both processes are informed by the *Akaike weight* associated with statistical models.

The Akaike weight of a model is based on the AIC, which was defined previously as an indicator of the trade-off between the accuracy and the complexity of a given model (i.e., a measure of the relative goodness of fit of the model to reality). When considering two nested alternative models (i.e., models that differ only in the presence or absence of one or more variables), the model with the lowest AIC should be chosen as the best model.

By contrast, when several models need to be compared, AIC weights are used—under the assumption that one of the models in the set is the optimal model. These models can be nested or nonnested (i.e., including alternative predictors). AIC weights, therefore, indicate the strength of the evidence in favor of a model in a particular set of competing models (Burnham and Anderson, 2003). The model with the highest AIC weight is taken as the one that best approximates the "true" process underlying the phenomenon under study, and the other models are evaluated in relation to that optimal model. This evaluation is based on Delta values, which correspond to the difference in AIC between the best model in the set and a particular competitor model. The following rules of thumb are usually applied to decide which models need to be considered along with the best model:

(14) AIC weights interpretation Burnham and Anderson (2003)
   a. A model with a Delta value within 1–2 of the best model has substantial support in the data, and should be considered along with the best model.
   b. A Delta value within only 4–7 units of the best model has considerably less support.
   c. A Delta value > 10 indicates that the worse model has virtually no support and can be omitted from further consideration.

This method can be used for variable selection, that is, to determine which predictor variable has the greatest influence, among a set of competitors (Burnham and Anderson, 2003). In our case, it can be used to compare the BPI with alternative language experience predictors. This is done by summing the Akaike weights of variables across all the models where the variables occur. The competing variables are then ranked using these sums. The larger this sum of weights, the more important the variable.

Following the method explained in the preceding text, we created a set of alternative models with the following alternative measures of language experience:

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

TABLE 11. Model selection for the effect of alternative measures of language experience (listed in column 1) on bilingual children's language proficiency (indexed by the SRep score) fit to 2,477 observations (i.e., total 2,468 degrees of freedom) with 3,105 null deviance. The weight indicates the probability that the model is the best one in the set. Delta is the AIC difference of a model compared with the best one.

	Resid. Df	Resid. Dev	Delta	Weight
Cumulative English Exposure	2,467	1,757.62	0.00	0.37
BPI	2,467	1,757.85	0.30	0.32
Cumulative English Use	2,467	1,757.87	1.30	0.19
Current English Exposure	2,467	1,757.35	2.70	0.09
Current English Use	2,468	1,758.06	4.70	0.04

(15) a. The BPI
b. Cumulative exposure the school language (English)
c. Current exposure to the school language (English)
d. Cumulative use the school language (English)
e. Current use of the school language (English)

As all the children in this sample were exposed to two languages only, SL measures and HL measures are equivalent from a statistical modeling perspective (albeit yielding opposite coefficients). Here we report on the measures calibrated on the SL (English).

Using the optimal model for the bilingual children's SRep data (reported in Table 7), we replaced the BPI with each of the preceding predictors in turn (i.e., refitting the model four times, once per predictor other than the BPI). As the models differed only with respect to the language experience measure, the one with the best fit can be considered the one with the most reliable predictor of language proficiency. The results of the comparison are summarized in Table 11, ordered by model weight.

There is a 37% probability that the model using Cumulative English Exposure as a predictor is the best in the set. When the difference in AIC between two models (dAIC or Delta) is < 2, it is reasonably safe to consider that both models have approximately equal weight in the data (Burnham and Anderson, 2003). There is strong evidence (Delta = 1.30) for the need to take Cumulative English Use into account alongside Cumulative English Exposure—which is in effect what the BPI does. With a Delta value of .30, the BPI can be considered to be roughly as good a predictor of English-language proficiency as Cumulative Exposure.[19] It has a 32% likelihood of being the best model in the set—which is not far off the 37% likelihood of the model based on Cumulative English Exposure.

The answer to first part of Research Question 2 (as to whether a simpler measure of language experience might fare better than a complex measure) is that the BPI, as a combined measure of cumulative language exposure and use, is not more informative as a predictor of SL proficiency, compared with a simple measure of cumulative language exposure. As explained in the preceding text, the BPI can be interpreted as a measure of cumulative exposure to the HL, adjusted for HL use: two children with the same amount of Cumulative HL Exposure will have a different BPI if one of them is relatively passive in the HL (i.e., if s/he tends to answer back in English when addressed in the HL).

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

However, the additional information encapsulated in the BPI (compared with a simple measure of cumulative language exposure) does not improve its predictive potential. This could be because language *use* does not have much of an effect over and above that of language *exposure*.

With respect to the cumulative-current comparison, cumulative measures are indeed more informative than current measures, as predictors of SL proficiency.

We now turn to the second part of Research Question 2, regarding the specific impact of language *use* as predictor of SL proficiency.

QUESTION 2.2: TEASING APART THE EFFECT OF EXPOSURE AND USE

If the amount of language use (estimated from the parental questionnaires) was perfectly correlated to the amount of language exposure, it would not be possible to disentangle their effect statistically as predictors of language proficiency. In the group under study, however, an imbalance between language use and exposure is observed: many children tended to be relatively "passive" in their HL, as shown in Figure 10. Passivity reflects the extent to which the child used their HL in relation to the extent they were exposed to it, and it can be calculated by dividing Cumulative HL Use by Cumulative HL Exposure. Children with the highest score (100%) are as active in their HL as they are exposed to it. Lower scores indicate relative passivity in the HL: the child uses that language less than they are exposed to it (0% indicating no output in the HL at all, which was the case for nine children). Importantly, a high score does not necessarily correlate with a high amount of HL exposure in our population sample: high usage is observed across the whole range of HL exposure.

Children who are relatively passive in their HL are (by implication) more active in their SL, as they answer back in English in some or all the contexts in which they are addressed in the HL. It is possible that this "extra" amount of English use confers a proficiency advantage in that language. This is what we investigate in the text that follows.

