# A study on the executive functioning skills of Greek–English bilingual children – a nearest neighbour approach

Athanasia Papastergiou[1] , Eirini Sanoudaki[1] , Marco Tamburelli[1] and Vasiliki Chondrogianni[2]

[1]School of Arts, Culture and Language, Bangor University, Bangor, UK and School of Philosophy, Psychology and [2]Language Sciences, University of Edinburgh, Edinburgh, UK

**CAMBRIDGE**
UNIVERSITY PRESS

## Abstract

Findings of bilingual participants outperforming their monolingual counterparts in executive functioning tasks have been repeatedly reported in the literature (Bialystok, 2017). However, uncontrolled factors or imperfectly matched samples might affect the reliability of these findings. This study aims to take into account a range of relevant variables in combination with innovative analyses to investigate the performance of one unstudied language group, Greek–English bilingual children in the north of England, compared to monolingual control groups. Our battery of executive function tasks taps into inhibition, updating and shifting. We use k-means nearest neighbour methods to match the groups and factor analysis to determine language proficiency. We find that bilinguals' accuracy is on a par with their monolingual peers – however, they are faster in inhibition and working memory tasks. Our study provides strong evidence for the presence of a bilingual advantage in these domains, while making important methodological contributions to the field.

## 1. Introduction

Many recent studies have focused on childhood bilingualism and executive control, showing that bilingual children outperform their monolingual peers on executive functioning tasks (see Adesope, Lavin, Thompson & Ungerleider, 2010; Bialystok, 2017). This is considered as a 'bilingual advantage' in executive functions (Bialystok, 2001; Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Craik & Ryan, 2006; Emmorey, Luk, Pyers & Bialystok, 2008) and has been observed in cognitive control tasks such as selective attention (Bialystok, 2001), cognitive flexibility (Poulin-Dubois, Blaye, Coutya & Bialystok, 2011) and working memory (WM) (Morales, Calvo & Bialystok, 2013). However, other studies have tended to show weaker or no effects of bilingualism (e.g., Valian, 2015).

The executive functioning system is a domain-general cognitive system, vital for the flexibility and regulation of cognition and goal-directed behaviour (Best & Miller, 2010). It is referred to as the most crucial cognitive achievement in early childhood (Bialystok & Craik, 2010). Children gradually master the ability to control attention, inhibit distraction, monitor sets of stimuli, and shift between tasks, while their working memory develops. More specifically, SHIFTING involves shifting back and forth between multiple tasks, operations, or mental sets (Monsell, 1996, as cited in Miyake et al., 2000). UPDATING includes monitoring and coding task-relevant information and replacing any no longer relevant information held in WM with the new, more relevant information (Morris & Jones, 1990). Lastly, INHIBITION is the ability to knowingly inhibit dominant, automatic, or prepotent information (Miyake et al., 2000).

The advantage in executive functions associated with bilinguals is operationalised as superior performance by bilinguals in tasks that are thought to require executive processing, which is the ability to monitor goal-setting cues, to switch attention to goal-relevant sources of information, and to inhibit those that are irrelevant or competing (Bialystok, 2006; Bialystok et al., 2004; Bialystok, Craik & Luk, 2008; Costa, Hernández & Sebastián-Gallés, 2008). These advantages are thought to be linked to the management of multiple languages and to the continuous monitoring of the appropriate language for each communicative situation (Bialystok, 2009). More specifically, bilinguals need to select the right language for each circumstance, attend to cues in order to select the right language, select the suitable lexical set and at the same time suppress the interference of the other language/s. This process is thought to generate executive functioning advantages (Bialystok, 2017).

There have been several meta-analytic reviews regarding the cognitive outcomes of bilingualism (e.g., Adesope et al., 2010; Hilchey & Klein, 2011; Hilchey, Saint-Aubin & Klein, 2015; Lehtonen, Soveri, Laine, Järvenpää, de Bruin & Antfolk, 2018) reporting mixed results in adults. More specifically, Adesope et al. (2010) analysed data from 63 studies and found

positive effects of bilingualism, including increased attention, working memory, metalinguistic awareness, and abstract and symbolic representation skills. However, there was high variability in terms of effect sizes. For inhibition, Hilchey and Klein (2011) found a global bilingual performance advantage, though insufficient evidence was provided for a bilingual effect in inhibition. Hilchey et al. (2015) in their re-analysis of the Hilchey and Klein (2011) study included more recent studies, and this time did not observe a global bilingual performance advantage.

Similar mixed findings are reported in studies examining the executive functioning skills of children. Overviews by Bialystok and colleagues (Bialystok, 2015; Bialystok, Craik & Luk, 2012) suggest that the bilingual advantage can be mostly observed in children and the elderly, possibly due to the fact that these two populations are not at the peak of their executive functioning skills as young adults are. Bialystok and colleagues agree with the idea that this advantage could be more general rather than linked to a specific executive domain such as inhibition (Bialystok, 2015; Bialystok et al., 2012). However, large-scale studies are not in line with this suggestion in other official bilingual settings such as the Basque country and Wales, where limited or no evidence of a bilingual advantage has been found (Antón, Duñabeitia, Estévez, Hernández, Castillo, Fuentes, Davidson & Carreiras, 2014; Duñabeitia, Hernández, Antón, Macizo, Estévez, Fuentes & Carreiras, 2014; Gathercole, Thomas, Kennedy, Prys, Young, Viñas-Guasch, Roberts, Hughes & Jones, 2014).

Inappropriate controlling strategies may also play a role in whether a bilingual advantage is detected (Papastergiou, Pappas & Sanoudaki, 2021). Paap and Greenberg (2013) have highlighted the need to control for an extensive number of variables within this context.

Based on the above, in the current study, we aim to answer if Greek–English bilingual children outperform two control groups of monolingual Greek and monolingual English children in executive functioning tasks tapping into inhibition, updating and shifting, by using a near-neighbour approach to control for a range of relevant variables.

## 2. Bilingual effect in children

Many studies have repeatedly reported a bilingual effect in executive functions. For example, Bialystok (1999) reports that bilingual children showed better attentional control involving shifting between different task criteria. This study investigated 30 English–Chinese bilingual and English monolingual children aged 3–5 years and 30 English–Chinese bilingual and English monolingual children aged 5–6 years using the Dimensional Change Card Sort (DCCS) task (Zelazo, Frye & Rapus, 1996). Results revealed that bilingual children gave more target responses compared to their monolingual counterparts, indicating higher levels of executive control, and suggesting that bilingualism aids the development of attentional control in task rule shifting. Similar findings were presented by Bialystok and Martin (2004). In another study, 24 bilingual and 24 monolingual 6-year-olds were comparable in identifying a simple shape hidden within drawings of complex objects in the Children's Embedded Figures Task, but the bilingual children were more able to change their interpretation of the two figures (e.g., the duck-rabbit) to acknowledge the other image in an ambiguous figures task (Bialystok & Shapero, 2005). Both tasks required perceptual

analysis, but only the ambiguous figures task required inhibiting the original meaning of the stimulus.

In line with the above, Carlson and Meltzoff (2008) aimed to investigate whether there was an advantage in executive functioning, previously observed in other languages, in 6-year-old Spanish–English bilingual children attending second-language immersion and traditional kindergartens. The bilingual children showed an advantage in executive-function tasks that require inhibition of attention to conflicting response options but not in tasks requiring inhibition of a habitual response to a familiar stimulus. Extending this pattern to infants, Kovács and Mehler (2009) investigated 40 preverbal 7-month olds; 20 infants raised in bilingual homes (14 infants exposed to Italian–Slovenian, 2 to Italian–Spanish, 2 to Italian–English, 1 to Italian–Arabic, 1 to Italian–Danish) and 20 in monolingual Italian homes. The infants brought up in bilingual homes were better able to switch responses after a change in the requirements of the task compared to their monolingual counterparts.

Additionally, Yang, Yang, and Lust (2011), in order to separate language effects and cultural effects, compared 15 Korean–American bilinguals, 13 Korean American (English-speaking) monolinguals, Korean monolinguals, and non-Korean-American (English-speaking) monolinguals, five years of age. Overall, the bilingual group was faster and more accurate compared to the monolinguals on all conditions of the Attentional Network Task (ANT), suggesting a bilingual advantage.

Finally, Poarch and van Hell (2012) found benefits of trilingualism on the Simon task and a bi- and trilingual advantage for the ANT. They investigated four groups of children 5–8 years of age using the Simon task: i) German-speaking monolingual children, ii) German speakers who were learning English as a second language (L2) in school (second language learners), iii) German–English bilingual children and trilinguals for whom either German or English was a native language along with a different language, and who were learning German or English or both at school. Findings for the Simon task provided evidence of a trilingual advantage compared to monolinguals and a strong trend towards a benefit for bilinguals compared to monolinguals. Bilinguals and trilinguals did not differ, nor did any other pairs. The L2 learners, the bilingual children and the trilingual children only took part in the ANT, six to eight months after the Simon task (Poarch & van Hell, 2012). Results showed no significant difference between bilingual and trilingual children; however, they both outperformed the L2 learners with regards to incongruent trials. There was no significant difference in response times across all children, irrespective of language status.

Large scale studies have tended to show weaker or no effects compared to smaller sample studies (Valian, 2015). For example, two recent large-scale studies, presented below, did not report any effects of bilingualism. More specifically, Antón et al. (2014) compared 360 bilingual Spanish and Basque children to Spanish monolingual children on the ANT. The researchers divided the children into three groups; i) children in 2nd and 3rd grade, ii) children in 4th and 5th grade, and iii) children in 6th and 7th grade. The first language of the bilingual children was Spanish and based on parental report the children were more fluent in Spanish compared to Basque. In addition, the bilingual children attended bilingual schools where Spanish and Basque were equally used as the languages of instruction. Their monolingual peers attended monolingual Spanish schools and they did not differ in age, reading and arithmetic skills, non-verbal IQ, and socio-economic status (SES) compared to the bilinguals. No differences

were found between the monolingual and bilingual groups. In their discussion, the authors noted that the absence of a bilingual advantage might be a result of uncontrolled factors, such as conditions associated with design and procedure.

