

Identifying a threshold for the executive function advantage in bilingual children

Article

Accepted Version

De Cat, C., Gusnanto, A. and Serratrice, L. ORCID:
<https://orcid.org/0000-0001-5141-6186> (2018) Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40 (1). pp. 119-151. ISSN 1470-1545 doi: 10.1017/S0272263116000486 Available at <https://centaur.reading.ac.uk/69130/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1017/S0272263116000486>

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Identifying a threshold for the executive function advantage in bilingual children

Cecile De Cat

Linguistics & Phonetics, University of Leeds

Arief Gusnanto

Statistics, University of Leeds

Ludovica Serratrice

Psychology and Clinical Language Sciences, University of Reading

Abstract

The literature exploring the executive function correlates of bilingualism is vast, but to date, few studies have concentrated on children, for whom the bilingual advantage appears even more inconsistent than for adults. We investigate a highly heterogeneous group of children (in terms of bilingual experience and socio-economic status) and identify the critical threshold of bilingual experience from which an advantage can be observed at group level. The modeling methods adopted allow the use of fine-grained, continuous factors for age and socio-economic status, thereby effectively controlling for their effect and isolating the specific effect of bilingual experience. We pioneer the use of Cox Proportional Hazard regression to analyze performance in the Simon task. This allows the modeling of *all* data points without transformation nor outlier removal, and captures both accuracy and reaction time within the same analysis, while also being able to handle multiple predictor variables.

Introduction

In the past fifteen years, the amount of research investigating the cognitive correlates of bilingualism has increased exponentially. Much research has focused on the effect of bilingualism on domain-general cognitive skills, and in particular executive functions — see for instance two recent keynote articles by Baum and Titone (2014) and Valian (2015).

This research was funded by a grant from the Leverhulme Trust (RPG-2012-633), which is gratefully acknowledged. Warm thanks to Sanne Berends and Furzana Shah for the data collection and coding, to the two anonymous reviewers for their constructive comments and challenges, to the many schools who opened their doors to our project, to the participating children for their enthusiasm and their efforts, and to their parents for filling lengthy questionnaires.

Miyake and colleagues’ seminal work (Friedman et al., 2006; Miyake et al., 2000) originally divided executive functions into three separable components: the updating and monitoring of working memory representations, the shifting between tasks and mental sets, and the inhibition of unwanted responses. These three components are dissociable (featuring unique variance) but inter-related (featuring common variance) (Miyake et al., 2000). This means that in practice, it is impossible to test one component to the complete exclusion of others. This tripartite view has now been superseded by a revised model in which two specific components are distinguished (i.e. *updating* and *shifting*, as defined above), in addition to a third component that is common to both of them, defined as the “ability to actively maintain task goals and goal-related information and use this information to effectively bias lower-level processing” (Miyake & Friedman, 2012, 12). Valian (2015) proposes to interpret this common factor variable as *inhibition*. We will adopt this label in the remainder of the paper.

Executive function (henceforth EF) development has been shown to span several important stages of maturation from early childhood (Garon, Bryson, & Smith, 2008) to adolescence (Brocki & Bohlin, 2004). In adults, EF has been shown to decline with age (Bryan & Luszcz, 2000; Buckner, 2004).

A growing body of evidence suggests that bilingual experience confers an EF advantage. Specifically, the constant need to control one language while using the other (Kroll, Dussias, Bogulski, & Valdes Kroff, 2012) is claimed to exercise the neurological mechanisms underlying attention control, which results in the enhancement of bilinguals’ abilities in that domain (see e.g. Bialystok, 1999). Abutalebi et al. (2011) show that early bilinguals use fewer brain resources for conflict monitoring, causing changes in brain structure in a region used for language control and for the resolution of non-verbal conflict (see Abutalebi & Rietbergen, 2014 for a review). Bilingual experience has been argued to enhance structural connectivity, thereby contributing to the constitution of a cognitive reserve protective against decline in old age (Luk, Bialystok, Craik, & Grady, 2011).

However, the bilingual literature features a discrepancy between studies that found an executive function advantage for bilinguals and those that have not, using seemingly comparable methodologies. This contribution focuses specifically on children, for whom the bilingual advantage appears even more difficult to pin down than for adults (see Valian, 2015 for a review). It explores alternative analytical methods that allow disentangling the effect of bilingual experience from other factors influencing executive function (“thinking multifactorially from the outset” — Titone, Pivneva, Sheikh, Webb, & Whitford, 2015, 44), using a gradient, composite measure of bilingual experience.

Updating is standardly operationalized as working memory (WM), which is the ability to retain information while performing mental operations (Baddeley, 2000). Out of the three EF components identified above, it is the one for which a bilingual advantage has been the most inconsistent. On the one hand, a few studies do report a bilingual advantage in terms of WM. Bilingual 5- to 7-year-olds have been shown to outperform monolingual controls on tasks requiring different levels of working memory (Morales, Calvo, & Bialystok, 2013). The advantage has also been shown to extend to sequential bilingual children of similar age but low socio-economic status (SES), if SES and vocabulary were controlled (Blom, Küntay, Messer, Verhagen, & Leseman, 2014) — although the advantage was only found in the 6-year-old group, not the 5-year-old group. On the other hand, a bilingual

advantage was not found by Engel De Abreu (2011) in a study comparable to that of Morales et al. (2013), involving 5- to 6-year-old children of similar SES and using similar (visuo-spatial) tasks.

Shifting concerns the ability to shift between mental sets or tasks (Miyake et al., 2000). It is typically investigated with card sorting tasks, in which the sorting has to be done according to different dimensions (e.g., shape and color) and the relevant dimension changes across trials, with dimensional change patterns varying across blocks of trials (i.e., single-task vs. mixed-task blocks). Timed implementations of such task-switching paradigms allow the evaluation of switching costs and mixing costs. Switching costs correspond to longer reaction times in switching trials compared with repetition trials in the mixed-tasks blocks (Monsell, 2003). Mixing costs correspond to longer reaction times in repetition trials in the mixed-task blocks compared with trials in the single-task blocks (Los, 1996). Switching costs are assumed to reflect transient control mechanisms, whereas mixing costs are assumed to reflect global sustained control mechanisms. Bilingual children have been shown to perform significantly better than monolingual controls in the Dimension Change Card Sort task (described in detail below). The advantage was observed in highly bilingual 3- to 6-year-olds, controlling for vocabulary and working memory (Bialystok, 1999; Bialystok & Martin, 2004), and in highly bilingual 5- to 7-year-olds (Carlson & Meltzoff, 2008), controlling for vocabulary, and in spite of lower socio-economic status of the bilingual group. By contrast to these positive findings, Mueller Gathercole et al. (2014) did not find a consistent bilingual advantage in their large-scale study involving 650 participants from 7 age groups who “grew up in the same context” (p.3) in North Wales. Participants were classified according to language dominance, which was determined on the basis of the patterns of language use in their home (Only Welsh Homes; Welsh and English Homes; English Only Homes). Card sorting tasks of age-adapted complexity level did not reveal an overall bilingual advantage in terms of accuracy or reaction time. An analysis based on difference scores only revealed a possible bilingual advantage in the teen group from Only Welsh Homes compared with their peers. In terms of absolute scores, the same group showed inferior accuracy compared with their peers, but faster reaction times.

Inhibitory control is the ability to resist interference (Dempster, 1992). This EF component has received the most attention in the literature on bilingualism (see Hilchey & Klein, 2011 for an in-depth, critical review). This was sparked by a seminal paper by Green (1998), which argued that language control in bilinguals resembles the control of action and proposed the Inhibition Control model to account for the control of language in bilinguals specifically. This model assumes that inhibitory control is required to suppress the irrelevant language (in a given task). Subsequent research has shown that the suppression of a bilingual’s dominant language was more costly than that required to suppress their weaker language (Meuter & Allport, 1999). Under the assumption that the constant need to control their languages exercises bilinguals’ domain-general inhibitory control abilities, it was then hypothesized that bilinguals would show an advantage over monolinguals in non-verbal tasks probing inhibition. This is standardly assessed using the Simon task (see below) or the flanker task (Eriksen & Eriksen, 1974). The incongruent condition in both types of task requires participants to suppress irrelevant visual information (such as the position of a colored block in the Simon task, or arrows pointing in the opposite direction to the target arrow in the flanker task). The inhibitory control hypothesis predicts that bilinguals

will show an advantage in the incongruent condition only. The so-called Simon effect is the difference in response speed between the incongruent condition and the congruent condition. Bilinguals are expected to experience a smaller Simon effect than monolinguals as they are better at controlling the conflicting information.