As explained in the "Methods of Analysis" section, it is not possible to include strongly correlated predictors in the same regression model, as is the case for our exposure and use measures. To remedy this problem, we derived a measure of the *difference* between Cumulative SL Exposure versus use (by subtracting Cumulative SL Exposure from Cumulative SL Use). This new measure, which we will call SL Difference, is significantly but only marginally correlated with Cumulative SL Exposure ($r = -.19$, $p < .0001$), which makes it possible to consider the two together in the same regression model.

We refitted the optimal model of the SRep accuracy data (by item), with Cumulative SL Exposure and SL Difference as language experience predictors. The effect of SL Difference was not significant ($\beta = -.005$, $p = .84$), and allowing for an interaction between Cumulative SL Exposure and SL Difference led to nonconvergence.

Given the lack of convergence with linear regression, we turned to a more powerful algorithm to investigate the relationship between SL exposure and use as predictors of SL proficiency.

The data was reanalyzed with a generalized additive mixed model (GAMM— Baayen, Vasishth, Kliegl, & Bates, 2017; Wood, 2006) using the R-package "mgcv" (version 1.8-

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

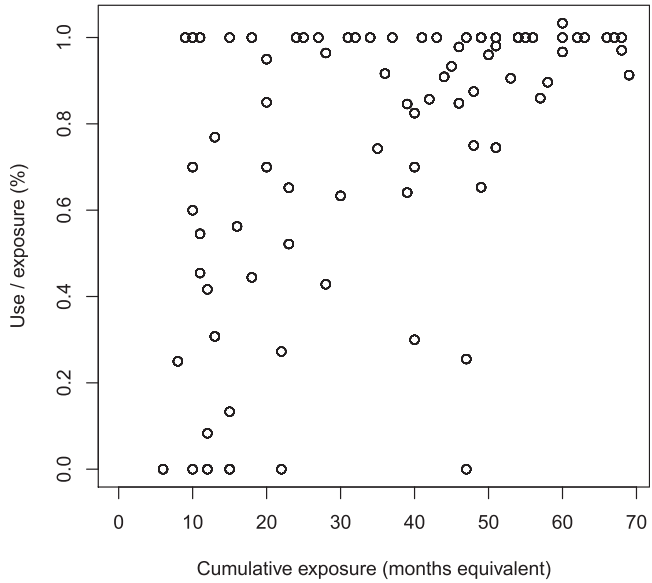


FIGURE 10. (Im)balance between cumulative use and cumulative exposure of the HL, ranging from an equal amount of both (100% active usage) to no use (total "passivity").

28). GAMMs are a variant of the linear models introduced in the preceding text. Nonlinear regression models operate according to the same broad principles as their linear variants, except that the regression line fitted through the scatterplot of observations can feature a certain amount of "wiggliness" instead of a straight line. The linear effects are included as *parametric terms* in the model (as previously mentioned), and the nonlinear effects are included as *smooths*.[20] The model has an in-built optimizer to avoid overfitting by allowing too much "wiggliness" in the smooths. In the model reported in the following text, the random effects were modeled as smooths. To probe the relationship (or lack thereof) between SL cumulative exposure and SL Difference, we allowed a nonlinear interaction of the two HL measures (Cumulative Exposure and HL Difference).

Allowing the nonlinear interaction of Cumulative SL Exposure and SL Difference did yield a significant improvement in model fit, compared with a nonlinear model without that interaction. The interaction between Cumulative SL Exposure and SL Difference reached a statistical significance ($\chi^2 = 10.45$, edf $= 3$, $p = .015$), suggesting a crossover effect of exposure and use affecting mainly children with a high amount of cumulative exposure to the SL.

To test the robustness of the modest significance level of the interaction between language exposure and use, we fitted an alternative model allowing our other main environmental predictor of interest (i.e., SES) to interact nonlinearly with language exposure.

This alternative model revealed a much more important nonlinear interaction at play in the SRep data, which renders the impact of SL passivity nonsignificant (even in interaction with cumulative SL exposure): the nonlinear interaction between SES and

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

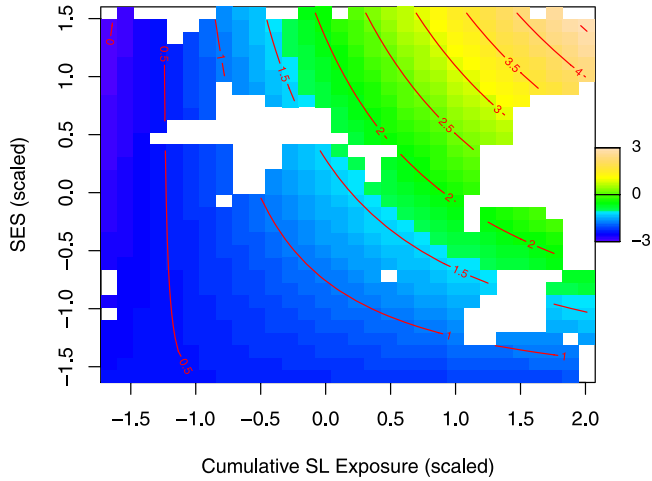


FIGURE 11. Nonlinear interaction between Cumulative SL Exposure and SES as predictors of bilingual children's accuracy in the SRep task estimated through a GAMM. Average SRep performance corresponds to 0 on the color gradient.

cumulative SL exposure turns out to be the strongest predictor of accuracy in the SRep test ($\chi^2 = 30.69$, edf = 3, $p = .000001$). Figure 11 shows that, at relatively low levels of cumulative exposure to the SL, SES does not confer an advantage in the SRep test; but at medium-to-high levels of cumulative SL exposure, accuracy in the SRep test is strongly correlated with SES.

SUMMARY

In answer to Research Question 2, we have demonstrated that cumulative exposure to the SL is the best predictor of proficiency in that language, both compared with current measures and compared with a combined measure of cumulative exposure and use. In attempting to tease apart the impact of language exposure and language use, we discovered that language use was not a robust predictor of SL proficiency. The marginally significant nonlinear interaction between language exposure and use was trumped by a very strong nonlinear interaction between language exposure and SES. Children with below-average exposure to the SL did not benefit from an SES advantage (in terms of proficiency score). But in children with above-average exposure to the SL, performance improved exponentially as the levels of language exposure and SES increased.

We now turn to the "catch up" question (Research Question 3): From what amount of cumulative exposure to the SL do bilingual children perform within the monolingual range?