In line with the above findings, Duñabeitia et al. (2014) used a non-verbal and a verbal Stroop task in the Spanish language to compare 504 monolingual Spanish and bilingual Spanish–Basque children. The children were enrolled in the 3$^{rd}$ to 8$^{th}$ grade. The findings suggested that the participants did show a cost of incongruence; however, the two groups of participants had similar performances. Additionally, the distribution of reaction times, overall reaction times and error rates were parallel for both bilinguals and their control group. Finally, in the regression analyses there was no effect of language status, teachers' judgments of children's reading, arithmetic, or attention skills, or IQ scores. In their discussion, the authors stated that they covered factors such as age, scores from teachers regarding reading, mathematics, and attention, general IQ test, and SES. Therefore, their groups differed only in linguistic profile; more specifically one group of children was immersed in bilingual (academic) context and the second consisted of purely monolingual children. No evidence of a bilingual advantage was observed (see also Paap & Greenberg, 2013).

Similar to the above study, Goldman, Negen, and Sarnecka (2014) recruited 32 English monolingual children and compared them to 40 bilingual children who were exposed to two languages other than English at home and to 20 bilingual children who were exposed to one extra language in addition to English. The children took part in a numerical discrimination task, tapping inhibitory control. The findings revealed no differences between the groups. In line with the above results, Kapa and Colombo (2013) found no group differences using the Flanker task with early and late Spanish–English bilingual children as well as their English monolingual control group aged 6–15 years.

Additionally, mixed results were presented by Poulin-Dubois et al. (2011). In this study, a partial bilingual advantage was observed in the shape Stroop task, a conflict task, one of the five tasks (two delay and three conflict tasks) used to measure executive functions in 33 bilingual and 30 monolingual two-year-olds. This suggested that a bilingual advantage in executive functions is first expressed in conflict inhibition. A bilingual effect was not found in the other two conflict tasks, possibly due to increased demands of those tasks or to them requiring both inhibitory control and working memory. An advantage in inhibitory control was found in simultaneous 7-month-old bilinguals when readily supressing the previously learned response and updating their predictions according to the changing requirements of the task, compared to monolinguals (Kovács & Mehler, 2009). Advantages in other executive functions were observed in slightly older children, 3–4 ½ years of age (Bialystok, Barac, Blaye & Poulin-Dubois, 2010) suggesting that it might be possible that more language experience is necessary to observe a bilingual advantage in switch-tasks due to the fact that the experience of infants has been primarily in receptive language rather than expressive language.

## 3. Possible reasons underlying contradictory findings

As shown in the previous sections, while there is a large body of research showing bilingual advantages (see Valian, 2015 for an overview), the field has not reached a consensus due to inconsistent findings. Several factors have been found to be relevant to this

bilingual effect in executive functions. Some studies show bilingual advantages in particular tasks, conditions of those, or in measures such as accuracy or reaction times, but not both (Valian, 2015). Results seem dependent on types of stimuli (e.g., verbal–nonverbal; Moreno-Stokoe & Damian, 2020). Also, the participants might get different amounts of physical exercise or might have had some other beneficial experience (e.g., musical training; Valian, 2015), or differ in terms of SES. Another, very important factor is the actual definition of bilingualism and how this is determined in each study. Bilinguals might differ in many aspects related to age of acquisition, language use, proficiency in each language, medium of education, bilingual experiences, culture (e.g., Adesope et al., 2010; Antoniou, Grohmann, Kambanaros & Katsos, 2016; Carlson & Meltzoff, 2008; Paap, Johnson & Sawi, 2016). Finally, De Bruin, Treccani, and Della Sala (2015) found a publication bias to report a bilingual effect.

### SES

Bilinguals might differ from monolinguals or other bilingual participants in socioeconomic factors, such as education, immigrant status and profession (Paap, Johnson & Sawi, 2015). The observed correlation between SES and executive functions may be due to the link of SES with the provision of emotional and academic resources in childhood (Linver, Brooks-Gunn & Kohen, 2002). Morton and Harper (2007) argued that previous studies did not appropriately match participants on SES, with the consequence that higher-SES children were being compared with monolingual children from low socioeconomic backgrounds. Some studies matching language groups on SES report a bilingual effect. For example, Engel de Abreu, Cruz-Santos, Tourinho, Martin, and Bialystok (2012) compared 40 Portuguese–Luxembourgish bilinguals and 40 Portuguese monolinguals from low-income immigrant families using flanker interference tasks. In line with Bialystok (1991, 2001, 2009), Engel de Abreu et al. (2012) found that regardless of the low-income background, this continuous use of executive functioning skills to resolve language conflict strengthened these processes in bilinguals. The results suggest that the higher the control demand of the task, the more likely it is that a bilingual effect will emerge.

Similarly, Calvo and Bialystok (2014) divided children from eight public schools into four groups which were: i) working-class monolinguals (n = 20), ii) working-class bilinguals (n = 44), iii) middle-class monolinguals (n = 46), iv) middle-class bilinguals (n = 65) based on questionnaire data on SES and on language status. The children spoke English at school and another language at home. The tasks included an intelligence test, language tests, a working memory task and a flanker task (Calvo & Bialystok, 2014). Middle-class children outperformed working-class children on all measures, and monolingual children outperformed bilingual children on language tests. Bilingual children scored higher than monolingual children on the executive functioning tasks.

Other studies closely matching bilingual and monolingual participants on SES found no bilingual advantage (Farah & Noble, 2005; Morton & Harper, 2007; Noble, Norman & Farah, 2005; Paap et al., 2015).

### Linguistic factors of bilingualism

Namazi and Thordardottir (2010) suggested that the way in which bilingualism is defined might vary across studies making

them difficult to compare. Other factors that might yield different findings might be the language background of the participants, including language exposure and language use, language of schooling, and proficiency in both languages (e.g., Bialystok & Barac, 2012; Crespo, Gross & Kaushanskaya, 2019; Iluz-Cohen & Armon-Lotem, 2013; Kubota, Chevalier & Sorace, 2020; Kuzyk, Friend, Severdija, Zesiger & Poulin-Dubois, 2020).

Language exposure and language use can be linked to the frequency of input and output a child might receive and produce (number of hours in a day, percentage of use of language and in which context). It has been shown that reduction in exposure to the L2 contributed to smaller improvement in monitoring and updating abilities – however, it did not affect the inhibition domain (Kubota et al., 2020).

In terms of language of schooling, Purić, Vuksanović, and Chondrogianni, (2017) compared Serbian children in Year 2 attending a high exposure L2 immersion program (about 5 hours of daily exposure for one year), a low exposure immersion program (about 1.5 hours of daily exposure for one year), and a monolingual control group. The high exposure group outperformed the other two groups in working memory tasks, but there were no group differences for the inhibition and shifting domain. Similarly, initial findings of a recent pilot study based in Wales suggest that children receiving minimal exposure to Welsh for a year are faster than their English monolingual counterparts in a backwards digit recall task tapping on working memory (Papastergiou, Sanoudaki & Collins, 2019). Based on Purić et al. (2017), working memory (updating) may be specifically linked to these early stages of intensive L2 learning.

Biliteracy and attending a bilingual educational setting have also been found to affect performance in cognitive tasks, such as updating and verbal working memory tasks (e.g., Andreou, Dosi, Papadopoulou & Tsimpli, 2020; Dosi & Papadopoulou, 2019; Dosi, Papadopoulou & Tsimpli, 2016). For example, Andreou et al. (2020) find that good levels of biliteracy, established by a bilingual educational setting that equally supports both languages, positively affect linguistic and cognitive skills.

Language proficiency has also been linked to executive functions. Iluz-Cohen and Armon-Lotem (2013) investigated the effect of language proficiency on executive functioning skills. They found that there is a positive relationship between language proficiency and inhibition and shifting abilities, with significantly lower performance among low language proficiency bilinguals. However, Kubota et al. (2020) found that proficiency did not affect the development of executive functioning skills in childhood.

### Fluent bilingual settings and minority and majority languages

Gathercole et al. (2014) propose that it might not be a coincidence that fluent bilinguals within bilingual communities such as Welsh–English bilinguals (Gathercole et al., 2014) and Basque-Spanish bilinguals (Antón et al., 2014; Duñabeitia et al., 2014) showed either no or mixed bilingual effects. These bilinguals are brought up with both languages as part of everyday life in their respective bilingual communities in Wales and the Basque country. It has been suggested by Lam and Dijkstra (2010) that these populations have strong between-language links and a great automaticity of the linguistic knowledge in both languages. As a result, the daily switch between both languages might not require the same cognitive effort and control, consequently not leading to bilingual effects in executive

functions. However, other studies including participants speaking minority languages within bilingual communities (e.g., Sardinian and Italian; Garraffa, Beveridge & Sorace, 2015) do show advantages, but in most cases only one test was used to tap one executive function, thus not allowing to extrapolate general theoretical implications.