The Simon effect has hardly ever been reported in children. An exception is Poarch and van Hell (2012), who found a significant Simon effect in trilingual 5- to 7-year-olds, and a marginally significant Simon effect in bilingual 5- to 7-year-olds, but not in the monolingual controls. A number of studies in which no Simon effect was observed nonetheless report a bilingual advantage in terms of reduced reaction times in *both* congruent and incongruent conditions, so long as the task was sufficiently demanding and included response competition. This global RT advantage has been found in five-year-olds (Bialystok, Martin, & Viswanathan, 2005; Martin-Rhee & Bialystok, 2008). A similar global RT advantage was also observed in the flanker task in 8-year-old bilinguals (Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012) and in 4-year-old bilinguals compared with three monolingual control groups based on culture (i.e. imposing self-regulation of behavior or not) (Yang, Yang, & Lust, 2011).

By contrast to the positive results reported above, Morton and Harper (2007) found no bilingual advantage in the Simon task in a group of 6- to 7-year-olds. They attribute this lack of difference to the control for socio-economic status. In a design that carefully controlled for age, non-verbal IQ, SES (indexed by maternal education), receptive vocabulary, receptive morphosyntax, and short-term memory, Namazi and Thordardottir (2010) also found no bilingual advantage in the Simon task in terms of accuracy or speed. Instead, both were predicted by the children's working memory scores on a pattern recall task. A number of large-scale studies using a variety of tasks probing inhibitory control did not find a bilingual advantage either (e.g. Duñabeitia et al., 2015; Antón et al., 2014 — see Paap, 2014 for an in-depth review).

It has been claimed that the bilingual advantage in inhibitory control is particularly elusive in young children (Hilchey & Klein, 2011). In their meta-analysis of 39 studies, Donnelly, Brooks, and Homer (2015) ascertained the existence of a moderate but statistically significant bilingual advantage overall on global reaction time and interference costs from conflict resolution tasks. However, different effects emerged at different ages, and in children, larger effect sizes for global reaction times were observed, compared with interference costs.

In the 34 studies they review, Paap, Johnson, and Sawi (2014) observe that the bilingual advantage tends to be found in studies with small sample sizes and argue that the lack of a significant effect in large-sample studies coupled with significant effects in small-sample studies is worrisome (as it is unexpected given the properties of null hypothesis statistical testing). Donnelly et al. (2015) also reveal a significant lab effect across studies, which they speculate could be due to socio-linguistic differences of subject pools or different data-preparation methods such as outlier removal.

How should one interpret the contradictory findings regarding the hypothesized bilingual advantage in executive function abilities? The extent of the lack of replicability of findings is not known, given that null results tend not to be published (Adesope, Lavin, Thompson, & Ungerleider, 2010; de Bruin, Treccani, & Della Sala, 2015). It is therefore impossible to draw a conclusion based on the balance between results and null results. Matters are complicated further by the methodological differences across studies, and in particular

by differences in how possible confounds are taken into account (Paap & Sawi, 2014).

Socio-economic status (SES) might have acted as a confound in some studies (as suggested by Hilchey & Klein, 2011). Socially advantaged 4- to 7-year-old children have been shown to perform faster than their age-matched peers in terms of accuracy and speed in a flanker task (Mezzacappa, 2004), and parental level of education has been shown to predict children's development of executive functions (Ardila, Rosselli, Matute, & Guajardo, 2005; Merz et al., 2015). A number of studies controlling for SES indeed found no bilingual advantage (e.g. Morton & Harper, 2007; Paap & Greenberg, 2013; Duñabeitia et al., 2015). However, evidence also exists of a bilingual advantage in socio-economically deprived children (Engel de Abreu et al., 2012; Calvo & Bialystok, 2014).

Differences between children and adults are to be expected, given the protracted development of executive functions. Important stages of executive function maturation have been identified in the preschool years (Hongwanishkul, Happaney, Lee, & Zelazo, 2005) and between the ages of 6 and 13 (Brocki & Bohlin, 2004). For instance, this might contribute to explaining the paucity of evidence for the Simon effect in children but not in older adults (e.g., Bialystok et al., 2005).

Various types of experience that are not usually controlled for in experiments probing executive function performance also have a significant effect. Acute physical activity, for instance, has been shown to affect executive function performance positively in 6- to 10-year-olds (Best, 2012). Musical training (Moreno et al., 2011) and physical exercise (Diamond & Lee, 2011) have also been shown to have a positive effect. Psycho-emotional factors such as lack of sleep, stress and loneliness also impact executive function performance, in this case negatively (Diamond, 2011).

Most importantly, the level of bilingualism of the populations tested is not reported consistently across studies. This is partly due to the lack of a general consensus as to how bilingual experience should be measured (see e.g. Bedore et al., 2012 on assessing language dominance). Several aspects known to vary in bilingual experience have been argued to have an effect on executive function performance, such as proficiency (Gutiérrez-Clellen, Calderón, & Ellis Weismer, 2004; Iluz-Cohen & Armon-Lotem, 2013; Rosselli, Ardila, Lalwani, & Vélez-Urbe, 2015), language dominance (Mueller Gathercole et al., 2014; Weber, Johnson, Riccio, & Liew, 2015), language exposure (Brito, Sebastián-Gallés, & Barr, 2015), and patterns of language switching (de Bruin et al., 2015; Verreyt, Woumans, Vandelandotte, Szmalec, & Duyck, 2015; Scaltritti, Peressotti, & Miozzo, 2015; Soveri, Rodríguez-Fornells, & Laine, 2011). A number of researchers are now calling for the integration of the variability of bilingual experience in the analyses (e.g. Kaushanskaya & Prior, 2015), to account for the whole spectrum of language experience, from monolingualism to the highest levels of bilingualism (Luk, 2015). Our study takes up this challenge.

Aims of the present study

This study aims to investigate the executive function correlates of childhood bilingualism in a highly heterogeneous sample in terms of bilingual experience and socio-economic status. We sought to embrace the heterogeneity of bilingual experience rather than control for it as an inclusion criterion; this was achieved by using a gradient measure of bilingual experience that encompasses both exposure (input) and production (output). We also sought to disentangle the effect of bilingualism from that of other factors (such as socio-economic

	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 = 42)	5;0	7;0	6;0	0;7

Table 1

Distribution in gender and age (in years;months)

status, proficiency in English, short-term memory, working memory, and gender); this was achieved by incorporating them into the analysis. In an attempt to achieve maximum comparability with other studies, we relied on widely-used tasks.

A key aim is to explore and exploit new methods of analysis to better achieve the objectives listed above. On the one hand, we exploit linear regression to model the additive effect of all the predictors, thereby isolating the specific effect of bilingual experience (or “adjusting” its effect in light of other significant factors). On the other hand, we exploit the Cox Proportional Hazard model to capture accuracy and speed within a single analysis of performance in the Simon task, while exploring the additive effect of all predictors. Two additional advantages of the Cox PH model are (i) that it does not require the exclusion of any observation, and (ii) that it is able to take into account the auto-correlation between trials.

The use of modeling techniques and the operationalization of bilingualism as a continuum allow us to identify the threshold from which the performance of bilingual children differs from that of monolingual children. In that way, we aim to contribute to bridging the gap between studies identifying a bilingual advantages and studies reporting null results, thereby starting to answer the question as to how bilingual one needs to be to benefit from a cognitive advantage.

Methods

Participants

We recruited 174 children attending year 1 or 2 of primary school in the North of England, including 87 monolinguals and 87 children with some amount of exposure to a home language other than English. For convenience, we will refer to these groups as the “monolinguals” and the “bilinguals” respectively — even though the language experience of children at the lower end of the bilingualism spectrum is hardly distinguishable from that of monolinguals. Forty one children were sequential bilinguals.¹ Table 1 summarizes the distribution of the two groups in gender and age. All children were developing typically and did not have any known hearing deficit.

The language of schooling was English for all the children. The bilingual children were also exposed to another language (henceforth the *home language*) in varying degrees (see below). There was a total of 28 home languages in our sample:² Arabic (9%), Bengali,

¹There is a significant correlation between age of initial bilingual exposure and BPI score (see De Cat and Serratrice (to appear) for a model of how the BPI relates with various estimates of bilingual experience).

²Percentages are given for those languages representing more than 5% of the total sample.

	Secondary school	Further education	University
High	0	13	61
Mid	14	16	8
Low	23	30	7

Table 2

SES measures (reduced to 3 levels): education by occupation. Correspondence with NS-SEC categories listed in Table 14: High < 7, Low ≥ 10.

Cantonese, Catalan, Dutch, Farsi, French (8%), Greek, Hindi, Italian, Kurdish, Mandarin, Marathi, Mirpuri, Nepalese, Pashto, Polish, Portuguese, Punjabi (21%), Shona, Somalian, Spanish (6%), Swedish, Tamil, Telugu, Thai, Tigrinya, and Urdu (17%).

Socio-economic profiles

The socio-economic status of the children’s families was estimated on the basis of information gathered via a parental questionnaire. Two measures were obtained: one for parental level of education (1), 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).