QUESTION 3: CRITICAL THRESHOLDS OF LANGUAGE EXPERIENCE

Even within the same age group, monolingual children vary in terms of language proficiency (see Kidd, Donnelly, & Christiansen, 2018 for a recent review). Identifying

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

what counts as monolingual-like performance needs to take that variability into account. One approach in the literature has been to rely on a threshold of 1.25 standard deviation below the mean to identify monolingual children at risk of a Developmental Language Disorder (Conti-Ramsden, 2003; Tomblin, Records, & Zhang, 1996). Any proficiency score above that threshold is considered "normal," which suggests a substantial amount of variation within the "normal" range. Our purpose here is not to identify Developmental Language Disorder among bilinguals, but to identify bilinguals who have caught up with their monolingual peers in the SL. One might suggest that this would require them to score above the −1.25 SD threshold (defined in relation to the mean achieved by their monolingual peers).

Instead of relying on an arbitrary threshold on raw scores (such as −1.25 SD), we propose to use an objective, data-driven method to identify the threshold in language experience beyond which bilingual children can be expected to perform within the monolingual range in terms of proficiency in the SL. This method will allow us to control for the other factors that we know have a significant impact on language proficiency, such as SES and cognitive factors.

In the following text, we carry out the threshold analysis based on cumulative exposure to the SL (i.e., an additive measure) and on language dominance (i.e., a proportion measure) in turn, to address the role of cumulative exposure and language dominance.

CUMULATIVE EXPOSURE THRESHOLDS

Our starting point is the optimal (nonlinear, mixed-effect) model for the SRep data, which had been fitted to the target accuracy data for the bilingual children only. We rely on cumulative exposure as our reference measure for bilingual language experience, as this proved the best predictor in the population under study. We first refit this model to the entire cohort of children (including monolinguals)—using exposure to English as a predictor. When the model is fitted to the entire cohort using that predictor, age no longer accounts for any unique variance. We therefore exclude it from the model. After confirming the significance of the nonlinear interaction between SES and Cumulative SL exposure in the whole-cohort model (not reported here), we perform the threshold analysis, as explained in the following text.

The procedure for the threshold analysis consists in replacing the continuous SL exposure predictor with a binary one, for which the cutoff point would be determined objectively, based on Information Theory. To identify the optimal cutoff point, we refit the model recursively, using each time a different SL exposure value as the binary cutoff point. As the cumulative SL experience measure features 70 different values in the group under study, the model needs to be fitted 70 times. Plotting the AIC values for all the models (shown in Figure 12) shows the fluctuations in model fit depending on SL exposure threshold. The lowest AIC value identifies the model with the best fit. The cutoff point for that model is interpreted as the amount of exposure beyond which a bilingual child can be expected to perform within the monolingual range.

The optimal cutoff point on the continuum of SL cumulative exposure is at 33 months equivalent (where the dashed line falls in Figure 12). The model therefore predicts that a (5- to 7-year-old) bilingual child who has been exposed to the SL for the equivalent at least 33 months can be expected to perform within the monolingual range (after controlling for SES and memory abilities) in the SRep task.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

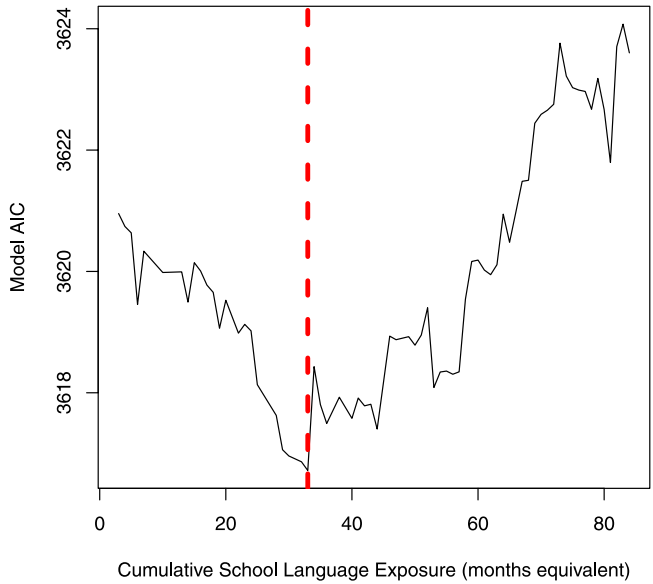


FIGURE 12. Optimal threshold of cumulative SL exposure above which bilinguals perform within the monolingual range in the SRep task. The continuous line represents fluctuations in AIC (indicating model fit) depending on where the cutoff point is situated along the SL exposure continuum. The dashed line indicates the lowest AIC, corresponding to the best-fitting model, and hence the optimal cutoff point.

The same procedure is then repeated for the other aspects of English proficiency we assessed in this study.[21] This results in the following thresholds of cumulative SL exposure: lexical semantics (composite DELV score): 44 months equivalent; discourse semantics: 32 months equivalent (see Figure 13).

The overall picture, when combining the information from the three preceding analyses, is that from 32 or 33 months equivalent of cumulative exposure to the SL, bilingual children (between 5 and 7 years of age) perform within the range of monolingual children in *some* aspects of language proficiency (indexed by the lexical semantics and the SRep tasks). The discourse semantics results suggest a more conservative threshold of 44 months equivalent (with respect to lexical semantics).

In the next section, we compare our threshold analysis to a more traditional approach based on language dominance.