### Publication bias

Finally, a study by De Bruin et al. (2015) examined abstracts from conferences between 1999 and 2012. The authors observed that studies which reported a full bilingual advantage in executive control were most likely to be published, followed by those either supporting or challenging this bilingual advantage. In contrast, those that found no bilingual advantage were the least likely to be published. This did not have any relation to differences in sample size, tests used, or statistical power, thus suggesting the existence of a publication bias. This is in line with Paap et al. (2015), who raised the concern that the literature based on executive control in bilinguals may be influenced by this bias to report a bilingual advantage. As a result, many studies that have not found evidence suggesting a bilingual advantage might have not reached publication and their hypotheses and methodologies have not enhanced our knowledge on executive functioning.

## 4. Current study

It is evident from the previous section that matching bilinguals with a monolingual control group/s has proven challenging, especially due to the variability within bilingual groups. Despite numerous studies investigating the cognitive effects of bilingualism, it is still not clearly understood which factors influence executive functioning and in what way. In the current study, we aim to control for relevant variables using innovative analyses in order to investigate the performance in executive functioning tasks of one unstudied language group of Greek–English bilingual children in the north of England. Our battery of executive function tasks taps into inhibition, updating and shifting, as operationalised by Miyake et al. (2000).

Bearing in mind previous studies on bilingualism and executive functions, we compare our Greek–English bilingual group to two monolingual control groups from both language backgrounds; namely, a control group of monolingual Greek-speaking children and a control group of monolingual English-speaking children. To the best of our knowledge, one more study has controlled for both languages of the bilingual groups of children (Torregrossa, Andreou, Bongartz & Tsimpli, 2021). Similarly, in our study we control for both languages, Greek and English, using factor analysis (Antoniou et al., 2016) to take as many variables as possible into consideration, such as language proficiency, language use and standardised vocabulary and grammar tasks. The group of bilingual children taking part in the current study attend a Greek complementary language school, a group not studied before in the U.K. for their executive functioning skills linked to language. The majority of these children are predominately exposed to Greek in the household and English at school (also see Papastergiou & Sanoudaki, 2021).

In combination with this, we use innovative analyses to control for as many variables as possible, a challenging issue in the study of bilinguals, and more specifically bilingual children. As a result, we aim to inform the debate and models of executive functions in relationship to bilingualism. More specifically, we aim to answer

**Table 1.** Participant information: parent questionnaires and scores on language and IQ tests (raw scores reported for tests)

| Variable | Participant Group | Bilinguals | Greek Control | English Control |
|---|---|---|---|---|
| **Age in months** | *N* | 19 | 15 | 25 |
| | *M (SD)* | 84.89 *(15.39)* | 88.47 *(13.69)* | 78.16 *(5.83)* |
| | Range | 63–108 | 68–108 | 67–88 |
| **Sex** | | 11f 8m | 7f 8m | 12f 13m |
| **PWFT** | *M (SD)* | 36.95 *(24.28)* | 77.6 *(8.36)* | |
| | Range | 0–82 | 58–88 | |
| **Adapted PPVT** | *M (SD)* | 34.23 *(17.85)* | 59.23 *(12.88)* | |
| | Range | 10.98–68.21 | 23.70–73.41 | |
| **CELF-4** | *M (SD)* | 51.50 *(22.93)* | | 62.07 *(12.69)* |
| | Range | 9.26–81.48 | | 33.33–83.33 |
| **BPVS3** | *M (SD)* | 53.35 *(11.08)* | | 55.93 *(6.71)* |
| | Range | 34.52–73.21 | | 40.48–70.83 |
| **DVIQ** | *M (SD)* | 79.12 *(12.87)* | 87.10 *(6.57)* | |
| | Range | 58.06–96.77 | 77.42–96.77 | |
| **Trog-2** | *M (SD)* | 60 *(15.26)* | | 61.40 *(12.79)* |
| | Range | 20–80 | | 40–90 |
| **K-BIT-2** | *M (SD)* | 104.26 *(10.83)* | 97.73 *(11.26)* | 99.96 *(18.06)* |
| | Range | 85–124 | 80–119 | 80–139 |
| **Language Use** | *M (SD)* | 52.46 *(22.33)* | 95.96 *(4.05)* | 0.28 *(0.91)* |
| | Range | 4.84–76.79 | 87.10–100 | 0–3.33 |
| **SES** | *M (SD)* | 77.96 *(16.32)* | 64.17 *(12.60)* | 77.25 *(13.84)* |
| | Range | 37.5–100 | 43.75–87.5 | 25–87.5 |

*Note. M* = Mean, *SD* = Standard Deviation, Age = participants' age in months, PWFT = Greek expressive vocabulary score, Adapted PPVT = Greek receptive vocabulary score, CELF-4 = English expressive vocabulary score, BPVS3 = English receptive vocabulary score, K-BIT-2 = non-verbal intelligence standardised score, DVIQ = Greek receptive grammar score, Trog-2 = English receptive grammar score, Language Use = Percentage of language use with 0% being only English and 100% being only Greek (For English monolingual group 100% being language other than English), SES = the average percentage of mother and father education.

the following research question: Do Greek–English bilingual children outperform two control groups of monolingual Greek children and monolingual English children in executive functioning tasks tapping into inhibition, updating and shifting, when closely matched on recently identified relevant variables?

## 5. Method

### 5.1 Participants

Nineteen Greek–English bilingual children, 15 Greek monolingual children and 25 English monolingual children, aged 63–108 months took part in this study. Details of the groups are presented in Table 1. The bilingual children were competent in both Greek and English languages to varying degrees. The Greek–English bilingual children lived in England and were recruited if at least one of their parents used Greek with them. The mean age of acquisition was 7 months (*SD* = 1 year and 3 months) for Greek and 2 years and 1 month (*SD* = 1 years and 9 months) for English. Four children had one English speaking and one Greek speaking parent and 15 children had only Greek speaking parents. We have excluded any trilingual participants. A further three children took part but were subsequently excluded because they did not meet the language

criteria (they were exposed to a third language). Also, children's scores were included in the analysis if their nonverbal intelligence score was within normal range (over 80; Kaufman & Kaufman, 2004). In this case, all children had standardised scores over 80 (*M* = 100.77, *SD* = 14.44). Children included had limited or no musical training. Based on parental and teacher reports the children did not have any hearing, behavioural, emotional, or mental impairment.

Bilingual Greek–English children were recruited from a Greek supplementary school in the north-west of England. The school offers a Greek-speaking supplementary program for 2.5 to 3.5 hours a week to enhance the reading, listening, speaking and writing skills in the Greek language and to offer knowledge around Greek culture. This programme is supplementary to the mainstream English education that these children attended. Eight of the bilingual children were born in Greece and had lived in England for at least two years at the time of the study, while the remaining bilingual children were born in England. The English monolingual control group was recruited from an infant school in the north-west of England and all the children were born in England. The Greek monolingual control group participated in Greece and all children in this group were born in Greece.

Ethical approval was granted by the University's Research Ethics Committee. Information sheets were sent to the head teachers and to parents before the study began in order to obtain informed consent. Teachers, parents, and children were provided enough time to ask any questions about the nature of the study. Parents and children were informed that they could withdraw at any time and were subsequently debriefed after the study.

## 5.2 Materials

### Parental questionnaire

The children's language experience was investigated through the Language and Social Background Questionnaire for Children (LSBQ; Luk & Bialystok, 2013). The LSBQ was forward and backward translated in Greek and was completed by the parents in their most convenient language (Greek or English)[1]. It consisted of information about the child's age, sex, country of birth, and age of acquisition of each language. Children's SES was measured as the mean of the highest attained educational level of both parents rated on an 8-point scale, which was then converted into percentages (questions 12 and 13). Parental education is the most commonly used index of SES, is highly predictive of other SES indicators (e.g., income, occupation), and is a better predictor of cognitive performance than other SES indicators (see Calvo & Bialystok, 2014).

In Section B, the child's speaking and understanding in Greek, English, or another language was rated by the parent on a 5-point scale ranging from Poor to Excellent (questions 14 and 15). A Greek proficiency parental score was derived from both scores for speaking and understanding in Greek and was included in the analysis. Similarly, both scores for speaking and understanding in English was used as the English proficiency parental score included in the analysis. Section B also included four questions about exposure to Greek and English educational settings, four questions about language acquisition and age of onset, and one question about experience with any musical instrument. Section C of the questionnaire included questions about general language use throughout the child's lifetime with parents, siblings, grandparents, neighbours, friends, and caregivers in various situations was measured on a 7-point scale ranging from 1 (only English) to 7 (only Greek/or other language).

### Non-verbal intelligence

Non-verbal intelligence was assessed using the Kaufman Brief Intelligence Test, Second Edition (K-BIT-2; Kaufman & Kaufman, 2004). The test consists of 46 items including a series of abstract images, such as designs and symbols, and visual stimuli, such as pictures of people and objects. Participants are required to understand the relationships between the presented stimuli and complete visual analogies by either pointing to the answer or saying which letter it corresponds to. All items include an option of at least five answers thus reducing chance guessing. The Matrices non-verbal subtest is individually administered, and standardised scores were calculated for the purposes of the screening, while raw scores were used in the analyses.

### Language measures

To assess the proficiency of the bilingual children in their languages, receptive and expressive vocabulary measures in each language were administered along with receptive grammar assessments. Raw scores converted to percentages were used in the analysis.

### English language measures

The British Picture Vocabulary Scale, Third Edition (BPVS3; Dunn & Dunn, 2009) was used to assess the receptive vocabulary of the bilingual and monolingual children in the English language. It is an individually administered, standardised test of Standard English receptive vocabulary for children ranging from 3 years to 16 years and 11 months. In this task, children are asked to select, out of four coloured items in a 2 by 2 matrix, the picture that best corresponds to an English word read out by the researcher. The assessment consists of 14 sets of 12 words of increasing difficulty (e.g., ball, island, fictional). The administration is discontinued when a minimum of eight errors is produced in a single set.