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

The occupational data were scored using the reduced method of the National Statistics Socio-economic Classification (simplified NS-SEC — see Table 14 in the Appendix), which is based on the Goldthorpe Scheme of sociological classification. For each family, the highest current occupation was selected, as recommended by the NS-SEC guidelines, and entered into a 48-point ranking scale. The score obtained was reversed for ease of interpretability (being positively correlated with educational level).

The proxy measures for socio-economic status are strongly and significantly associated, as shown in Table 2 ($\chi^2(4, N = 174) = 83.57, p < 0.0001$). We subsequently used the simplified NS-SES measures in all our analyses as it afforded the possibility of capturing SES as a continuous variable (based on the 48-point scale).

Quantifying bilingual experience

To quantify bilingual experience as a continuum, we used the Bilingualism Profile Index (De Cat & Serratrice, to appear). The BPI encapsulates children’s cumulative exposure to an “additional” language (to the language of schooling) as well as their active (cumulative) use of that language.³ Because English is the language of schooling, we assumed it was (or had become) a language actively used by all children. Greater variability

³The measures of cumulative input and output were derived from information collected via parental questionnaires. Full details of the procedure and calculations are provided in De Cat and Serratrice (to appear).

was therefore expected in the exposure to and use of the home language (which, for many children, would be the weaker language). The BPI intends to apprehend that variability of bilingual experience, and can be interpreted as a gradient measure of language dominance (De Cat & Serratrice, to appear).

The BPI is a composite score of cumulative input and cumulative output in the home language, derived by a standard measure of dimensionality reduction, i.e. Principal Component Analysis. The PCA of cumulative input and cumulative output yields two principal components, the first of which captures most of the variability (given the strength of the correlation between the two cumulative measures). In our data, the first component captured 98% of the variability, so it can be interpreted as a reliable measure combining the two cumulative (input and output) measures. The BPI scores correspond to the loadings of that first component.⁴ It can be interpreted as a measure of the child's experience of their home language (with a score of 0 indicating monolingualism).

Figure 1 shows the correlation of the BPI with current and cumulative measures of input and output in the home language. The amount of dispersion from the linear relationship between the BPI and each of the cumulative measures shows that it is not reducible to either of these measures. The BPI is also very strongly correlated with the age of onset of exposure to an additional language ($r = .85, p < 0.0001$). This correlation is much stronger than that between current input and onset of exposure to an additional language ($r = .51, p < 0.0001$).

In De Cat and Serratrice (to appear) we provide independent validation for the BPI by predicting the BPI scores (via linear regression) from factors not used to derive the BPI but standardly assumed to predict levels of bilingualism, such as home environment (e.g. one-parent one-language), bilingualism onset and length of exposure. We also demonstrate that, although based on estimates of experience in the home language, the BPI predicts proficiency in the language of schooling (English).

In summary, our sample ranges across an evenly distributed continuum of bilingual language experience, as indexed by the BPI score. 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 home language until their first significant exposure to English at primary school. Current exposure to a language other than English ranges from 9% to 89% in our sample.

A weak but significant negative correlation was observed between the level of bilingualism and socio-economic status (based on the occupational classification): $r = -.25, p = 0.0009$.

Proficiency

The executive function tasks used in this study did not rely on language production from the participants (except for the use of numbers in the memory tasks), but the instructions were given in English. It is therefore possible that children's proficiency in English could have an impact. For that reason, we assessed their global proficiency with a sentence

⁴Loadings are the linear combinations of coefficients that correspond to the covariances/correlations between the original variables and the unit-scaled components. We reversed their sign for ease of interpretation (i.e. a positive rather than negative correlation between the BPI and the amount of experience in the home language).

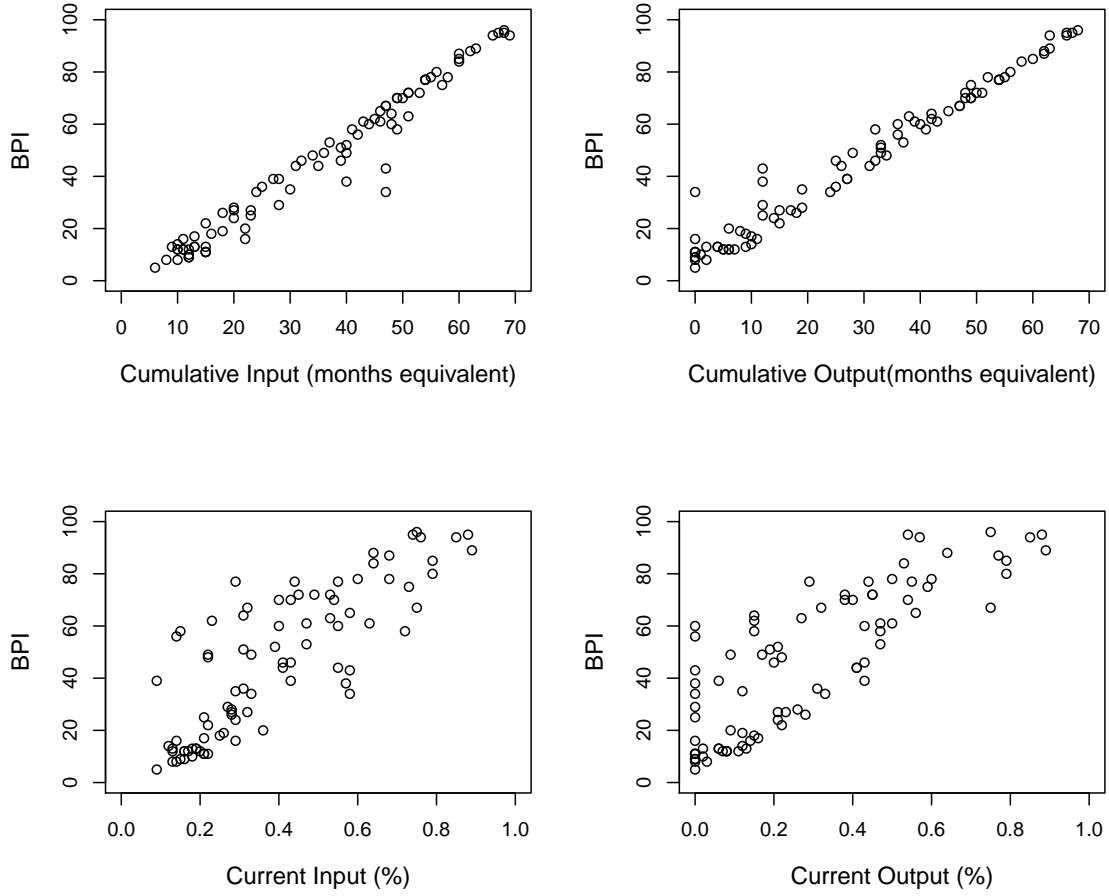


Figure 1. The Bilingual Profile Index and its relationship with input and output measures in the home language

repetition task (the short version of the School-Age Sentence Imitation Test — Marinis, Chiat, Armon-Lotem, Gibbons, & Gipps, 2010), which was originally designed to identify children at risk of Specific Language Impairment in bilingual populations.⁵ The distribution of scores is shown in Table 3. See De Cat and Serratrice (to appear) for analysis of this and other proficiency tests in the present sample of participants.

Measures of executive function abilities

Updating. Measures of short-term and working memory 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

⁵The SASIT was subsequently renamed as the LITMUS test (Marinis & Armon-Lotem, 2015).

	Group	Mean	St.dev
1	Mono	0.8077	0.1688
2	Bi	0.6540	0.2618

Table 3

Distribution of scores in the sentence repetition task, by group

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 were four trials per digit span.

The Forward Digit Span measure was used as a proxy for children’s episodic buffer capacity (i.e. short term memory). The Backward Digit Span measure was used as a proxy for children’s updating abilities (i.e. working memory).

Shifting. To assess children’s flexible use of rules to govern behavior, we administered the Dimensional Change Card Sort task. 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 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.

Inhibition. Inhibition was assessed with computer-based version of the Simon task (Simon, 1969), programmed and run via E-Prime software. This task manipulates the compatibility of stimulus and response according to two dimensions: color and position. In order to perform accurately on critical trials, participants have to inhibit the prepotent association of to irrelevant position information and focus on color information exclusively. Children were sat in front of a computer (Toshiba Satellite L-855 laptop with a 15.6” screen) with an E-prime serial response button box featuring a green button and a red button.⁶ They were asked to press the button whose color matched that of a square as fast as possible after it appeared on the screen. In the matching condition, the color of the square was aligned with the position of the button (i.e. left for red, right for green). In the mismatching condition, the square was aligned with the button with the wrong color. Children were asked to respond as fast as possible, and their reaction time was recorded. There was no neutral condition (in which the colored square would appear centrally). Children started with an 8-trial practice session, immediately followed by 48 test trials. Comparison with other protocols is summarized in Table 4.