LANGUAGE DOMINANCE THRESHOLDS

In the literature on bilingualism, it is common practice to consider that, below 20% of current exposure to an additional language, children should be considered "functionally monolingual" in their dominant language (e.g., Bedore et al., 2012; Gutiérrez-Clellen and Kreiter, 2003; Pearson et al., 1997). As its name indicates, functional monolingualism implies that the child should perform within the monolingual range in terms of language proficiency.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

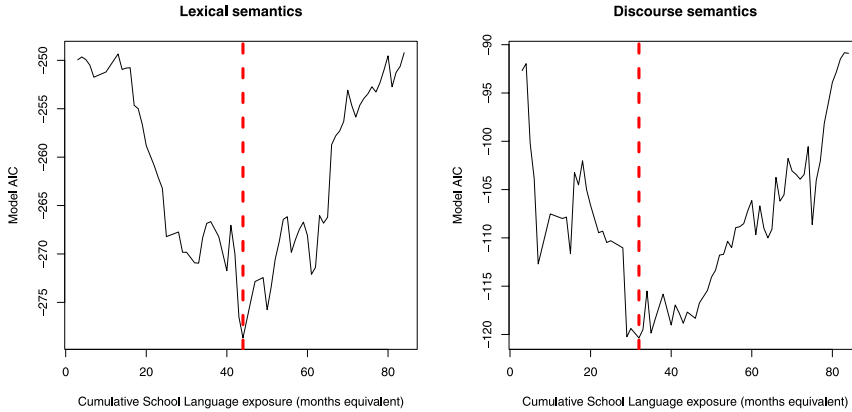


FIGURE 13. Optimal threshold of cumulative SL exposure above which bilinguals perform within the monolingual range in the lexical semantics and discourse-semantics tasks. The continuous lines represent fluctuations in AIC (indicating model fit) depending on where the cutoff point is situated along the SL exposure continuum. The dashed lines indicate the lowest AIC, corresponding to the best-fitting models.

Assuming that cumulative language exposure is a reliable proxy for language dominance (as demonstrated by Unsworth, 2015), we operationalized language dominance as the proportion of cumulative exposure to the SL out of the child's lifetime (in months). We then replicated the threshold analyses above using that estimate of language dominance as our predictor of interest.

The threshold analysis of the SRep data using language dominance as predictor of interest suggests a cutoff point of 57% dominance in the SL (in 5 to 7 year olds).

Figure 14 compares the threshold based on the additive measure (corresponding to the horizontal line) with the threshold based on the proportional measure (cf. the vertical line). A number of data points are found in the top left quadrant, corresponding to the children that are predicted to perform within the monolingual range according to the additive threshold, but not according to the proportional threshold. The threshold analysis based on language dominance is therefore more conservative than that based on the cumulative measure of language exposure.

Figure 15 compares Bedore et al.'s categories with our objective threshold for monolingual-like performance.

SUMMARY AND DISCUSSION

The discussion is organized around our three research questions, which are repeated in the following text for clarity.

Research Question 1: Does a gradient measure of bilingual language experience (combining exposure and use) and a gradient measure of SES (as a proxy for the richness of the child's language environment) significantly predict the following aspects of SL proficiency: comprehension and production of complex sentences, lexical semantics, and discourse-semantics? We proposed a gradient measure combining cumulative HL exposure and use (the BPI) and demonstrated that it

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

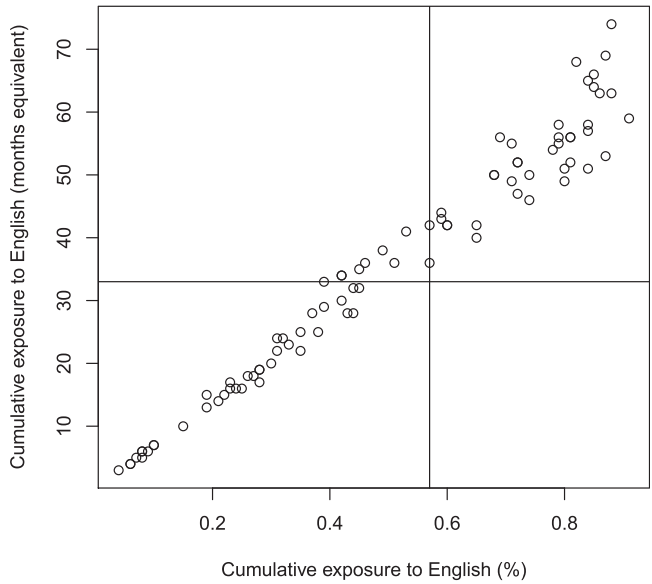


FIGURE 14. Cumulative exposure to the SL expressed in months equivalent (y-axis) and as a proportion of total language exposure (x-axis). The lines indicate the threshold of exposure required for monolingual-like performance in the SRep test.

significantly predicts 5- to 7-year-old bilingual children's proficiency in the SL. This finding was replicated with respect to different aspects of English proficiency, including morpho-syntax (indexed by the LITMUS SRep test), lexical semantics (indexed by the DELV lexical tests), and discourse-semantics (indexed by the DELV articles test). Our results are in line with a substantial body of research demonstrating the impact of language exposure on language proficiency in bilinguals (see Paradis, 2017 for a recent review), and expand on the bulk of these studies by considering a range of proficiency measures.

Some aspects of language were found to be more sensitive than others to language exposure. Functional errors in the SRep task were predicted by language experience more strongly than lexical errors, while inflectional errors were not predicted by language experience at all, suggesting that a sufficient amount had already been received by most children in this study to acquire those aspects of English grammar. This is in line with studies showing that the nonuniform impact of language exposure on different aspects of grammar reflects their acquirability (Schulz & Grimm, 2019; Tsimpli, 2014).

Children's SES was a significant source of substantial individual differences, over and above the effect of the cognitive predictors (which are independently known to correlated with SES—see e.g., Gathercole et al., 2016). Along with many others (e.g., Bohman et al., 2010; Hoff, 2013; Meir and Armon-Lotem, 2017), we interpret SES as a proxy for the richness of the (linguistic) environment experienced by the child, which results in qualitative and quantitative differences in children's language experience. Further research will be required to identify which of the correlates of SES genuinely explains the