The Clinical Evaluation of Language Fundamentals, Fourth UK Edition – CELF-4UK (Semel, Wiig & Secord, 2006) is an individually administered standardised language measure which is used for the comprehensive assessment of a student's language skills by combining core subtests with supplementary subtests. The expressive vocabulary subtest was used here to assess the participants' expressive vocabulary in the English Language. This measure is designed for children and adolescents ranging from 5 to 16 years of age. Expressive vocabulary was screened through the Expressive Vocabulary subtest for children. Children were asked to look at a picture and name what they see or what is happening in each picture (e.g., a picture of a girl drawing, the child should give the targeted response 'colouring' or 'drawing' to score 2 points or the response 'doing homework' to score 1 point). The administration is discontinued after seven consecutive zero scores.

The Test for Reception of Grammar – Version 2 (TROG-2; Bishop, 2003) was used to assess receptive grammar. It is an individually administered standardised test for children and adults and it comprises 80 items of increasing difficulty with four picture choices. Children are asked to select the item that corresponds to the target sentence read out by the researcher. For each grammatical element there is a block of four target sentences. A block is considered to be failed unless all four items of each block are established by the child. The sentences include simple vocabulary of nouns, verbs and adjectives. If a child fails five consecutive blocks the administration is terminated.

### Greek language measures

A standard Modern Greek version of the Peabody Picture Vocabulary Task (PPVT; Dunn & Dunn, 1981) was adapted and used based on the Greek adaptation by Simos, Sideridis, Protopapas and Mouzaki (2011). The children clicked on the image, out of four possible choices, that best corresponded to the target word they heard, such as nouns, verbs, or adjectives. There were 173 items of increasing difficulty. If eight incorrect responses were provided to ten consecutive items, then the task was stopped. The answers were scored as correct (1) or incorrect (0).

The Picture Word Finding Test (PWFT; Vogindroukas, Protopapas & Sideridis, 2009a) is an individually administered standardised measure used to assess standard Modern Greek expressive vocabulary. It is a tool norm-referenced for Greek adapted from the English Word Finding Vocabulary Test – 4th Edition (Renfrew, 1995). The children are presented with 50 black and white images consisting of nouns in developmental order. The words included originate from objects, categories of

---

[1]Questionnaire in Greek can be accessed here: https://drive.google.com/file/d/1fxvoVhE6JwJApSJqTtn5aXd2HQr2weO5/view?usp=sharing

objects, television programs and fairy-tales very familiar to children. A score sheet is used to record the responses provided during testing which are later scored as correct (1) or incorrect (0). The children are asked to name the objects they saw and when they are ready, they move to the following one. The assessment is discontinued after five consecutive wrong replies.

The Developmental Verbal Intelligence Quotient (DVIQ; Stavrakaki & Tsimpli, 2000) was used to assess standard Modern Greek receptive grammar. It consists of five subtests used to measure children's language abilities in expressive vocabulary, understanding metalinguistic concepts, comprehension and production of morphosyntax, and sentence repetition. For this study, only the subtest measuring comprehension of morphosyntax (e.g., two/three elements, negative, passive voice, comparative) was used for both Greek monolingual and Greek–English bilingual children and it was administered individually. Each child was given a booklet with 31 pages, each including 3 images. The researcher read out a sentence and each child was asked to point to the picture that best represented the situation in the sentence. For example, this might have been *"το ψηλότερο δέντρο"* (the tallest tree) and the correct answer was the picture of the tallest tree out of three trees. An answer sheet was used to record the child's answers (as A, B, or C) during testing which were later scored as correct (1) or incorrect (0).

### Cognitive measures

All tasks were administered on a 15.6-inch laptop screen using the experimental software E-Prime 2.0 (Schneider, Eschman & Zuccolotto, 2002). Accuracy and reaction times (RTs) were calculated automatically through E-Prime.

### Attention task

The ATTENTIONAL NETWORK TASK (ANT; Fan, McCandliss, Sommer, Raz & Posner, 2002) was used to evaluate three different attentional networks: i) alerting; ii) orienting, and iii) executive control (Posner & Petersen, 1990). Similar to the flanker task, participants were asked to indicate the direction (left or right) that the target stimulus (centre fish here) pointed to. The child's distance between his/her head and the centre of the screen was approximately 50 cm. The child's task was to press either the right or left key button on the mouse (with the right or left index finger) corresponding to the direction in which the middle fish was swimming. The child was presented with a training block of 16 trials and 128 trials distributed in four experimental blocks. There were breaks in between the four experimental blocks. The task's length was approximately 20 minutes. Auditory feedback was offered to the child during both the training and experimental blocks.

### Working memory tasks

The first working memory task was a COUNTING RECALL TASK, which was an adaptation of the Automated Working Memory Assessment (Alloway, 2007). The children were presented with a varying number, between four and seven, of red circles and blue triangles on the laptop screen. The children were asked to count and memorise the number of red circles in each block of trials. During the recall phase the children typed the number of red circles in each trial of that block. The number of trials increased in each block, reaching seven numbers. If the child failed to correctly recall three trials in a block the task stopped.

The second working memory task was a BACKWARD DIGIT SPAN TASK (BDST) and it was adapted from Huizinga, Dolan, and van

der Molen (2006). The children began with two training trials in order to understand the task and were instructed to type the reverse order of the numbers presented. For example, if a child heard the number 7 and 4 they should type 4 and 7. The sequence begins with four trials of two numbers gradually reaching eight numbers. Similar to the above task, if the child failed to correctly recall three trials in a block the task stopped.

### Inhibition task

The NONVERBAL STROOP TASK was adapted from Lukács, Ladányi, Fazekas, and Kemény (2016) and the stimuli consisted of arrows pointing upwards, downwards, left and right. Three experimental blocks of 60 trials each were presented to the children. The aim was to select the direction that the arrows indicated regardless of their position on the screen. The children used the arrow buttons on the laptop's keyboard. The task began with the control block, where arrows were presented in the middle of the screen. In the second block, which was the congruent block, the direction of the arrows matched their position on the screen (e.g., an arrow indicating upwards was presented at the top of the screen). Finally, the third experimental block was the incongruent block. Here the direction of the arrows was the opposite compared to their position on the screen (e.g., an arrow indicating upwards was presented at the bottom of the screen).

For accuracy measures, the number of correct answers for the incongruent items was subtracted from the number of correct answers for the congruent items. The difference in RT for congruent and incongruent trials represents the inhibition cost.

### Shifting task

All children were also administered one shifting task, the COLOUR-SHAPE TASK (Purić et al., 2017). This task included three blocks each, where children were presented with two shapes (triangle, circle) coloured either red or blue. The same buttons, one for the left hand and one for the right, corresponded to one of the choices (circle-triangle, red-blue). In the first two experimental blocks, the children's task was to either recognise the shape of the stimulus and ignore their colour or the reverse. The shape stimuli were presented in the top half and the colour stimuli in the bottom half of the screen. In the third block children were required to alternate between identifying colour and shape depending on the object's location on the screen. Cues directing the participant to the relevant dimension are presented simultaneously with the stimuli on all trials, in all blocks. The first two blocks contained 32 trials each, while the third block contained 64. The number of shifting and non-shifting sequences within the third block was balanced. The difference in RT for the first two (non-shifting) and the third (shifting) block represents the shifting cost.

### 5.3 Procedure

The children were tested individually in a quiet school classroom setting, during one session in Greek for the Greek monolingual children and one session in English for the English monolingual children that lasted 40 minutes on average. The bilingual children were tested in two separate sessions; the English language session was conducted within one month of the Greek language session. The researcher informed the children that they would play some games. Parents were administered the questionnaire (LSBQ) and returned it to the classroom teacher, the school head teacher, or directly to the researcher.

**Table 2.** Results of factor analysis on the four language variables for Greek–English bilinguals and Greek monolinguals.

| Measure | Factor Loadings |
|---|---|
| | Factor 1 |
| DVIQ | **.708** |
| Greek proficiency parental score | **.750** |
| Adapted PPVT | **.955** |
| PWFT | **.935** |
| % of variance | 71.27% |

*Note.* Factor loadings over .40 are presented in bold.

**Table 3.** Results of factor analysis on the four language variables for Greek–English bilinguals and English monolinguals.

| Measure | Factor Loadings |
|---|---|
| | Factor 1 |
| CELF-4 | **.655** |
| BPVS3 | **.977** |
| TROG-2 | **.818** |
| English proficiency parental score | **.401** |
| % of variance | 55.31% |

*Note.* Factor loadings over .40 are presented in bold.

### Greek session

The bilingual participants began with the Greek language session. Each child completed the tasks in the following order: i) Greek adapted PPVT, ii) ANT, iii) Picture Word Finding Test, iv) Colour shape task, v) Nonverbal Stroop task, and vi) DVIQ. A pilot study was conducted with four children before the actual data collection. As a result of the pilot study, the choice of the above fixed order of tasks was such so the children did not feel tired or uninterested.

### English session

The second session for the bilingual participants was the English session. Each child completed the tasks in the following order: i) K-BIT-2, ii) BDST, iii) BPVS, iv) counting recall task, v) CELF-4, and vi) TROG-2.

After the end of each session the researcher thanked the child for their participation. All children participated enthusiastically.