⁶All children had previously been screened for color blindness using the Ishihara Color Test (Ishihara, 1994).

Study	Training	Test
Bialystok, Craik, Klein, and Viswanathan (2004)	8	28
Morton and Harper (2007)	2	28
Martin-Rhee and Bialystok (2008)	8	40
Paap and Greenberg (2013)	20	40
(Poarch & van Hell, 2012)	24	84 ⁷
our study	8	48

Table 4

Number of critical trials in the Simon task: cross-study comparison

Analysis

For the results of the working memory task and the switching task (DCCS), we fitted linear models using packages `lme4` (version 1.1.11) and `ordinal` (2015.6.28) in R (3.2.4). The models were built by adding factors incrementally; factors were retained only if they improved the fit of the model, yielding a significant reduction in AIC⁸ and a significant R-squared value for the model. This allowed the data to determine which variables and which interactions were justified in the model, through systematic testing. The same procedure of incremental model building was adopted in the Cox Proportional Hazard analysis of the Simon task results (which is explained in detail in the relevant section below). All model summary tables report the statistics for the optimal model. Statistics for each non-significant variable or interaction of interest were computed by adding it back to the optimal model. These are reported in the text only.

The following variables were tested for inclusion in all models:⁹ *BPI*, *SES* (occupational score), *Short Term Memory* (forward digit recall score), *Proficiency* (Sentence Repetition score), *Age* (in months), *Gender*, as well as interactions between these variables. Additional, task-specific predictors were also considered, as explained in the relevant sections below. *Age* was centered in all analyses to make the value at the intercept interpretable (i.e. not zero months but the average age) and mitigate its correlation with other variables.

Results

Updating

Three (bilingual) children failed the working memory test: they did not manage to reverse even two digits. One of them also failed the DCCS test. The three children were excluded from the analysis of the memory measures. The raw results prior to this exclusion are summarized in Tables 5 and 6.

To account for the fact that memory increases with age, we divided the forward digit span score and the backward digit span score by age (in months).¹⁰ This turned them into

⁷In addition to these critical trials, there were 42 items in a neutral condition, in which the colored square was presented centrally.

⁸The Akaike Information Criterion is a measure of fit that penalizes a model for including an excessive number of variables.

⁹Variable names will be italicised for clarity.

¹⁰Without this transformation, the model reported in Table 7 would not be able to distinguish the effect of *STM* from the effect of *Age* as a confounder. Dividing both memory measures by *Age* allows us to

Digits	3	4	5	6	7
Children (N)	22	68	72	10	2

Table 5

Forward Digit Recall (digits correctly recalled)

Digits	0	2	3	4
Children (N)	3	66	98	7

Table 6

Backward Digit Recall (digits correctly reversed)

continuous measures capturing *STM* and *WM* regardless of *Age* (assuming that increment in *STM* and *WM* is linear with *Age*). These new measures were used as dependent variables in the linear regression analyses reported below.

As shown in the optimal model reported in Table 7, the only significant predictor of the working memory score was *STM* ($t(181) = 54.15, p < 0.0001$). Neither *BPI* ($t(181) = -0.29, p = 0.77$) nor *SES* ($t(181) = -0.40, p = 0.69$) reached significance, and there were no significant interactions. The model accounted for 94% of the variance ($R^2 = 0.94$).

As the strongest predictor of working memory, the short-term memory (*STM*) performance is interesting to model in its own right. It was found to be predicted by socio-economic status ($t(181) = 2.94, p = 0.004$) and gender, with boys performing less well than girls ($t(181) = -2.18, p = 0.03$), as shown in Table 8. *BPI* was not a significant predictor ($t(181) = 0.80, p = 0.42$). The model accounted for 6% of the variance ($R^2 = 0.06$).

Shifting

We adopted the method recommended by Zelazo (2006) to analyze the DCCS data. It assigns a score between 0 and 3 based on a pass-fail criterion on each trial block. Passing the first two blocks requires sorting at least 5 out of 6 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 pre-switch block, a score of 1 if they pass the pre-switch block only, a score of 2 if they also pass the post-switch block, and a score of 3 if they pass all three blocks.

Little variability is observed, as shown in Table 9, as many children performed at ceiling. Only one (bilingual) child did not pass the first block, and was excluded from analysis (as he was not able to perform the task).

carefully include *STM* as predictor in the WM model, making it possible to interpret parameter estimates independently. See the Supplementary Online Material for a more detailed explanation.

	Coefficient	Std.Error	t-value	p
Intercept	-0.0440	0.1150	-0.3823	0.7027
Forward Digit Recall / Age (months)	0.5315	0.0098	54.1544	0.0000

Table 7

Coefficients of the linear regression model fitted to the Backward Digit Recall score divided by Age

	Coefficient	Std.Error	t-value	p
Intercept	0.0698	0.0021	32.6650	0.0000
SES (occupation)	0.0007	0.0002	2.9444	0.0037
Gender	-0.0037	0.0017	-2.1794	0.0307

Table 8

Coefficients of the optimal linear regression model fitted to the Forward Digit Recall scores divided by Age

DCCS score	0	1	2	3
Participants	1% (N=1)	8% (N=14)	37% (N=65)	54% (N=94)

Table 9

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

We fitted an ordinal Cumulative Link Model to the DCCS overall 0-3 score (to capture the order of the dependent variable's levels). The optimal model is summarized in Table 10. Performance in the DCCS was predicted by (i) the child's socio-economic status *in bilingual children only* ($Z = 1.98, p < 0.05$), (ii) the age of the child ($Z = 3.58, p < 0.001$), (iii) the child's proficiency in English ($Z = 4.02, p < 0.001$). *Working memory* approached significance ($Z = 1.65, p = 0.10$). *BPI* did not reach significance ($Z = 0.38, p = 0.70$).¹¹ Importantly, in spite of a significant correlation between *BPI* and *Proficiency* score ($r = -.49, p < 0.001$), *BPI* does not turn out as a significant predictor even in the absence of *Proficiency*, or in interaction with it.

	Coefficient	Std.Error	z value	p
1 2	6.5818	2.0627	3.1908	0.0014
2 3	9.9425	2.1960	4.5275	0.0000
Age (months)	0.1119	0.0313	3.5809	0.0003
Proficiency (SASIT)	9.0021	2.2377	4.0229	0.0001
Working memory	0.4014	0.2426	1.6550	0.0979
SES (occupation)	-0.0766	0.0813	-0.9422	0.3461
Bilingual:yes	0.8906	0.8241	1.0807	0.2798
SES x Bilingual:yes	0.2051	0.1036	1.9808	0.0476

Table 10

Coefficients of an ordinal Cumulative Link Model fitted to the DCCS overall score. Reference level: Bilingual=no.

Inhibition

As expected, incongruent trials in the Simon task resulted in lower accuracy in average (as shown in Table 11), and slower reaction time (shown in Table 12 for accurate responses only).

¹¹An alternative analysis on the cumulative raw score (including the 3 blocks of trials), based on a Generalized Linear Model with Poisson distribution, showed a similar pattern of results, with a significant effect of working memory. The BPI remained non-significant.

Condition	Accuracy mean	Accuracy st.dev
congruent	0.9638	0.1868
incongruent	0.8998	0.3004

Table 11

Accuracy of responses in the Simon task

Condition	RT mean	RT st.dev
congruent	1116.9572	772.4815
incongruent	1214.2377	624.8028

Table 12

Speed of accurate responses in the Simon task

Issues with traditional analyses. Reaction time data is known to feature a positively skewed distribution, as the scale is bounded on the left (i.e. one cannot respond faster than x number of milliseconds) but not to the right (i.e. response can be delayed for a variety of reasons). Performance in the Simon Task is usually analyzed by comparing mean reaction times across conditions (congruent vs. incongruent) and across groups (e.g. bilingual vs. monolingual) — as done in all the relevant studies cited above. This assumes that the data is normally distributed (which it isn’t, as explained above). The traditional solution is to remove observations further than 2 standard deviations from the mean (for each participant).

In addition to this, auto-correlation between trials is usually not taken into account by traditional analyses. On the one hand, self-monitoring is likely to have an effect, whereby the participant might slow down after noticing they had answered incorrectly on a particular trial. On the other hand, the trials form part of a time series, and there might be an effect of e.g. habituation or tiredness. Removing the trial immediately following an erroneous response (as done by e.g. Poarch & van Hell, 2012) results in further loss of data, and does not take into account the time-series effect.

Finally, and most importantly, traditional approaches cannot apprehend accuracy and response time within the same analysis. Any trial eliciting an incorrect response is removed from the reaction time analysis. If a group is less accurate, the number of data points in the reaction time analysis will be reduced, but the analysis cannot take this disparity into account. In addition to this, the information carried by the incorrect trials is lost.