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

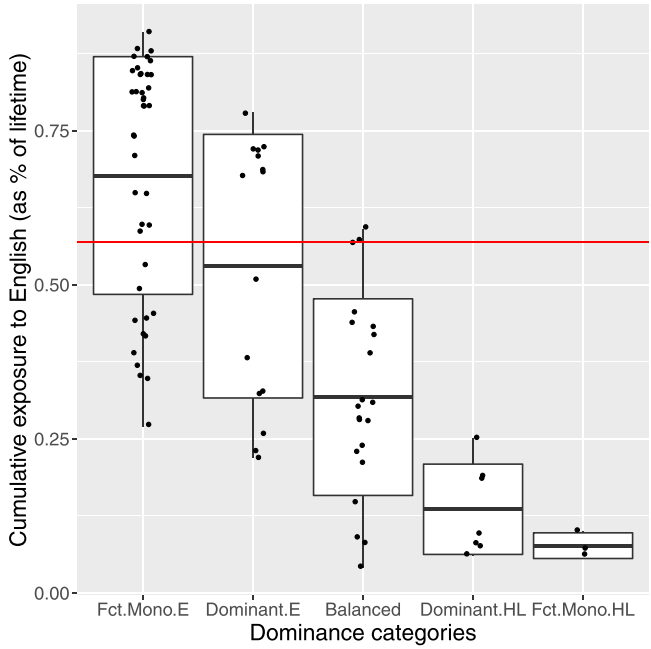


FIGURE 15. Comparison of Bedore et al.'s language dominance categories and a continuous measure of language dominance based on cumulative exposure to the SL (in months equivalent). The horizontal line indicates the threshold of exposure required for monolingual-like performance in the SRep test.

nature of its effect. One potentially important factor that we did not consider in the present study is the home literacy environment and the extent to which reading and writing in the HL and/or the SL correlates with SES and in turn affects children's language skills. Brinchmann, Braeken, and Lyster (2019), for instance, argue that the home literacy environment accounts for individual differences in vocabulary and grammar in preschool monolingual children. A hypothesis for future research is whether (and to what extent) the amount and diversity of literacy activities explain the relationship between SES and language proficiency in bilingual children.

SES was found to predict children's performance in the SRep task and in the discourse semantics tasks, but not in the lexical semantics tasks. We speculate that this could be due to the very strong impact of cognitive predictors on the latter measures. SES is a strong predictor of performance on cognitive flexibility (as demonstrated by De Cat, Gusnanto, & Serratrice, 2018 for this group of children), and might therefore have only an indirect impact on performance in the lexical semantics task.

An important finding of this study is that SES interacts with language experience in a complex way, which we modeled as the nonlinear interaction of SES and cumulative exposure to the SL: the extent to which SES confers an advantage in terms of SL proficiency depends on the amount exposure to that language. In our sample, children with below-average cumulative exposure to the SL hardly benefited from any SES advantage, whereas at higher levels of cumulative exposure to the SL, that advantage grew exponentially in relation to SL exposure. In other words, SES was only associated

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

with a higher proficiency score in the children who had received a substantial amount of exposure to the SL over their lifetime. What our cross-sectional study was not able to investigate is the possibility that children from higher SES might be able to catch up faster with their monolingual peers. This will need to await future research based on longitudinal data.

Research question 2 sought to identify the optimal way of modeling language experience, as a predictor of SL proficiency. We investigated it in three steps.

How does a combined measure of language experience (combining exposure and use) compare with simpler measures (i.e., exposure *or* use), as predictors of SL proficiency? Adopting an information-theoretic approach, we found that a more precise measure of cumulative language experience (the BPI, which encompasses exposure and use) did not improve predictive accuracy compared with a simpler measure of cumulative HL exposure. This finding will be discussed in relation to the issue of language exposure versus use in the following text.

Are *cumulative* measures more accurate than current measures, in predicting SL proficiency? To our knowledge, this is the first study to systematically compare different measures of language experience as predictors of language proficiency in bilingual children (i.e., current exposure, current use, cumulative exposure, cumulative use, and a combined measure of cumulative exposure and use). Our information-theoretic comparison demonstrated that cumulative experience measures fare better as predictors of language proficiency than current experience measures. Future research will be needed to determine if this is the case in older age groups as well.

What is the specific impact of language *use* as a predictor of SL proficiency? Analyses using decorrelated measures confirmed that there was no robust evidence for an effect of language use over and above the effect of exposure to the SL. In other words, if a child uses the SL even in (some of) the contexts in which they are addressed in the HL, that "extra" use does not confer a tangible advantage in terms of SL proficiency. This is likely due to the leveling effect of schooling, and/or possibly to the fact that the children's English utterances in HL contexts are not elaborated upon in English by their interlocutor. We hypothesize that language use will, however, be a significant predictor of proficiency in the HL (as many children were relatively "passive" in that language). The BPI might thus prove to be more accurate in predicting *home* language proficiency. This could not be investigated in this study, given the number of HLs represented in our sample.

Research question 3 focused on the identification of functional thresholds in bilingual language experience in relation to SL proficiency: What is the critical amount of SL experience required for bilingual children to perform within the monolingual range? In other words: At what point can one expect the gap between bilinguals and monolinguals to be "closed" in the SL? We exploited an objective, data-driven method to investigate this based on (a) an additive measure of language experience and (b) a language dominance measure.

Once SES and cognitive abilities have been taken into account, the threshold of exposure required for bilingual children to perform within the monolingual range was found to vary depending on the proficiency measure. The measures of discourse semantics (indexed by the mastery of definiteness distinctions) and complex morphosyntax (indexed by the SRep task) showed that children performed within the monolingual

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

range from 32 or 33 months-equivalent of exposure to the SL. A more conservative threshold of 44 months-equivalent was suggested by the lexical semantics measure. It is not surprising to find that different aspects of language proficiency show different levels of sensitivity to language exposure, as already alluded to previously. Jia and Paradis (2017) show that morpho-syntactic structures encoding whether a referent is new or old information can be acquired by bilingual children even with limited language exposure. This is consistent with our finding that the lowest amount of exposure to English required for performance within the monolingual range was observed in the discourse-semantics task (which tested children's knowledge of the *a/the* distinction in English). However, the higher threshold observed with respect to the lexical semantics measure needs to be interpreted with caution. That test proved the most challenging both for bilinguals and monolinguals (with a mean overall score of 63% in the monolinguals, compared with 81% in the SRep test, 83% in the sentence comprehension test, and 77% in the discourse-semantics test), and it is possible that the higher threshold might in fact be due to age as a latent variable (which in turn could be due in part to the cognitive demands of the task).

If expressed in terms of language dominance (rather than cumulative experience), our results indicate that, in this age group, children with at least 57% of exposure to the SL (averaged over their lifetime) are likely to perform within the monolingual range. This aligns closely with the 60% threshold suggested for toddlers by Cattani et al. (2014), but is more lenient than an 80% language exposure cutoff point commonly used for "functional monolingualism."