## 6. Results

### 6.1 Preliminary data analyses

#### Outlier analysis

Response accuracy and RTs were recorded for all the executive function tests. All RTs shorter than 200 ms and all RTs for incorrect trials were excluded from the analysis; thus, only analyzing RTs from correct responses (e.g., Purić et al., 2017). Furthermore, in order to prevent extreme RTs from influencing participants' mean scores, we established ±3 standard deviation values both between and within participants. Every value that surpassed ±3 standard deviations away from the mean RT was substituted by the established lower and upper bound RTs (see also, Miyake et al., 2000). The inhibition cost for the nonverbal Stroop task was calculated as the difference between congruent and incongruent mean RTs. Local shifting costs (LSC) were calculated in the third block as the difference between the average RT for the shift trials and the average RT for the non-shift trials. General shifting costs (GSC) were calculated as the difference between the average RT for the third block and average RT for the first and second block together.

#### Factor analysis

In order to reduce the number of control variables included in the analysis, Greek and English language measures together with the proficiency scores from the parental questionnaires were submitted to a factor analysis. The analysis was conducted between the two groups of Greek–English bilinguals and Greek monolinguals and between the two groups of Greek–English bilinguals and English monolinguals. For the Greek–English bilinguals and Greek monolinguals the following four independent measures were entered into the analysis: PWFT, DVIQ, adapted PPVT, Greek proficiency parental score. For the Greek–English bilinguals and English monolinguals the following four independent measures were entered into the analysis: BPVS3, TROG-2, CELF-4, English proficiency parental score.

A Maximum Likelihood factor method was applied to the four variables for each of the two cases. Based on the analysis it was observed that participants' scores in the PWFT, DVIQ, adapted PPVT, Greek proficiency score (based on the parental report) and the BPVS3, TROG-2, CELF-4, English proficiency score (based on the parental report) clustered on one component, which represented the proficiency in each language. The analysis showed that the Greek proficiency factor explained 71.27% of the variance and the English proficiency factor 55.31% of the variance. Tables 2 and 3 summarise the Maximum Likelihood results. Table 4 indicates the correlations between the control background variables.

#### Matching method

For the analysis of the data we applied k:1 nearest neighbour matching (Rubin, 1973). The idea behind matching methods is to compare the outcomes ($Y$) of subjects that are as similar as possible to a number of covariates ($X$), with the sole exception of the treatment status. In our case, we would like to compare the executive function accuracy and response time of a monolingual with those of a bilingual child as long as they have similar values in other background scores; namely, the Age in months, Sex, K-BIT-2, SES, English proficiency factor, Greek proficiency factor. Only then can we be sure that any difference in the outcome variable is a consequence of the action rather than of the correlation between a test and the outcome.

For a single covariate, like the PWFT, identifying a pair of comparable children is simple. Adding a second covariate that is binary (e.g., Sex) or categorical (e.g., SES) would require more effort on our behalf and a larger dataset. However, if we want to consider more covariates, particularly if they are continuous (e.g., K-BIT-2), then finding matches becomes a daunting task. To circumvent this problem, a similarity measure or similarity index may be constructed, which quantifies how close two observations (i.e., scores from two children) are. Two well-established methods are the k-means nearest neighbour matching and the propensity score matching.

The k-means nearest neighbour matching calculates the "distance" between pairs of observations with regard to a set of

**Table 4.** Correlations between the control background variables

|            | Greek prof | PWFT | adapted PPVT | DVIQ |
|------------|------------|------|--------------|------|
| Greek prof | 1          | .70  | .69          | .56  |
| PWFT       | .70        | 1    | .89          | .66  |
| adapted PPVT | .69      | .89  | 1            | .71  |
| DVIQ       | .56        | .69  | .71          | 1    |

covariates ($X$'s) and then "matching" each subject to comparable observations that are closest to it. For example, suppose that a bilingual participant has a PWFT score of 65.7 and we also have information on two monolingual children – monolingual A and B – where A has a PWFT score of 55.3 and B of 64.1. Naturally, monolingual B represents a closer match to the bilingual, and B would therefore be selected by the k-means nearest neighbour matching. In this case, the distance is simply $d = |65.7 - 64.1| = 1.6$, which is also known as the Eucleidian distance. If more than one variable is used to match, then the distance statistic that is used is the Mahalanobis, which takes into account the correlation between the covariates and the fact that they may be measured on different scales.

The k-means nearest neighbour matching does not use a formal model for either the outcome or the treatment status and this makes it very flexible. However, when matching on more than one continuous covariate, the k-means nearest neighbour estimator must be augmented with a bias-correction term (Abadie & Imbens, 2006, 2011)[2].

The k-means nearest neighbour matching relies on some distance function. For example, initially assume a single covariate – the PWFT score. In the general form we can denote this variable as $x$. Then the distance between two individuals $i, j$ where the i individual is bilingual and the j individual is not can be given as

$$|x_i - x_j| = \frac{(x_i - x_j)(x_i - x_j)}{Cov(x, x)}$$

We can generalise this formula for when we have p number of covariates using matrix algebra. Assume that $x = \{x_1, x_2, \ldots, x_p\}$ and that each individual, $i$, has the following set of covariates $\mathbf{x}_i = \{x_{1,i}, x_{2,i}, \ldots, x_{p,i}\}$. The distance between individuals i, j is now given as:

$$||\mathbf{x}_i - \mathbf{x}_j|| = ((\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{x}_j))^{1/2}$$

where $\mathbf{S}$ is the variance-covariance matrix of the covariates.

Coming back to observation $i$, we can define the following set of nearest-neighbor index

$$\Omega(i)^x = \{j | t_j = 1 - t_i, ||\mathbf{x}_i - \mathbf{x}_j|| < ||\mathbf{x}_i - \mathbf{x}_l||_\mathbf{S} , \ t_l = 1 - t_i \forall l \neq j\}$$

where $i$ is the observation (i.e., the participant) who is bilingual and for whom we want to find a matching monolingual. $j$ denotes the matching monolingual (only one in this case) and $l$ denotes another monolingual candidate. $t$ denotes the treatment effect and takes the value of 1 for bilinguals, zero otherwise. $||\mathbf{x}_i - \mathbf{x}_j||$ and $||\mathbf{x}_i - \mathbf{x}_l||$ denote the distance between i, j and i, l respectively

[2]Further information is about the matching estimators are included in the Technical Appendix.

and in the formula above we require that the distance between $i, j$ is smaller than $i, l$ (since we select the matching $j$ participant as our match). The notation $t_j = 1 - t_i$ and $t_l = 1 - t_i$ implies that our $i$ participant who is bilingual (hence $t_i = 1$) needs to be matched with some monolingual participant for whom $t_j = 1 - 1 = 0$ or $t_l = 1 - 1 = 0$

The above can be generalised for $m$ matching participants

$$\Omega(i)_m^x = \{j_1, j_2, \ldots, j_m | t_{j_k} = 1 - t_i, ||\mathbf{x}_i - \mathbf{x}_{jk}||_\mathbf{S} \\ < ||\mathbf{x}_i - \mathbf{x}_{jk}||_\mathbf{S} , \ t_l = 1 - t_i \forall l \neq j_k\}$$

The structure of $\mathbf{S}$ depends on our initial assumption and can be one of Euclidean, Mahalanobis or inverse variance. Formally

$$\mathbf{S} = \begin{cases} \mathbf{I}_p & \text{for the Euclidean case} \\ \dfrac{(\mathbf{X} - \bar{\mathbf{x}}'\mathbf{1}_n)' W(\mathbf{X} - \bar{\mathbf{x}}'\mathbf{1}_n)}{\sum_i^n w_i - 1} & \text{for the Mahalanobis case} \\ \text{diag}\left\{\dfrac{(\mathbf{X} - \bar{\mathbf{x}}'\mathbf{1}_n)' W(\mathbf{X} - \bar{\mathbf{x}}'\mathbf{1}_n)}{\sum_i^n w_i - 1}\right\} & \text{for the inverse variance case} \end{cases}$$

where $\mathbf{1}_n$ is an $n \times 1$ vector of ones, $\mathbf{I}_p$ is the identity matrix of order p, same as the number of covariates used. $w_i$ is the frequency weight for the $i$ observation, $\bar{x} = \sum_i^n w_i x_i / \sum_i^n w_i$ which denotes a weighted mean and $\mathbf{W}$ is an $n \times n$ diagonal matrix containing the frequency weights.

For the prediction of the potential outcomes we use the following: $y_{1,i}$ is the potential outcome of the $i$ individual that has received the treatment or in our case is bilingual ($t = 1$). Conversely, $y_{0,i}$ is the potential outcome of the $i$ individual that has not received the treatment or in our case is monolingual ($t = 0$). As we have discussed, the problem posed by the potential-outcome model is that only $y_{1,i}$ or $y_{0,i}$ is observed, never both. The k-means nearest neighbours can predict the potential outcome for the $i$ observation as follows:

$$\hat{y}_{t,i} = \begin{cases} y_i & \text{if } t_i = t \ \text{ for } t \in \{0, 1\} \\ \dfrac{\sum_{j \in \Omega(i)} w_j y_j}{\sum_{j \in \Omega(i)} w_j} \end{cases}$$

The first is the case where the outcome of the individual ($y_i$) is observed whether he is bilingual (t = 1) or monolingual (t = 0). The second case is the counterfactual outcome which does not exist and is estimated as the outcome of the closest match (or matches).