The combination of these three issues results in the removal of informative data, which means that the estimates of the model parameters are heavily biased and cannot really estimate the truth. This jeopardizes the interpretation of the parameter estimates.¹²

A solution: Cox Proportional Hazard Model. To address these concerns, we propose the use of a Cox Proportional Hazard (PH) Model to analyze children’s responses to the Simon Task. We start by describing how this type of model allows us to account for

¹²If observations are systematically removed based on response accuracy, this results in a bias of the estimates: in the Simon task, response accuracy is significantly associated with the BPI and with (shorter) reaction times (as confirmed by logistic regression analysis).

the difference between efficient and less efficient inhibitors before explaining the technique in detail.

Imagine a “good” inhibitor takes x amount of time to answer correctly on a particular trial in the incongruent condition. A “bad” inhibitor is expected to either take longer to answer correctly, or possibly the same amount of time (or shorter) but with a likely incorrect response. It is most unlikely that the bad inhibitor would be able to respond correctly in that trial in a shorter amount of time than the good inhibitor.¹³ The Cox PH model enables us to capture this by including the time to an incorrect response as a *censored* observation (as explained below). Censored observations are interpreted as the *minimum* amount of time it would have taken to produce a *correct* response in that trial (even though we do not know how long it would actually have taken). In that way, all responses are included in the analysis (i.e., both the time taken to answer correctly, and the time taken to answer incorrectly), while being interpreted differently.

The Cox PH technique (also known as Cox regression in the context of survival analysis) is commonly employed to model time-to-event data (Armitage, Berry, & Matthews, 2008; Collett, 2015). In the case of the Simon Task, the ‘event’ is defined as ‘*correct* response to a trial’ (as opposed to an incorrect response). This definition has two consequences for how the data is treated. First, the time span from stimulus presentation to correct registered response is treated as an *uncensored* observation, in the sense that the time to end-point (i.e. correct response) is known/observed. Second, when the children gave a wrong response, the time leading to it is treated as *censored*, as the end point (i.e. correct response) is not observed. Censored observations are taken into account in the model as they carry valuable information: they indicate that the amount of time to a correct response would have taken *at least* as long as that of the censored observation. Ignoring that information would result in biased estimates. The Cox PH model takes into account the censored times because it does not model the response time directly, but rather the *hazard rate*, which is defined as the probability (known as the ‘(instantaneous) hazard’) of responding correctly at any particular time, given the covariates or predictors. The hazard rate is modeled as a function of *baseline* hazard and covariates as predictors. The baseline hazard can be interpreted as a ‘typical’ rate of correct response over time (independent of the predictors). The effect of covariates is to modify the baseline hazard multiplicatively. This means that the baseline ‘hazard’ function effectively serves as the intercept (as in linear regression models).

In the context of the Cox PH model, longer reaction times are not considered as outliers (indeed, they are expected) and therefore are included in the analysis. As a consequence, it is not necessary to log-transform reaction times prior to analysis.

To account for the auto-correlation between trials due to self-monitoring, we included accuracy at the previous trial as a dummy-coded factor. Furthermore, the observations between items are expected to be correlated as they are taken in sequence. To take this correlation into account, we consider *Item* as a random effect in the Cox PH model. This allows the model to take into account the excess ‘risk’ or ‘*frailty*’ for *Item*, over any measured covariates (Therneau, Grambsch, & Pankratz, 2003): a random intercept is calculated for each item. The normality assumption shrinks the random effects estimates towards

¹³In our study, accurate responses took on average 204 ms longer than inaccurate responses: accurate responses: mean = 1164 ms (SD = 707 ms); inaccurate responses: mean = 960 ms (SD = 828 ms). This fits perfectly with the assumptions of the Cox PH model.

the population mean (Lee, Nelder, & Pawitan, 2006). Hence, as an indication that the correlation between items is taken into account, we expect (and, in the present case, find) a reduction in the degrees of freedom of fit in random effects relative to the case in which we would include *Item* as fixed effects.

Participant could not be included as a random effect because individual variation is accounted for by the fixed effects in this model. Participant variability is collinear with participant-related variables such as *BPI*, *SES* etc,¹⁴ as shown by the following diagnostic: when we include *Participant* as a fixed effect to check its impact on the other variables in the predictors,¹⁵ collinearity with the other variables in the model makes some parameters unestimable. Furthermore, excluding *Participant* from the random effect structure is sound in terms of modelling, as confirmed by the following diagnostic: when we include *Participant* as a random effect, the reduction in the degrees of freedom is minimal (only 8%), indicating that the (random) effects due to individual participants are relatively independent (e.g. in terms of background or genetic profile). This shows that there is no participant-induced structure that has to be accounted for by the model in addition to the baseline hazard function and their covariates as modifying factors.¹⁶

Implementation of the Cox PH regression analysis is explained in the Appendix.

Results of the Cox PH regression analysis. The results of the optimal Cox PH model for our data are summarized in Table 13. As expected, performance was poorer in the incongruent condition ($\chi^2(1) = 117.54, p < 0.0001$). The strongest predictor of performance was *Age* ($\chi^2(1) = 813.92, p < 0.0001$), followed by self-monitoring — resulting in better performance in a given trial if the response to the previous trial had been accurate ($\chi^2(1) = 119.81, p < 0.0001$). Socio-economic status (*SES*) had the opposite effect in monolingual children vs. bilingual children as a group: whereas higher *SES* seemed to yield marginally poorer performance in monolingual children ($\chi^2(1) = 4.48, p = 0.03$), it conferred a pronounced advantage in bilingual children, when considered as a group ($\chi^2(1) = 40.60, p < 0.0001$). An alternative interpretation (suggested by an anonymous reviewer) is that as *SES* increases, a bilingual difference emerges. There was also a trend for the bilingual group as a whole to do better than the monolingual group ($\chi^2(1) = 3.8, p = 0.05$). Over and above that effect of group, the amount of bilingual experience conferred an overall advantage ($\chi^2(1) = 12.13, p = 0.0005$). That advantage was not significantly greater in the incongruent condition ($\chi^2(1) = 0.07, p = 0.78$) — i.e., bilingualism did not give rise to the so-called “Simon effect” given the absence of a significant interaction between *BPI* and Condition.¹⁷

As shown in Figure 2, participants tended to slow down as the experiment progressed. Figure 2 presents the random effects estimates for items 1 to 48 in the experiment. Considering the sequence of items, the figure indicates that the estimates have a significant downward linear trend ($p = 0.013$). Considering that the random effects estimates modify the baseline hazard, a lower estimate corresponds to a longer time in responding correctly

¹⁴The systematic between-subject variability in our data is *captured* by participant-related variables, resulting in a model of more scientific value than an alternative model with *Participant* as a random effect (as this would have to exclude participant-related variables).

¹⁵This is relatively safe to do, considering that the participants are randomly selected.

¹⁶See (Hougaard, 2000) for a demonstration of how the standard Cox PH model can capture individual-induced variability in that way, without the need for including *Participant* as a random effect.

¹⁷There was no significant interaction between Condition and Group (i.e., monolingual vs. bilingual) either.

	Coefficient	Coefficient (scaled)	Std.Error	Chi-sq	DF	p
Condition: incongruent	-0.2470	-0.2471	0.0228	117.5452	1.0000	0.0000
Age (months)	0.0520	0.3360	0.0018	813.9154	1.0000	0.0000
Accuracy at previous trial	0.5243	0.5243	0.0479	119.8091	1.0000	0.0000
BPI	0.0027	0.1026	0.0006	17.7801	1.0000	0.0000
SES (occupation)	-0.0100	-0.0839	0.0047	4.4800	1.0000	0.0342
Bilingual: yes	0.1151	0.1151	0.0589	3.8238	1.0000	0.0507
Gender	-0.0071	-0.0071	0.0233	0.0938	1.0000	0.7625
Item				7.8878	8.1732	0.4622
SES x Bilingual: yes	0.0415	0.3497	0.0065	40.5996	1.0000	0.0000

Table 13

Coefficients of the optimal Cox Proportional Hazard model fitted to the Time to Correct Response (with Item as random effect; reference levels: Condition = “congruent”, Accuracy at previous trial = “inaccurate”, Bilingual = “no”, Gender = “female”). Coefficient (scaled) are for the same model fitted using scaled predictors for BPI, SES and Age. The values in the other columns remain identical for both models.

to the Simon tasks.