The approach adopted by Bedore et al. (2012) is more lenient too, but for a different reason. For them, functional monolingualism obtains if there is at least 80% exposure to *or* use of a language (and 20% or less exposure or use of another language). This disjunctive criterion gives equal weight to exposure and use: if either falls below 20%, the child is categorized as functionally monolingual in their other language. According to the criteria proposed by Bedore et al. (2012), 41 of the 87 bilinguals in our sample would be classified as functionally monolingual in English. Yet, many children in that group perform below the monolingual range, as identified by our threshold analyses (see Figure 15).

One reason for this discrepancy could be that Bedore et al.'s classification of language dominance is based on current language experience, which we have seen is less informative a predictor of language proficiency than cumulative language experience in our sample. Another, likely more important, reason is that their classification assigns equal weight to language exposure and language use, as explained in the preceding text. While we agree it is important to take language use into account, it appears not to have much impact on proficiency *in the SL*, as pointed out above. Further research will be required to determine objectively the relative importance of language exposure and use as components of bilingual language experience. For instance, language use is likely to have a significant impact on proficiency in the HL, and possibly on executive functions in the cognitive domain. The investigation of phenomena that are sensitive to language use (over and above language experience) will provide fertile grounds to evaluate how it should be modeled. Should language use and exposure have equal weight, as proposed by Bedore et al. (2012)? Or should it have a weaker impact, as modeled in our BPI? We leave it to future research to evaluate the merit of each approach, including whether the language of the conversational partner has an impact.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Finally, three cautionary notes are in order, regarding the interpretation of our findings.

First, while critical thresholds can be useful for practical purposes (e.g., selection criteria for experimental studies, or the identification of children in need of support with their SL), we agree with Treffers-Daller (2015), Luk (2015), and many others that language dominance and bilingualism are best treated as continua rather than discrete categories.

Second, the functional thresholds identified need to be interpreted in light of other significant predictors, that is, SES and cognitive factors. For instance, a child with more than 60% exposure to the SL over her lifetime will not necessarily score within the monolingual range of proficiency if she comes from a deprived background. All the proficiency tests used in this study required a certain amount of cognitive effort, and this induced a significant amount of individual variation in the scores. STM had a significant impact on performance in the SRep test[22] and the lexical-semantic tests. WM had a significant impact on performance in the lexical semantic tasks, but not the SRep task nor the discourse semantic task. The Sentence Structure test of the CELF proved to be more sensitive to children's cognitive abilities than to their mastery of complex syntax, due to an overwhelming confound caused by the cognitive demands of the task (see De Cat, 2019). It was therefore not discussed in this article.

Third, the children in this study were all in the early stages of formal education, and most bilinguals will have caught up with their monolingual peers by the end of primary school—except if part of an at-risk group (which we were not able to investigate here; see Strand & Demie, 2005). Although bilingualism might result in an apparent delay in terms of SL proficiency in the early years of formal education, it tends to turn into an asset once the proficiency gap between monolinguals and bilinguals has narrowed sufficiently (or closed). If bilingual children are compared with monolinguals of similar levels of English proficiency, the bilingual group already shows an advantage in social, emotional, and behavioral functioning in Reception year (around 5 years of age), and a functional advantage in meeting curriculum targets at Year 2 (around 7 years of age)—see for example, Dowdy, Dever, DiStefano, and Chin (2011); Halle, Hair, Wandner, McNamara, and Chien (2012); and Whiteside, Gooch, and Norbury (2017).

Investigating language development in its linguistic and cognitive complexity in a heterogeneous population of children presents nontrivial challenges. The analytical approach taken here goes some way to propose a solution to dealing with complex data and to making sense of the multifaceted bilingual experience that is common in many schools today.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0272263119000597

NOTES

[1]Percentages are given for those languages representing more than 5% of the sample.

[2]The Goldthorpe class scheme allocates individuals and families into categories of social class, based on their income, degree of economic security, and chances of economic advancement, as well as on their degree of autonomy in performing their work tasks and roles.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

[3]Parents were asked "How well do you speak English?," and could choose between "not at all," "not well," "quite well," and "very well."

[4]If there was more than one interlocutor in a time window, the time window was divided by the number of interlocutors.

[5]At the time, the version we used was called the *School-Age Sentence Imitation Test*.

[6]After three questions, the prompt sentence is repeated to help the child remember it.

[7]We refer the interested reader to De Cat et al. (2018) for the analysis of the cognitive measures used in this study.

[8]The fit of a model estimates how closely it matches the observed values in the dataset. The AIC is an estimate of the model fit, penalized for overfitting (i.e., the inclusion of too many parameters). The smaller the AIC, the better the model.

[9]The R-squared of a model expresses how much variance is captured by the model.

[10]Standardized scores, also known as z-scores, are rescaled so that they have a mean of zero and a standard deviation of one. Standardized scores were derived for all children as one group.

[11]Current HL measures: $r = 0.86$, $p < .0001$; cumulative HL measures: $r = 0.93$, $p < .0001$.

[12]This PCA did **not** include any other variable than the cumulative measures of HL exposure and use.

[13]Loadings are the linear combinations of coefficients that correspond to the covariances/correlations between the original variables and the unit-scaled components. They are used to interpret principal components. We reversed the scores for ease of interpretability, so that a high value corresponds to a high amount of HL experience.

[14]The BPI correlates significantly with current exposure (Pearson's product-moment correlation $r = 0.93$, $p < 0.0001$), current use (Pearson's product-moment correlation $r = 0.89$, $p < 0.0001$), cumulative exposure (Pearson's product-moment correlation $r = 0.99$, $p < 0.0001$), and cumulative use (Pearson's product-moment correlation $r = 0.99$, $p < 0.0001$) in the HL.

[15]Dispersion is relatively modest in the case of cumulative measures. We come back to this in the comparative analyses that follow.

[16]As seen in Figure 4, performance according to the Full Accuracy analysis appears much poorer in both bilingual and monolingual children, with bilingual children only achieving a mean score of 35% (compared with 55% for the monolinguals) and several bilingual children obtaining an overall score of zero.

[17]The estimates are expressed as the log of the odd ratio.