Once the above are estimated we can define the following quantities of interest; namely, the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATET). These are defined as

$$ATE = \tau_1 = E(y_1 - y_0)$$

$$ATET = \delta_1 = E(y_1 - y_0 | t = 1)$$

and obviously as $y_{1,i}$ and $y_{0,i}$ are realisations of the $y_1$ and $y_0$ random variables respectively, $y_1$ is the average of all $y_{1,i}$ and the equivalent holds for $y_0$

**Table 5.** Descriptive Statistics – RTs in Executive Function Tasks

| Tasks | Bilinguals | | | | Greek Monolinguals | | | | English Monolinguals | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | Min | Max | M | SD | Min | Max | M | SD | Min | Max |
| GSC | −291.42 | 240.80 | −757.72 | 79.56 | −268.753 | 204.2933 | −573.7827 | 68.11292 | −331.217 | 229.9095 | −823.735 | 109.9965 |
| LSC | −125.91 | 151.42 | −452.32 | 98.30 | −128.332 | 196.4123 | −526.9701 | 93.01868 | −156.176 | 142.9145 | −475.535 | 142.9883 |
| Inhibition Cost | −233.9352 | 186.9592 | −546.9808 | 188.338 | −272.37 | 227.1293 | −568.037 | 204.0914 | −381.233 | 191.0973 | −765.184 | 95.6483 |
| BDST | 865.1175 | 313.8921 | 273 | 1612.429 | 1089.913 | 816.3311 | 401.1667 | 3818.859 | 1920.779 | 717.0673 | 911.875 | 3518.739 |
| Count Recall | 2915.154 | 1278.187 | 1033.887 | 4986.751 | 2374.906 | 1252.296 | 1039.916 | 4766.141 | 2844.503 | 1456.564 | 952.1733 | 6717.39 |
| ANTcong | 1000.159 | 208.7245 | 686.1579 | 1394.72 | 987.3016 | 187.3585 | 717.0995 | 1394.581 | 1107.997 | 152.3785 | 775.7869 | 1399.675 |
| ANTincong | 1124.714 | 243.9346 | 809.6393 | 1668.079 | 1099.647 | 224.9456 | 745.6984 | 1516.116 | 1216.706 | 167.2387 | 895.7541 | 1525.547 |
| Stroop cong | 994.5229 | 357.4001 | 501.1017 | 1681.476 | 946.3075 | 272.5045 | 445.0204 | 1498.74 | 1116.352 | 221.8276 | 493.5714 | 1475.074 |
| Stroop incong | 1239.598 | 329.6895 | 575.4286 | 1991.7 | 1238.133 | 320.0497 | 708.1667 | 1700.744 | 1497.585 | 316.3162 | 397.9231 | 1956.878 |
| Colour-Shape cong | 901.0499 | 130.4189 | 658.6734 | 1126.833 | 961.6368 | 125.1502 | 747.3365 | 1176.129 | 951.1903 | 166.3187 | 622.6957 | 1263.273 |
| Colour-Shape incong | 1187.341 | 276.124 | 686.3704 | 1694.677 | 1229.012 | 167.6315 | 948.1464 | 1547.466 | 1280.2 | 279.1717 | 687.2692 | 1739.979 |

Note: GSC = Global shifting cost; LSC = Local shifting cost; Inhibition cost = The difference RT for congruent and incongruent trials; BDST = Backward digit span task; Count Recall = Counting recall task; cong = congruent trial; incong = incongruent trials.

## 6.2 Analyses of executive functions

Tables 5 and 6 report descriptive statistics for the accuracy and RT measures from each executive function task for each group. In the case of accuracy in the two working memory tasks (BDST and Counting Recall tasks) a higher score indicates better performance, whereas for the RT a lower score indicates better performance. Similarly, for the accuracy in attention, switching and inhibition tasks (ANT, Arrow Stroop & Colour-Shape tasks) a higher score indicates better performance, whereas a lower RT score indicates better performance. We performed comparisons between the three groups of children. Table 7 and Table 8 show the results of the monolingual and bilingual groups on the attention and working memory tasks.

### Comparison 1

The first comparison was between the bilingual group and the Greek monolingual group. Participants were matched via nearest neighbour matching as described above. The matching variables were Age in months, Sex, K-BIT-2, SES, Greek proficiency factor. There were no differences between the bilingual group and the Greek monolingual group based on RTs on the Arrow Stroop. No group difference was found for the inhibition accuracy scores. Similarly, no significant group differences were found for the remaining tasks, where the groups performed comparably (see Table 7 for p-values).

### Comparison 2

The second comparison was between the bilingual group and the English monolingual group. Nearest neighbour matching was again applied to match participants for the same matching variables; namely, Age in months, Sex, K-BIT-2, SES, English proficiency factor.

The differences between the groups based on RTs emerged on the inhibition task; namely, Arrow Stroop, where the bilingual group was faster compared to the monolingual group. In addition, there was a significant Stroop effect $(\hat{b} = 139.728, p = .033)$. However, no group difference was found for the inhibition accuracy scores. A significant group difference was also found for the BDST where the bilinguals were faster compared to their monolingual counterparts $(\hat{b} = -1021.77, p < .001)$. In the remaining tasks, the groups performed comparably (see Table 8 for p-values).

## 7. Discussion

The present study investigated differences in the executive functioning skills of Greek–English bilingual children compared to two groups of Greek and English monolingual children. We investigated the executive functioning scores using a battery of tests assessing inhibition, shifting, and updating, and matching closely for language proficiency, SES, language use, vocabulary and grammar scores, and non-verbal intelligence. Our aim was to see if the Greek–English bilingual children would outperform their monolingual counterparts in line with multiple previous findings (see Bialystok, 2017), once a large number of potentially confounding variables was controlled for using innovative analyses, and therefore to contribute methodologically to the debate on whether a bilingual advantage exists and/or how reliable it is.

To achieve this, the bilingual children were compared to two closely matched monolingual control groups, one consisting of Greek monolinguals and the other of English monolinguals. We

**Table 6.** Executive Function Tasks Descriptive statistics – accuracy

| Tasks | Bilinguals | | | | Greek Monolinguals | | | | English Monolinguals | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | SD | Min | Max | M | SD | Min | Max | M | SD | Min | Max |
| BDST | 0.59 | 0.10 | 0.29 | 0.73 | 0.56 | 0.12 | 0.25 | 0.73 | 0.57 | 0.07 | 0.45 | 0.73 |
| Count Recall | 0.42 | 0.15 | 0.21 | 0.75 | 0.32 | 0.14 | 0.14 | 0.61 | 0.35 | 0.09 | 0.14 | 0.50 |
| ANTcong | 0.92 | 0.08 | 0.77 | 1.00 | 0.84 | 0.18 | 0.43 | 1.00 | 0.91 | 0.08 | 0.64 | 1.00 |
| ANTincong | 0.82 | 0.21 | 0.22 | 1.00 | 0.62 | 0.29 | 0.03 | 0.97 | 0.87 | 0.10 | 0.59 | 0.98 |
| Stroop cong | 0.93 | 0.07 | 0.78 | 1.00 | 0.84 | 0.18 | 0.43 | 1.00 | 0.86 | 0.16 | 0.22 | 1.00 |
| Stroop incong | 0.74 | 0.31 | 0.10 | 1.00 | 0.62 | 0.29 | 0.03 | 0.97 | 0.64 | 0.20 | 0.27 | 0.97 |
| Colour-Shape cong | 0.89 | 0.08 | 0.69 | 0.98 | 0.83 | 0.09 | 0.69 | 0.97 | 0.88 | 0.06 | 0.75 | 0.97 |
| Colour-Shape incong | 0.67 | 0.13 | 0.42 | 0.89 | 0.68 | 0.11 | 0.48 | 0.88 | 0.63 | 0.11 | 0.38 | 0.77 |

**Table 7.** Executive functioning tasks: comparison of bilingual group and Greek control group

| Executive Function Task | Nearest – Neighbour Matching | |
|---|---|---|
| Working Memory | *Coef* | *p* |
| BDST Overall ACC Score | .010 | .799 |
| Counting Recall Overall ACC Score | .098 | .115 |
| BDST Overall RT Score | −207.52 | .354 |
| Counting Overall Recall RT Score | 327.49 | .641 |
| **Attention** | | |
| ANT Overall ACC Score | −.010 | .821 |
| ANT Overall RT Score | 42.014 | .566 |
| ANT Congruent RT Score | 29.418 | .695 |
| ANT Incongruent RT Score | 54.100 | .497 |
| ANT Congruent ACC Score | −.020 | .399 |
| ANT Incongruent ACC Score | −.001 | .992 |
| **Inhibition** | | |
| Arrow Stroop Overall RT Score | −1.5231 | .986 |
| Arrow Stroop Congruent RT Score | 105.12 | .293 |
| Arrow Stroop Incongruent RT Score | 39.61 | .744 |
| Inhibition Switch Cost | 55.040 | .473 |
| Arrow Stroop Overall ACC Score | .015 | .792 |
| Arrow Congruent ACC Score | .030 | .531 |
| Arrow Incongruent ACC Score | .037 | .774 |
| **Shifting** | | |
| Colour -Shape Congruent ACC Score | .050 | .168 |
| Colour – Shape Incongruent ACC Score | −.024 | .511 |
| Colour – Shape Congruent RT Score | −7.674 | .902 |
| Colour – Shape Incongruent RT Score | 52.96 | .559 |
| Colour – Shape GSC | −64.33 | .598 |
| Colour – Shape LSC | −29.08 | .777 |

*Note.* Matching variables: Non-Verbal Ability (K-BIT-2), Age, SES, Sex, Greek Proficiency Factor.