Identifying a critical threshold of bilingual advantage

The use of a continuous variable indexing bilingualism makes it possible to identify the amount of bilingual experience¹⁸ above which children’s performance on the Simon task starts to differ significantly from that of monolingual children. This can be done by partitioning children in two groups (insufficiently bilingual vs. sufficiently bilingual) recursively at each point of the BPI scale, and fitting the optimal Cox PH model (reported in Table 13) using each binary partition for bilingualism in turn. The model with the lowest AIC value is the one with the best fit, and is therefore interpreted as the one with the optimal BPI partitioning score. The AIC value for each of the 98 resulting models is plotted in Figure 3. The model with the lowest AIC value was based on a binary bilingualism partitioning at BPI = 59.¹⁹

In our sample, 33 children had a BPI score of 59 or above, which means they were likely to benefit from an inhibitory advantage compared with monolinguals, *all other things being equal*. This represents 38% of the group of children who were exposed to more than one language. Their profiles are summarized in a table in the Supplementary Online Material.

It is important to stress that being above the BPI threshold is only one among many factors predicting enhanced performance in the Simon task. Figure 4 illustrates this by plotting the bilingual children’s modeled scores against three of the significant predictors in our model (*Age*, *SES* and *BPI*, taking care of scaling them beforehand),²⁰ highlighting

¹⁸Strictly speaking, the BPI indexes the amount of experience in the Home Language, ranging from 0 (English monolinguals) to 96.

¹⁹The model based on the optimal group partition does not yield a significant Simon effect either, but a general advantage (as in the present model).

²⁰The coefficients of the model fitted on the basis of the scaled variables are given in Table 13, column 3.

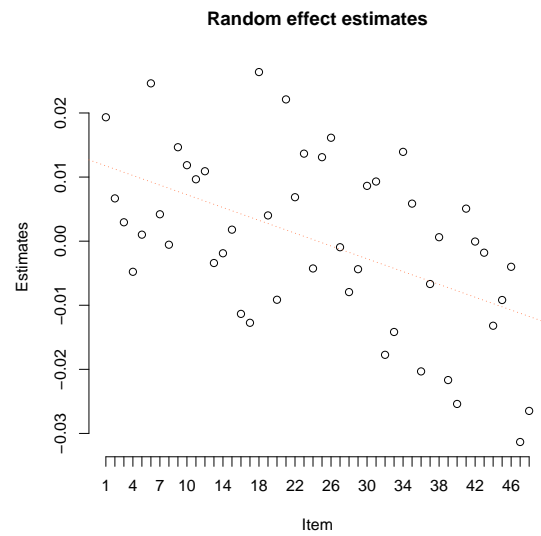


Figure 2. Random effect for item in the Cox PH model fitted to the Time to Correct Response

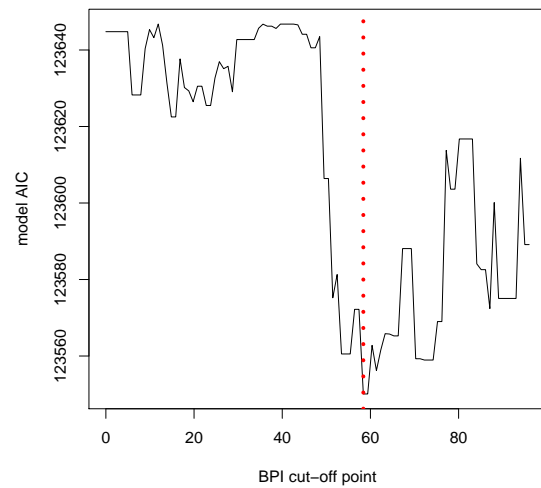


Figure 3. Model fit at each BPI cut-off point, identifying a threshold for the bilingual advantage

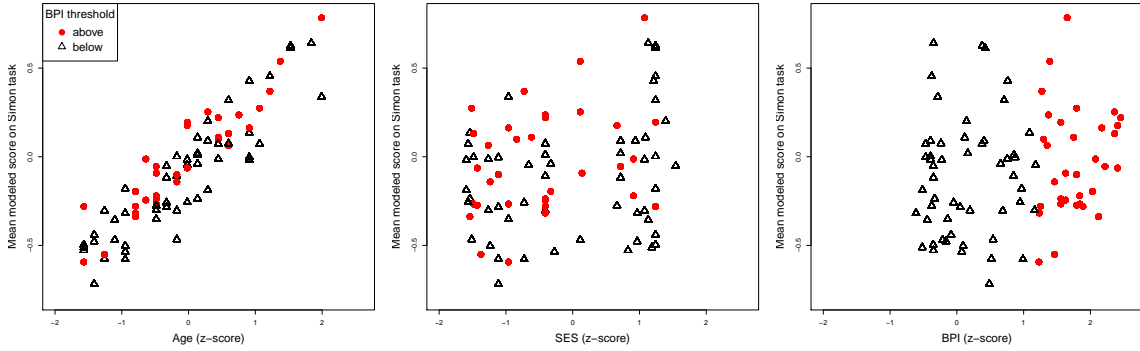


Figure 4. Bilingual children’s modeled scores in the Simon task (as estimated by the Cox regression model), according to scaled predictors for (i) *Age*, (ii) *Socio-Economic Status*, and (iii) bilingual experience (*BPI*). A higher value on the y-axis indicates better performance (i.e. more accurate and faster).

whether or not the children were above the critical BPI threshold. This plot also shows that *Age* is much more strongly correlated with a high score than *SES* or *BPI*.

Discussion

The results reveal a mixed picture regarding the impact of bilingual experience on executive function performance, as indexed by the three tasks in our study. This is not unexpected “because of the uncertainty about what components of executive functions different tasks tap, and uncertainty about the importance of task features that are orthogonal to executive function” (Valian, 2015, 9).

Results from the working memory tests showed very little variation, making them insufficiently discriminative to be able to detect the influence of likely predictors. Most of the variance in working memory was predicted by short term memory. Because of this, we suggest that the predictors of short term memory could be interpreted as indirect predictors of working memory. This includes *Age*, *SES* (conferring an advantage) and *Gender* (with girls outperforming boys). Given the lack of substantial variation, our study remains inconclusive as to the effect of bilingualism on updating skills. Other studies have reported a bilingual advantage at similar ages (Blom et al., 2014) or younger (Brito et al., 2015).

Although considered to be non-verbal, the DCCS switching task included constant verbal reminders of the sorting rules. This was intended to control for the effect of short-term memory, but there remained a trend for better performance in children with better short-term memory. Crucially, it appears the frequent verbal instructions induced an effect of language proficiency on performance. The amount of bilingual experience did however not have an effect (even if *Proficiency* was not included as a predictor). Instead, only the children in the (highly heterogeneous) bilingual group benefited from a socio-economic advantage. A limitation of our study is that reaction times were not recoded.

Using Cox PH regression to model the Simon task data allowed us to capture accuracy and response speed within the same analysis (i.e. not discarding inaccurate responses at any stage of the analysis), and to consider all the data points (i.e. including “outliers”)

without violating model assumptions. The retention of all the potentially informative data allows the model parameter estimates to be interpreted with confidence. All children were found to have taken progressively longer over the course of the experiment. Self-monitoring was evident in longer response time in the trial following an incorrect response. Bilingual children performed better than monolingual children, in the sense that they were faster at responding accurately. The BPI predicted better performance overall, over and above a significant group advantage. The advantage was a global one, as there was no significant Simon effect in favor of the bilinguals: no interaction was observed between bilingualism and condition (congruent vs. incongruent). However, a global advantage might still indicate superior inhibition abilities if performance in congruent trials is affected by the overall context of the task. Support for this interpretation comes from the fact that this global advantage is only found in tasks that are sufficiently demanding and include response competition (Martin-Rhee & Bialystok, 2008). Other studies have also reported a global advantage for bilingual children over their monolingual peers in the Simon task (Bialystok et al., 2005) or the flanker task (Engel de Abreu et al., 2012; Yang et al., 2011). The only study to date reporting a Simon effect in favor of bilingual children is Poarch and van Hell (2012). That study included twice as many trials in the Simon task compared with other children studies, and it included a neutral condition (in which the colored block appeared centrally). The effect of these aspects of the design and of the substantial quantity of discarded data²¹ will need to be scrutinized in future research.

Our analysis controlled for age, socio-economic status, gender, and self-monitoring (i.e. accuracy at the previous trial). This gives us confidence that the bilingual advantage observed is unlikely to be due to confounds arising from these particular factors.

The bilingual advantage

As called for by many (e.g. Abutalebi & Rietbergen, 2014; Baum & Titone, 2014; Kaushanskaya & Prior, 2015; Luk, 2015), we have integrated the variability of bilingual experience in our model, alongside other relevant factors, and this has allowed us to better capture the specific impact of bilingualism on executive function. The BPI (Bilingual Profile Index) indexes children’s cumulative experience in their home language, thereby taking variations in their bilingual history into account. It is derived from the combination of cumulative measures of input *and* output measures, and thus measures the extent to which a child actively uses their home language — a dimension which might require additional neuro-cognitive resources compared with language switching for comprehension only (see Abutalebi & Rietbergen, 2014). There is a strong and significant correlation or association between the BPI and each of the following measures: age of first exposure, current input, current output, and household type (defined by patterns of language use). Although derived from measures of exposure to the home language, the BPI has also been shown to predict proficiency in the language of schooling (De Cat & Serratrice, to appear). It is therefore a useful index of bilingual experience overall, which can be interpreted as varying quantitatively and qualitatively. As such, it makes it possible to capture *globally* the effect of bilingual experience on executive function performance. Our study therefore answers the

²¹In that study, a trial was discarded if (i) the response was incorrect, (ii) it immediately followed a trial with an incorrect response, or (iii) its reaction time was above or below two standard deviations from the mean.

question as to how much bilingual experience²² is enough for 5- to 7-year-olds to benefit from an inhibition advantage (Luk, 2015).