[18]Whether sequential bilingualism also affects language development qualitatively remains a controversial issue—see e.g., Meisel (2009) and Unsworth (2013a). We do not address it in this study.

[19]The results for HL measures (not reported here) are exactly identical to their English counterpart reported in Table 11, so the fact that the BPI was based on HL measures does not make a difference in this respect.

[20]A smooth is a function that determines the wiggliness of the line (or surface, in the case of interactions).

[21]This was not done for the CELF Sentence Structure test, as language experience was not found to be a significant predictor of performance in the bilinguals.

[22]This was entirely expected, as this test can be used as a verbal measure of WM. In our study we scored children's performance based on the correct repetition of the target structure rather than exact word-for-word repetition. This alleviated the impact of memory factors.

REFERENCES

Avons, S. E., Wragg, C. A., Cupples Wragg, L., & Lovegrove, W. J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics*, *19*, 583–601. doi: 10.1017/S0142716400010377

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

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

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cognitive Science*, *4*, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173. doi: 10.1016/S1364-6613(00)01538-2

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, *83*, 413–422. doi: 10.1111/j.1467-8624.2011.01707.x

Bedore, L., Peña, E., Summers, C., Boerger, K., Resendiz, M., Greene, K., Bohman, T. M., & Gillam, R. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition*, *15*, 616–629. doi: 10.1017/S1366728912000090

Bernardini, P., & Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The Ivy hypothesis. *Bilingualism: Language and Cognition*, *7*, 49–69. doi: 10.1017/S1366728904001270

Bohman, T. M., Bedore, L., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism*, *13*, 325–344. doi: 10.1080/13670050903342019

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning*, *63*, 211–242. doi: 10.1111/lang.12003

Brinchmann, E. I., Braeken, J., & Lyster, S.-A. H. (2019). Is there a direct relation between the development of vocabulary and grammar? *Developmental Science*, *22*, e12709. doi: 10.1111/desc.12709

Burnham, K., & Anderson, D. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer Science & Business Media.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*, 278–288. doi: 10.1016/j.cognition.2013.11.015

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition*, *20*, 13–16. doi: 10.1017/S1366728915000863

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*, 649–671. doi: 10.1111/1460-6984.12082

Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research*, *59*, 1179–1189. doi: 10.1044/2016_JSLHR-L-15-0293

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *46*, 1029–1037. doi: 10.1044/1092-4388(2003/082)

Cook, V. (1994). Universal grammar and the learning and teaching of second languages. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 25–48). Cambridge: Cambridge University Press.

De Cat, C. (2019). What does the Sentence Structure component of the CELF-IV index, in monolinguals and bilinguals? doi: 10.31219/osf.io/8nkwc

De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, *40*, 119–151. doi: 10.1017/S0272263116000486

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*, 411–424. doi: 10.1017/S0142716407070221

De Houwer, A. (2009). *Bilingual first language acquisition.* Clevedon, UK: Multilingual Matters.

De Houwer, A. (2011). Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, *2*, 221–240. doi: 10.1515/9783110239331.221

Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. K. (2011). Screening for emotional and behavioral risk among students with limited English proficiency. *School Psychology Quarterly*, *26*, 14–26.

Engel de Abreu, P., Baldassi, M., Puglisi, M., & Befi-Lopes, D. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language and Hearing Disorders*, *56*, 630–642. doi: 10.1044/1092-4388(2012/12-0079)

Engel de Abreu, P., Gathercole, S., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, *21*, 569–574. doi: 10.1016/j.lindif.2011.06.002

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). *Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance.* Hoboken, NJ: Wiley.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Friedmann, N., & Rusou, D. (2015). Critical period for first language: The crucial role of language input during the first year of life. *Current Opinion in Neurobiology*, *35*, 27–34. doi: 10.1016/j.conb.2015.06.003

Gathercole, V. C. M., Kennedy, I., & Môn Thomas, E. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition*, *19*, 51057–51078. doi: 10.1017/S1366728915000504

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, *22*, 611–632. doi: 10.1017/S0305000900009971

Goldstein, B., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology*, *19*, 238–247. doi: 10.1044/1058-0360(2010/08-0086)

Goldthorpe, J. (1980). *Social mobility and class structure in modern Britain.* Oxford: Clarendon.

Grosjean, F. (1989). Neurolinguists, beware! the bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15. doi: 10.1016/0093-934X(89)90048-5

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

Grüter, T., & Paradis, J. (Eds.). (2014). *Input and experience in bilingual development.* Amsterdam, The Netherlands: John Benjamins.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, *24*, 267–288. doi: 10.1017/S0142716403000158

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early vs. later English language proficiency among English language learners. *Early Childhood Research Quarterly*, *27*, 1–20. doi: 10.1016/j.ecresq.2011.07.004

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*, 55–88. doi: 10.1016/j.dr.2005.11.002

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*, 4. doi: 10.1037/a0027238

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*, 1–27. doi: 10.1016/j.cogpsych.2010.08.002

Huttenlocher, J., Waterfall, H., Vasilyeva, M., & Vevea, J. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *16*, 343–365.

Jardak, A., & Byers-Heinlein, K. (2018). Labels or concepts? The development of semantic networks in bilingual two-year-olds. *Child Development*, *90*, e212–e229. doi: 10.1111/cdev.13050

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131–161. doi: 10.1017/S0142716403000079

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, *23*, 599–621. doi: 10.1017/S0142716402004058

Jia, R., & Paradis, J. (2017). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, *18*, 737–752. doi: 10.1017/S1366728914000728

Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition*, *18*, 27–28. doi: 10.1017/S1366728914000510

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension. *Topics in Language Disorders*, *33*, 208–223. doi: 10.1097/TLD.0b013e31829d623e

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

Kupisch, T. (2007). Determiners in bilingual German–Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition*, *10*, 57–78. doi: 10.1017/S1366728906002823

Laloi, A., de Jong, J., & Baker, A. (2017). Can executive functioning contribute to the diagnosis of SLI in bilingual children? *Linguistic Approaches to Bilingualism*, *7*, 431–459. doi: 10.1075/lab.15020.lal

Lanza, E. (2004). *Language mixing in infant bilingualism: A sociolinguistic perspective.* Oxford: Clarendon Press.