**Table 8.** Executive functioning tasks: comparison of bilingual group and English control group

| Executive Function Task | Nearest – Neighbour Matching | |
|---|---|---|
| Working Memory | *Coef* | *P* |
| BDST Overall ACC Score | −.001 | .974 |
| Counting Recall Overall ACC Score | .0354 | .487 |
| BDST Overall RT Score | −1021.77 | **.000** |
| Counting Overall Recall RT Score | −1802.183 | .511 |
| **Attention** | | |
| ANT Overall ACC Score | −.041 | .365 |
| ANT Overall RT Score | −90.959 | .184 |
| ANT Congruent RT Score | −89.746 | .177 |
| ANT Incongruent RT Score | 72.879 | .267 |
| ANT Congruent ACC Score | −.019 | .573 |
| ANT Incongruent ACC Score | −.061 | .361 |
| **Inhibition** | | |
| Arrow Stroop Overall ACC Score | .029 | .604 |
| Arrow Stroop Congruent RT Score | −155.147 | .148 |
| Arrow Stroop Incongruent RT Score | −294.875 | **.031** |
| Inhibition Switch Cost | 139.728 | **.033** |
| Arrow Congruent ACC Score | .044 | .435 |
| Arrow Incongruent ACC Score | .061 | .489 |
| **Shifting** | | |
| Colour -Shape Congruent ACC Score | .041 | .087 |
| Colour – Shape Incongruent ACC Score | .057 | .170 |
| Colour – Shape Congruent RT Score | 57.263 | .900 |
| Colour – Shape Incongruent RT Score | −71.264 | .504 |
| LSC | 41.671 | .551 |
| GSC | 76.280 | .430 |

*Note.* Matching variables: Non-Verbal Ability (K-BIT-2), Age, SES, Sex, English Proficiency Factor.

used a factor analysis on four indicators of language proficiency to reveal one factor that we interpreted as proficiency in English and Greek and closely matched the participants using the k-means nearest neighbour matching. This close matching gives us greater confidence in the results taking into consideration a large number of relevant variables. The results showed that Greek–English bilinguals were faster than the English monolinguals in two executive function measures in terms of RT – namely, i) the bilingual children were faster in the incongruent inhibition trials and demonstrated a lower inhibition switch cost, in the inhibition task (Stroop); ii) the bilinguals were faster than the English monolinguals, in the backward WM digit span (BDST). In all the other executive function measures the bilingual children were comparable to the English monolingual children. The bilingual children showed no difference in their performance compared to the Greek monolingual control group.

These findings support the hypothesis that bilingualism influences the development of executive functions and extend previous research (Blom, Boerma, Bosma, Cornips & Everaert, 2017; Bosma, Hoekstra, Versloot & Blom, 2017; Costa, Hernández, Costa-Faidella & Sebastián-Gallés, 2009; Garraffa et al., 2015; Lauchlan, Parisi & Fadda, 2013). After controlling and closely matching this group of bilinguals to two monolingual control groups on a large number of relevant variables, a bilingual effect was observed in inhibition and working memory. The comparison between the bilingual group and the English monolingual group elicited a bilingual effect only in one working memory task and in the inhibition task. Our study is in line with previous research that has showed mixed findings in executive function tasks (Paap & Greenberg, 2013; Ross & Melinger, 2017).

In contrast to the previous comparison where a bilingual effect was found, the bilingual group was comparable in all the measures to the Greek monolingual group. The fact that there was no significant difference in any task between the bilingual and the Greek monolingual group may be linked to the fact that due to the Greek educational system, we could not avoid recruiting children in Greece that were exposed to English at least one hour a week starting in Year 1 and reaching three hours a week in Year 3 (Greek Ministry of Education and Religious Affairs, 2016). This is in combination with after school language classes, where children attend English classes at least two hours a week. It is possible that these few hours of English a week have affected the executive functioning scores. Other studies investigating dual language development and executive functions of bilingual children attending L2 education programs have found advantages in working memory after as little as one year of immersion education – for example, in a group of Serbian-speaking second-grade children (Purić et al., 2017). Nicolay and Poncelet (2015) found positive effects after 3 years of immersion education in alerting, auditory selective attention, divided attention, and mental flexibility, in line with Carlson and Meltzoff (2008) who reported a bilingual advantage on a battery of executive function tasks after 6 months of immersion.

In contrast, the bilinguals had faster reaction times in the inhibition and BDST working memory task compared to the monolingual English group. Based on Purić et al. (2017), working memory may be specifically linked to these early stages of intensive L2 learning. This finding is in line with previous research showing a bilingual advantage in working memory (Antoniou et al., 2016; Bialystok, 2011; Blom, Küntay, Messer, Verhagen & Leseman, 2014; Purić et al., 2017). The bilingual effect in the inhibition domain is also in line with previous research on

bilingualism (Bialystok, 2017). Based on Paap's (2018) Controlled Dose hypothesis, this bilingual advantage might be present due to the fact that the bilinguals are still in the process of learning how to control their languages and are constantly monitoring and inhibiting. The English control group did not have exposure to an L2 whereas the Greek control group did – however, it might be the case of other variables playing a role such as differences in the aspects of the curriculum across the two school systems, cultural effects between these two control groups (Yang et al., 2011) and other school activities and hobbies, such as playing video games and sports that have been found to affect executive functions (Paap, Anders-Jefferson, Mason, Alvarado & Zimiga, 2018; Valian, 2015; Vestberg, Gustafson, Maurex, Ingvar & Petrovic, 2012). Future research could take all these additional variables into account.

However, the other tasks, one tapping into working memory (Counting recall task) and one tapping into inhibition (ANT; only the CONFLICT INDEX was analysed here) revealed no significant differences on either accuracy or response times. This might be an issue linked to reliability and validity of commonly used executive function tasks. The view that these tasks are far from optimal is supported by many researchers in the field (e.g., Laine & Lehtonen, 2018; Paap & Greenberg, 2013; Paap & Sawi, 2014; Soveri, Lehtonen, Karlsson, Lukasik, Antfolk & Laine, 2018). This dissociation between tasks might also be linked to the lack of theory on the bilingual advantage and the lack of clarity in the architecture of executive functions despite the division by Miyake et al. (2000) into three interrelated components (shifting, inhibition, and WM). Even though the above tasks supposedly tap the same domain, that does not mean that they are correlated with each other (Jylkkä, Lehtonen, Lindholm, Kuusakoski & Laine, 2018; Laine & Lehtonen, 2018). Though some researchers have reported that forwards and backwards recall tasks load onto the same factor during factor analysis (e.g., Colom, Abad, Rebollo & Shih, 2005; Engle, Kane & Tuholski, 1999), others state that a reversal of order requires the involvement of executive-attentional resources (e.g., Elliott, Smith & McCulloch, 1997). On the other hand, Costa et al. (2008) and Pelham and Abrams (2014) found a significant bilingual conflict effect using the ANT when testing young adults. This might be linked to the engagement of the monitoring processes during an executive function task that may depend on several properties of the design, such as different type of stimuli. If, for example, a task involves one type of trials, monitoring processes may not be recruited as much (Costa et al., 2009). As Costa et al. (2009) hypothesise in their study, a bilingual advantage could be linked to a more efficient monitoring processing system, that checks which strategy should be applied in a specific trial. They found that in low-monitoring conditions no bilingual advantage was detected in contrast to high monitoring condition where a bilingual conflict effect was observed. Perhaps, the child-friendly version of the ANT used in the current study was not challenging enough. Similarly, in Antón et al. (2014) and Carlson and Meltzoff (2008) no difference was found in the children's version of the ANT task between the bilingual and monolingual children. The fact that we only found the significant difference in RT in the Stroop and the BDST tasks might be linked to the fact that a bilingual advantage in monitoring and updating may speed up performance, leading to not only overall faster RTs but also to a smaller conflict effect (Costa et al., 2009).

On the shifting task we did not find any bilingual effect. As Huizinga et al. (2006) stated, various executive function

components may develop asynchronously. This is in line with previous research not finding effects of bilingualism in any executive function tasks (Paap & Greenberg, 2013).

## 8. Limitations and future directions

Due to practical matters, we used non-standardised tasks to assess Greek receptive vocabulary and grammar skills in Greek monolingual and bilingual children as well as English tests that are not standardised for bilingual children. Future development of tests is needed in Greek and English, which should also include bi-mutilingual children (Babatsouli, 2019; Marinis, Armon-Lotem & Pontikas, 2017). Also, standardised Greek tests assessing language skills such as the Action Picture Test for Greek (Vogindroukas, Protopapas & Stavrakaki, 2009b) and more recently the Logometro (Mouzaki, Ralli, Antoniou, Diamanti & Papaioannou, 2017) can also be used to assess Greek grammar skills.

Future studies can shed light on the possibility that limited exposure to a second language could enhance executive functions. Pursuing this might clarify the reasons why no differences were identified between the Greek-speaking cohort and the Greek–English bilingual cohort as well as mixed findings in other studies. This finding has important educational implications especially for Greece, where there will be a pilot project of teaching English for two hours a week, as a compulsory topic, in state nurseries from September 2020 (Greek Ministry of Education and Religious Affairs, 2020). Additionally, the European Commission is working together with national governments aiming for all citizens to begin learning foreign languages at an early age (European Commission, 2019). Finally, in Wales similar findings were obtained in a pilot study where children receiving minimal exposure to Welsh for a year were faster in a working memory task than their English monolingual counterparts (Papastergiou et al., 2019). Future longitudinal studies can further investigate these groups with minimal exposure to a second language and how this interacts with executive functions.

The relatively small sample size of this study is one of its limitations. Nevertheless, our findings extend previous research and demonstrate that after controlling and closely matching this group of bilinguals to two monolingual control groups on related factors, a bilingual effect is observed in inhibition and working memory.

Based on these results and as a further step, the bilingual advantage debate on executive function could be approached holistically, using frontier methodologies that allow to jointly consider information from multiple domains of executive function (e.g., Papastergiou et al., 2021).