The bilingual advantage is hypothesized to arise from bilinguals' regulation of their two languages. The particular aspects of bilingual experience conferring a cognitive advantage are however not yet fully understood (Kroll & Fricke, 2014). Code-switching is likely to be a significant factor (Green & Abutalebi, 2013; Prior & Gollan, 2011), itself influenced by a number of other dimensions of the bilingual experience (Green, 2014; Genesee, 2014). Further research will be necessary to investigate this. In our study, bilingualism varied in terms of age of initial exposure (although this was below 5 years of age for all children), amount of exposure to a home language (ranging from 9% to 89% in terms of current input), amount of output produced by the child in their home language, household type (in terms of patterns of language use), and home language itself (from a range of 28). The opportunities for code-switching at school also varied, but unfortunately we did not have a reliable indicator to include in the analyses. In other respects, there was little variation: the language of schooling was English for all children, none of the children were in a bilingual education program. Although our approach does not allow us to ascertain if particular aspects of bilingual experience had significantly more impact, observing the characteristics of the children who were above the critical BPI threshold is a first step in that direction. The profile of these children is detailed in the Supplementary Materials. These children were exposed to their home language 58% of waking hours on average (the lowest was 23%), and were generally 'active' speakers of that language (with current output at 52% on average, and one child not producing any).²³ This suggests that on average, these children were balanced bilinguals. In terms of cumulative exposure, they had received the equivalent of 54 months exposure to their home language on average (minimum 43 months) and their cumulative output amounted to the equivalent of 52 months on average (minimum 36 months). Only one of them came from a one-parent-one-language household. Twenty-two came from a household where only the home language was reported to be used, and ten from a household in which both parents spoke the home language but children sometimes interacted in English.

Other significant factors

Executive function components cannot be fully isolated from each other. Children's updating performance (indexed by the working memory task) was however not a significant predictor of performance in the other tasks. This was expected in the DCCS (as the protocol controlled for the effect of memory). In the Simon task, our results seem at odds with those of Espy and Bull (2005), who found that in younger children (3- to 6-year old), performance on attention control tasks was predicted by their short-term memory span. Aside from the age difference between the two studies, methodological differences are likely to explain this apparent discrepancy in findings: we used age in months and short term memory as predictors and found that memory did not predict performance in the Simon

²²Recall that the BPI quantifies the amount of cumulative exposure and use of an "additional" language to the language of schooling.

²³The child with no current output in the home language was however an active user prior to starting school.

task over and above the effect of age; Espy and Bull (2005) compared children in three age groups and found that their memory span predicted their inhibition performance.

Socio-economic status is known to confer an executive function advantage (Ardila et al., 2005). The nature of that advantage is difficult to interpret, however, as SES is most likely a proxy for the opportunity to engage in challenging activities that lead to better executive function (Valian, 2015). It has been suggested that the impact of bilingualism might be stronger at certain SES levels (Woodard & Rodman, 2007). In our study, a positive interaction between bilingualism and SES was observed in the Simon task and the DCCS task (conferring an advantage), but this could be due to the moderate association between *Group* (monolingual vs. bilingual) and *SES* in our sample.

Age was a highly significant predictor of performance in all tasks. This is unsurprising as children experience critical phases in executive function development within the age range considered in our study (Brocki & Bohlin, 2004; Garon et al., 2008; Hongwanishkul et al., 2005).

In addition to the above, a number of experiences are known to be associated with superior executive function, including immigrant status, physical exercise, musical training, video games, time spent in leisure activities (see Valian, 2015 for a review). Personality variables may also play a role, and it is highly likely that other yet unidentified factors are also significant. Indeed, in our study, we found that individual variation dominates over the effect of participant-related variables. This suggests that, although significant, the bilingual advantage is likely to be small in comparison with other factors — as we clearly observed in the case of age. Further research will be essential to uncover these hitherto hidden factors and estimate the magnitude of their impact.

Methodological remarks

In terms of size, the population tested ($N=174$) compares favorably with most relevant studies involving children. In that sense, our study bucks the trend identified by Paap et al. (2014), according to which the bilingual advantage was confined to studies with small sample sizes.

The modeling methods adopted in our analyses allowed the use of fine-grained, continuous factors for age and SES, thereby effectively controlling for their effect and isolating (or “adjusting”) the specific effect of bilingual experience. By contrast, analyses relying on measures of central tendency are prone to confounds induced by e.g. age, SES, and amount of bilingual experience, as it is impossible to match children precisely on each of these variables. In particular, we found that age in months very strongly predicted performance in the Simon task, with each additional month in age having a clear effect. Discrepancies between studies could therefore have arisen partly from overly coarse matching of groups in terms of age.

Conclusion

The last couple of years have featured a lively and productive debate regarding the investigation of the executive function correlates of bilingualism. This debate is starting to crystallize into agreement as to the way forward in terms of the questions which remain to be addressed and the methods required to that effect.

Our contribution has been threefold. First, we have responded to the call for an operationalization of bilingualism as a continuum, to start answering the question as to how bilingual one has to be to benefit from executive function advantages. Second, we have used state-of-the-art statistical modeling techniques to analyze the data, including the pioneering use of Cox Proportional Hazard regression to capture accuracy and reaction time within a single analysis (in the Simon task). The use of regression in all analyses has enabled us to effectively control for socio-economic status, age and other potential confounds. Thirdly, we have demonstrated a significant correlation between inhibition performance and bilingual experience in 5- to 7-year-old children, on the basis of a relatively large sample (N=174) from a heterogeneous population. If inhibition indexes the common factor underlying all aspects of executive function (Miyake & Friedman, 2012; Valian, 2015), this can be interpreted as a general executive function advantage. Importantly, this advantage was observed at ages characterized by intense development of executive functions.

The Bilingual Profile Index enabled us to identify the critical threshold of bilingual experience from which an advantage can be observed at group level. To understand the exact cause of the advantage will require a different approach, in which the various aspects of bilingual experience are disentangled and allowed to vary, so that their effect and how they interact can be brought to light.

References

- Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., ... Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, 9, 2076-86.
- Abutalebi, J., & Rietbergen, M. J. (2014). Neuroplasticity of the bilingual brain: Cognitive control and reserve. *Applied Psycholinguistics*, 35, 895-899.
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, 80, 207-245.
- Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., ... Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, 5, 00398.
- Ardila, A., Rosselli, M., Matute, E., & Guajardo, S. (2005). The influence of the parents' educational level on the development of executive functions. *Developmental Neuropsychology*, 28, 539-560.
- Armitage, P., Berry, G., & Matthews, J. N. (2008). *Statistical methods in medical research* (Fourth ed.). Oxford: Wiley-Blackwell.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cognitive Science*, 4, 417-423.
- Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, 35, 857-894.
- Bedore, L., Peña, E., Summers, C., Boerger, K., Resendiz, M., Greene, K., ... Gillam, R. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, 15, 616-629.
- Best, J. R. (2012). Exergaming immediately enhances children's executive function. *Developmental Psychology*, 48, 1501-1510.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, Volume 70, 636-644.
- Bialystok, E., Craik, F., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, 19, .

- Bialystok, E., & Martin, M. (2004). Attention and inhibition in bilingual children: Evidence from the Dimensional Change Card Sort task. *Developmental Science*, 7, 325-339.
- Bialystok, E., Martin, M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9, 103-119.
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, 128, 105-119.
- Brito, N., Sebastián-Gallés, N., & Barr, R. (2015). Differences in language exposure and its effects on memory flexibility in monolingual, bilingual, and trilingual infants. *Bilingualism: Language and Cognition*, 18, 670-682.
- Brocki, K., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, 26, 571-93.
- Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology*, 22, 40-55.
- Buckner, R. L. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron*, 44, 195-208.
- Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, 130, 278-288.
- Carlson, S., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11, 282-298.
- Collett, D. (2015). *Modelling survival data in medical research* (Third edition ed.). Bristol: CRC press.
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism an example of publication bias? *Psychological Science*, 26, 99-107.
- De Cat, C., & Serratrice, L. (to appear). Predicting language proficiency in bilingual children. *Language Learning*, .
- Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, 12, 45-75.
- Diamond, A. (2011). Executive functions. *Annual Review of Psychology*, 64, 135-168.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959-964.
- Donnelly, S., Brooks, P. J., & Homer, B. (2015). Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th Annual Conference of the cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2015). The inhibitory advantage in bilingual children revisited. *Journal of Experimental psychology*, 61, 234-251.
- Engel De Abreu, P. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, 19, 529-537.
- 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, 630-42.
- Engel de Abreu, P., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor : Enhanced cognitive control in low-income minority children. *Psychological Science*, 23, 1364-1371.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and psychophysics*, 16, 143-149.
- Espy, K. A., & Bull, R. (2005). Inhibitory processes in young children and individual variation in short-term memory. *Developmental Neuropsychology*, 28, 669-688.