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Lauro, L. J. R., Reis, J., Cohen, L. G., Cecchetto, C., & Papagno, C. (2010). A case for the involvement of phonological loop in sentence comprehension. *Neuropsychologia*, *48*, 4003–4011. doi: 10.1016/j.neuropsychologia.2010.10.019

Luk, G. (2015). Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition*, *18*, 35–36. doi: 10.1017/S1366728914000625

Magimairaj, B., & Montgomery, J. (2012). Children's verbal working memory: Role of processing complexity in predicting spoken sentence comprehension. *Journal of Speech, Language and Hearing Disorders*, *55*, 669–682. doi: 10.1044/1092-4388(2011/11-0111)

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*, 95–119. doi: 10.1016/j.jecp.2005.07.005.

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–122). Bristol, UK: Multilingual Matters.

Marinis, T., Chiat, S., Armon-Lotem, S., Gibbons, D., & Gipps, E. (2010). *School-age sentence imitation test (SASIT).* Reading, UK: University of Reading.

McDonald, J. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language*, *35*, 2247–268. doi: 10.1017/S0305000907008367

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, *8*, 1442. doi: 10.3389/fpsyg.2017.01442

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*, 15–34. doi: 10.1515/ZFSW.2009.002

Montgomery, J. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research*, *38*, 187–199. doi: 10.1044/jshr.3801.187

Montgomery, J., Magimairaj, B., & O'Malley, M. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, *37*, 331–354. doi: 10.1007/s10936-008-9077-z

Montrul, S. (2015). Dominance and proficiency in early and late bilingualism. C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 15–35). Cambridge: Cambridge University Press.

Papagno, C., Cecchetto, C., Reati, F., & Bello, L. (2007). Processing of syntactically complex sentences relies on verbal short-term memory: Evidence from a short-term memory patient. *Cognitive Neuropsychology*, *24*, 292–311. doi: 10.1080/02643290701211928

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

Paradis, J. (2017). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism, Language and Cognition*, *20*, 27–28. doi: 10.1017/S136672891600033X

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French–English bilingual children with SLI: How do they compare with their monolingual peers? *Journal of Speech, Language, and Hearing Research*, *46*, 113–127. doi: 10.1044/1092-4388(2003/009)

Paradis, J., Rusk, B., Duncan, T. S., & Govindarajan, K. (2017). Children's second language acquisition of English complex syntax: The role of age, input, and cognitive factors. *Annual Review of Applied Linguistics*, *37*, 148–167. doi: 10.1017/S0267190517000022

Pearson, B., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*, 41–58. doi: 10.1017/S0142716400009863

Place, S., & Hoff, E. (2011). Proficiency Properties of dual language exposure that influence two-year-olds' bilingual proficiency. *Child Development*, *82*, 1834–1849. doi: 10.1111/j.1467-8624.2011.01660.x

Ribot, K., Hoff, E., & Burridge, A. (2017). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development*, *89*, 929–940. doi: 10.1111/cdev.12770

Romaine, S. (1989). *Bilingualism.* Oxford: Blackwell.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Schafer, R., & de Villiers, J. (2000). Imagining articles: What a and the can tell us about the emergence of DP. In S. C. Howell, S. Fish, & T. Keith-Lucas (Eds.), *Proceedings of BUCLD, vol. 24* (pp. 609–620). Sommerville, MA: Cascadilla Press.

Schulz, P., & Grimm, A. (2019). The age factor revisited: Timing in acquisition interacts with age of onset in bilingual acquisition. *Frontiers in Psychology*, *9*, 2732. doi: 10.3389/fpsyg.2018.02732

Semel, E., Wiig, E., & Secord, W. (2006). *Clinical evaluation of language fundamentals* (4th ed.). London: Harcourt Assessment.

Seymour, H. N., Roeper, T., & de Villiers, J. (2005). *DELV-NR (diagnostic evaluation of language variation) norm-referenced test.* San Antonio, TX: The Psychological Corporation.

Silva-Corvalán, C., & Treffers-Daller, J. (Eds.). (2015). *Language dominance in bilinguals: Issues of measurement and operationalization.* Cambridge: Cambridge University Press.

Strand, S., & Demie, F. (2005). English language acquisition and educational attainment at the end of primary school. *Educational Studies*, *31*, 275–291. doi: 10.1080/03055690500236613

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

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*, 1–16. doi: 10.1016/j.jcomdis.2012.08.002

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate assessment of two languages? *Journal of Multilingual Communication Disorders*, *4*, 1–21. doi: 10.1080/14769670500215647

Thorn, A., & Gathercole, S. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology Section A*, *52*, 303–324. doi: 10.1080/713755823

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *39*, 1284–1294. doi: 10.1044/jshr.3906.1284

Treffers-Daller, J. (2015). Language dominance: The construct, its measurement, and operationalization. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 235–265). Cambridge: Cambridge University Press.

Tsimpli, I. M. (2014). Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, *4*, 283–313. doi: 10.1075/lab.4.3.01tsi

Unsworth, S. (2013a). Assessing age of onset effects in (early) child L2 acquisition. *Language Acquisition*, *20*, 74–92. doi: 10.1080/10489223.2013.766739

Unsworth, S. (2013b). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*, 86–110. doi: 10.1017/S1366728912000284

Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization.* Cambridge: Cambridge University Press.

Unsworth, S. (2017). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Lifespan perspectives on bilingualism* (pp. 136–196). Berlin: Mouton de Gruyter/APA.

Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, *36*, 527–548. doi: 10.1093/applin/amt052

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

Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children: Manual.* San Antonio, TX: Psychological Corporation, Harcourt Brace Jovanovich.

Whiteside, K. E., Gooch, D., & Norbury, C. F. (2017). English language proficiency and early school attainment among children learning English as an additional language. *Child Development*, *88*, 812–827. doi: 10.1111/cdev.12615

Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, *9*, 349–363. doi: 10.1080/09658210143000155

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press

Wood, S. (2006). *Generalised additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall/CRC.

Zelazo, P. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297–301. doi: 10.1038/nprot.2006.46

https://doi.org/10.1017/S0272263119000597 Published online by Cambridge University Press