- Friedman, N., Miyake, A., Corley, R., Young, S., DeFries, J., & Hewitt, J. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172-179.
- Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, 134, 31-60.
- Genesee, F. (2014). Commentary on "Moving toward a neuroplasticity view of bilingualism": The early years. *Applied Psycholinguistics*, 35, 905-909.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, 1, 67-81.
- Green, D. W. (2014). Individual variability and neuroplastic changes. *Applied Psycholinguistics*, 35, 910-912.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515-530.
- Gutiérrez-Clellen, V., Calderón, J., & Ellis Weismer, S. (2004). Verbal working memory in bilingual children. *Journal of Speech, Language and Hearing Disorders*, 47, 863-876.
- Hilchey, M., & Klein, R. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin and Review*, 18, 625-658.
- Hongwanishkul, D., Happaney, K. R., Lee, W., & Zelazo, P. D. (2005). Assessment of hot and cool executive function: Age-related changes and individual differences. *Developmental Neuropsychology*, 28, 617-644.
- Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer.
- Iluz-Cohen, P., & Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism: Language and Cognition*, 16, 884-899. (10.1017/S1366728912000788)
- Ishihara, S. (1994). *The series of plates designed as a test for colour-blindness*. Tokyo: Kanehara and Co LTD.
- 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.
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Valdes Kroff, J. R. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 56, p. 229-262). Amsterdam: Elsevier.
- Kroll, J. F., & Fricke, M. (2014). What bilinguals do with language that changes their minds and their brains. *Applied Psycholinguistics*, 35, 921-925.
- Lee, Y., Nelder, J. A., & Pawitan, Y. (2006). *Generalized linear models with random effects: unified analysis via h-likelihood*. London: CRC Press.
- Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, 94, 145-188.
- Luk, G. (2015). Who are the bilinguals (and monolinguals)? *Bilingualism: Language and Cognition*, 18, 35-36.
- Luk, G., Bialystok, E., Craik, F. I. M., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *The Journal of Neuroscience*, 31, 16808-16813.
- 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* (p. 95-122). Bristol: Multilingual Matters.
- Marinis, T., Chiat, S., Armon-Lotem, S., Gibbons, D., & Gipps, E. (2010). *School-Age Sentence Imitation Test (SASIT)*. Reading: University of Reading.
- Martin-Rhee, M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 11, 81-93.
- Merz, E. C., Landry, S. H., Jeffrey M. Williams, Barnes, M. A., Eisenberg, N., Spinrad, T. L., ... School Readiness Research Consortium (2015). Parenting predictors of cognitive skills and

- emotion knowledge in socioeconomically disadvantaged preschoolers. *Journal of Experimental Child Psychology*, 132C, 14-31.
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language*, 40, 25-40.
- Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and socio-demographic correlates in an epidemiological sample of young, urban children. *Child Development*, 75, 1-14.
- Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21, 8-14.
- Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, 7, 134-140.
- Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114, 187-202.
- Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22, 1425-1433.
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, 10, 719-726.
- Mueller Gathercole, V. C., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., . . . Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon and metalinguistic tasks. *Frontiers in Psychology*, 5, 1-14.
- Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, 13, 597-616.
- Office for National Statistics. (2010). *Standard occupational classification 2010. volume 3: The national statistics socio-economic classification*. Basingstoke: Palgrave MacMillan.
- Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, 26, 242-255.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66, 232-258.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, 26, 615-639.
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in psychology*, 5, 00962.
- Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 113, 535-551.
- Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society*, 17, 682-691.
- Ripatti, S., & Palmgren, J. (2002). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56, 1016-1022.
- Rosselli, M., Ardila, A., Lalwani, L. N., & Vélez-Urbe, I. (2015). The effect of language proficiency on executive functions in balanced and unbalanced Spanish-English bilinguals. *Bilingualism: Language and Cognition, FirstView*, 1-15.

- Scaltritti, M., Peressotti, F., & Miozzo, M. (2015). Bilingual advantage and language switch: What's the linkage? *Bilingualism: Language and Cognition, FirstView*, 1-18.
- Simon, J. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology*, 81, 174-176.
- Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology*, 2, 183.
- Therneau, T. M., Grambsch, P. M., & Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics*, 12, 156-175.
- Titone, D., Pivneva, I., Sheikh, N. A., Webb, N., & Whitford, V. M. (2015). Doubling down on multifactorial approaches to the study of bilingualism and executive control. *Bilingualism: Language and Cognition*, 18, 43-44.
- Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, 18, 3-24.
- Verreyt, N., Woumans, E. V. Y., Vandelanotte, D., Szmalec, A., & Duyck, W. (2015). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition, FirstView*, 1-10.
- Weber, R. C., Johnson, A., Riccio, C. A., & Liew, J. (2015). Balanced bilingualism and executive functioning in children. *Bilingualism: Language and Cognition, FirstView*, 1-7.
- Wechsler, D. (1991). *WISC-III: Wechsler Intelligence Scale for Children: Manual*. San Antonio, TX: Psychological Corporation, Harcourt Brace Jovanovich.
- Woodard, E., & Rodman, A. (2007). *Socio-economic status as a mediating factor in the relationship between bilingualism and the cognitive development of pre-school children*. Paper presented at the annual meeting of the National Communication Association 93rd Annual Convention.
- Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, 14, 412-422.
- Zelazo, P. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, 1, 297-301.

Appendix

Implementation of the Cox PH regression

We denote the response time t of a child as a sample value of a random variable $T \geq 0$, which has an underlying probability density function $f(t)$, and the distribution function $F(t) = P(T < t) = \int_0^t f(u)du$. The survivor function $S(t)$ is defined as the probability that the response time is greater than or equal to t , and so $S(t) = P(T \geq t) = 1 - F(t)$. $S(t)$ can therefore be used to represent the probability that a child response correctly at least t (since the moment of showing the task on the screen).

The hazard function is defined as

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

which is interpreted as the instantaneous ‘risk’ of responding correctly to the task at time t , or the rate of responding correctly at time t . In Cox proportional hazard (PH) model (Cox, 1972), the hazard at time t of an individual with some covariates x_i , denoted $h_i(t|X)$, is modeled as a multiplication of the baseline hazard function that only depends on t , denoted $h_0(t)$, with the exponential of linear predictors $\exp\{x_i\beta\}$, where β is a vector of model parameters (without intercept).

1	Employers in large establishments
2	Higher managerial and administrative occupations
3	Higher professional occupations
4	Lower professional and higher technical occupations
5	Lower managerial and administrative occupations
6	Higher supervisory occupations
7	Intermediate occupations
8	Employers in small organizations
9	Own account workers
10	Lower supervisory occupations
11	Lower technical occupations
12	Semi-routine occupations
13	Routine occupations
14	Never worked and long-term unemployed
15	Full-time students

Table 14

NS-SEC operational categories (each further sub-divided according to level of responsibility and expertise to yield 48 categories in total). Source: Office of National Statistics (2010)

The (standard) Cox PH model does not take into account the correlation structure between items, but it can take into account different intercepts. We exploit this to include random effect for Item in the model. Specifically, let t_{ij} be the response time of the i -th child in task j , $i = 1, 2, \dots, n_c$ and $j = 1, 2, \dots, n_t$, δ_{ij} be the event indicator: $\delta_{ij} = 1$ if the event was uncensored and $\delta_{ij} = 0$ if the time was censored. Let $h_{ij}(t|X)$ be the hazard rate at time t for the i -th child in j -th item, given the matrix of covariates X . The extended Cox PH model can be expressed as

$$h_{ij}(t; X|\theta) = h_0(t) \exp\{X_i\beta + \theta_j\} \quad (1)$$

where θ_j is the random effect of j -th item in the sequence of tasks assumed to follow a normal distribution with mean zero and variance σ_θ^2 , i.e. $\theta_j \sim N(0, \sigma_\theta^2)$ for $j = 1, \dots, q$.

The estimation of parameters, both fixed effects and random effects, are based on the method proposed by (Ripatti & Palmgren, 2002), which are implemented in the R package ‘survival’ (Therneau et al., 2003).