## 9. Conclusion

The aim of this study was to examine the differences in the executive functioning skills of Greek–English bilingual children compared to two control groups of Greek and English monolingual children. The contribution of this study to the field is empirical and methodological; namely, we considered recently identified relevant variables in combination with innovative analyses and one unstudied language group of Greek–English bilingual children from the north of England. More specifically, we used k-means nearest neighbour methods to match bilingual to monolingual children on a wide array of variables, including age, SES, Greek and English proficiency. We used a factor analysis on

four indicators of language proficiency to reveal one factor which we interpret as proficiency in English and Greek, closely matching on language background information that we obtained from both objective and contextual factors. This close matching gives us greater confidence in the results that revealed a bilingual advantage in two domains, inhibition and working memory, compared to the English monolingual group, while the Greek monolingual group was comparable to the Greek–English bilingual group. The latter finding might be explained by Greek children's exposure to small amounts of English in Greece due to the nature of the Greek educational system or it could be clarified in the way executive function is divided and analysed. Our findings extend previous research on the effect of L2 exposure on executive functions.

## References

**Abadie A and Imbens GW** (2006) Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**, 235–267.

**Abadie A and Imbens GW** (2011) Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics* **29**, 1–11.

**Adesope OO, Lavin T, Thompson T and Ungerleider C** (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research* **80**, 207–245.

**Alloway TP** (2007) *Automated Working: Memory Assessment: Manual*. Pearson.

**Andreou M, Dosi I, Papadopoulou D and Tsimpli I-M** (2020) Heritage and non-heritage bilinguals: The role of bi-literacy and bilingual education. In Brehmer B, Treffers-Daller J and Berndt D (eds), *Lost in transmission: The role of attrition in heritage language development*. John Benjamins, pp. 172–196.

**Antón E, Duñabeitia JA, Estévez A, Hernández JA, Castillo A, Fuentes LJ, Davidson DJ and Carreiras M** (2014) Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology* **5**, 398.

**Antoniou K, Grohmann KK, Kambanaros M and Katsos N** (2016) The effect of childhood bilectalism and multilingualism on executive control. *Cognition* **149**, 18–30.

**Babatsouli E** (2019) A phonological assessment test for child Greek. *Clinical Linguistics & Phonetics* **33**, 601–627. https://doi.org/10.1080/02699206.2019.1569164

**Best JR and Miller PH** (2010) A developmental perspective on executive function. *Child Development* **81**, 1641–1660.

**Bialystok E** (1991) Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (Ed.), *Language Processing in Bilingual Children* (pp. 113–140). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620652.008

**Bialystok E** (1999) Cognitive complexity and attentional control in the bilingual mind. *Child Development* **70**, 636–644. https://doi.org/10.1111/1467-8624.00046

**Bialystok E** (2001) *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.

**Bialystok E** (2006) Bilingualism at school: Effect on the acquisition of literacy. In McCardle P and Hoff E (ed), *Childhood bilingualism: Research on infancy through school age*. Clevedon: Multilingual Matters, pp. 107–124.

**Bialystok E** (2009) Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition* **12**, 3–11.

**Bialystok E** (2011) Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology* **65**, 229–235 https://doi.org/10.1037/a0025406

**Bialystok E** (2015) Bilingualism and the Development of Executive Function: The Role of Attention. *Child Development Perspectives* **9**, 117–121. https://doi.org/10.1111/cdep.12116

**Bialystok E** (2017) The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin* **143**, 233.

**Bialystok E and Barac R** (2012) Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition* **122**, 67–73. https://doi.org/https://doi.org/10.1016/j.cognition.2011.08.003

**Bialystok E, Barac R, Blaye A and Poulin-Dubois D** (2010) Word Mapping and Executive Functioning in Young Monolingual and Bilingual Children. *Journal of Cognition and Development* **11**, 485–508. https://doi.org/10.1080/15248372.2010.516420

**Bialystok E and Craik FIM** (2010) Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science* **19**, 19–23. https://doi.org/10.1177/0963721409358571

**Bialystok E, Craik FIM, Klein R and Viswanathan M** (2004) Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging* **19**, 290.

**Bialystok E, Craik FIM and Luk G** (2008) Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34**, 859.

**Bialystok E, Craik FIM and Luk G** (2012) Bilingualism: Consequences for mind and brain. In *Trends in Cognitive Sciences* (Vol. **16**, Issue 4, pp. 240–250). https://doi.org/10.1016/j.tics.2012.03.001

**Bialystok E, Craik FIM and Ryan J** (2006) Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 1341.

**Bialystok E and Martin MM** (2004) Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science* **7**, 325–339.

**Bialystok E and Shapero D** (2005) Ambiguous benefits: the effect of bilingualism on reversing ambiguous figures. *Developmental Science* **8**, 595–604. https://doi.org/10.1111/j.1467-7687.2005.00451.x

**Bishop DVM** (2003) *TROG-2 test for reception of grammar-2.* London: Harcourt.

**Blom E, Boerma T, Bosma E, Cornips L and Everaert E** (2017) Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Frontiers in Psychology* **8**, 552. https://www.frontiersin.org/article/10.3389/fpsyg.2017.00552

**Blom E, Küntay AC, Messer M, Verhagen J and Leseman P** (2014) The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology* **128**, 105–119. https://doi.org/https://doi.org/10.1016/j.jecp.2014.06.007

**Bosma E, Hoekstra E, Versloot A and Blom E** (2017) The minimal and short-lived effects of minority language exposure on the executive functions of Frisian-Dutch bilingual children. *Frontiers in Psychology* **8**, 1453. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01453

**Calvo A and Bialystok E** (2014) Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition* **130**, 278–288.

**Carlson SM and Meltzoff AN** (2008) Bilingual experience and executive functioning in young children. *Developmental Science* **11**, 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

**Colom R, Abad FJ, Rebollo I and Shih PC** (2005) Memory span and general intelligence: A latent variable approach. *Intelligence* **33**, 623–624.

**Costa A, Hernández M, Costa-Faidella J and Sebastián-Gallés N** (2009) On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition* **113**, 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

**Costa A, Hernández M and Sebastián-Gallés N** (2008) Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition* **106**, 59–86.

**Crespo K, Gross M and Kaushanskaya M** (2019) The effects of dual language exposure on executive function in Spanish–English bilingual children with

different language abilities. *Journal of Experimental Child Psychology* **188**, 104663. https://doi.org/https://doi.org/10.1016/j.jecp.2019.104663

**De Bruin A, Treccani B and Della Sala S** (2015) Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science* **26**, 99–107.

**Dosi I and Papadopoulou D** (2019) The role of educational setting in the development of verbal aspect and executive functions: evidence from Greek-German bilingual children. *International Journal of Bilingual Education and Bilingualism.* https://doi.org/10.1080/13670050.2018.1539465

**Dosi I, Papadopoulou D and Tsimpli I-M** (2016) Linguistic and Cognitive Factors in Elicited Imitation Tasks: A Study with Mono- and Bi-literate Greek-Albanian Bilingual Children. In Scott J and Waughtal D (eds), *Proceedings of the 40th annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press, pp. 101–115.

**Duñabeitia JA, Hernández JA, Antón E, Macizo P, Estévez A, Fuentes LJ and Carreiras M** (2014) The inhibitory advantage in bilingual children revisited. *Experimental Psychology* **61**, 234–251.

**Dunn LM and Dunn DM** (2009) *The British picture vocabulary scale.* GL Assessment Limited.

**Dunn LM and Dunn LM** (1981) *Peabody Picture Vocabulary Test-Revised.* Circle Pines. MN: American Guidance Service.

**Elliott CD, Smith P and McCulloch K** (1997) *Technical manual British Ability Scales II.* Windsor, Berkshire: NFERNELSON Publishing Company.

**Emmorey K, Luk G, Pyers JE and Bialystok E** (2008) The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science* **19**, 1201–1206.

**Engel de Abreu PM, Cruz-Santos A, Tourinho CJ, Martin R and Bialystok E** (2012) Bilingualism enriches the poor: enhanced cognitive control in low-income minority children. *Psychological science* **23**, 1364–1371. https://doi.org/10.1177/0956797612443836

**Engle R, Kane M and Tuholski S** (1999) Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In Miyake A and Shah P (eds), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge: Cambridge University Press, pp. 102–134. https://doi:10.1017/CBO9781139174909.007

**European Commission**. (2019) Council recommendation on a comprehensive approach to the teaching and learning of languages. *Official Journal of the European Union.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.C_.2019.189.01.0015.01.ENG&toc=OJ%3AC%3A2019%3A189%3ATOC

**Fan J, McCandliss BD, Sommer T, Raz A and Posner MI** (2002) Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience* **14**, 340–347. https://doi.org/10.1162/089892902317361886

**Farah MJ and Noble KG** (2005) Socio-economic influences on brain development: a preliminary study. In Mayr U, Awh E and Keele SW (eds), *Developing individuality in the human brain: A tribute to Michael Posner.* Washington, DC: APA, pp. 189–208.

**Garraffa M, Beveridge M and Sorace A** (2015) Linguistic and cognitive skills in Sardinian–Italian bilingual children. *Frontiers in Psychology* **6**, 1898. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01898

**Gathercole VCM, Thomas EM, Kennedy I, Prys C, Young N, Viñas-Guasch N, Roberts EJ, Hughes EK and 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.

**Goldman MC, Negen J and Sarnecka BW** (2014) Are bilingual children better at ignoring perceptually misleading information? A novel test. *Developmental Science* **17**, 956–964. https://doi.org/10.1111/desc.12175

**Greek Ministry of Education and Religious Affairs** (2016) *Ωρολόγιο Πρόγραμμα – Υπουργείο Παιδείας.* https://www.minedu.gov.gr/publications/docs2016/ypourgiki_apofasi_olohmera.doc

**Greek Ministry of Education and Religious Affairs**, (2020) *18-06-20 Υπουργική απόφαση για τις δραστηριότητες στην αγγλική γλώσσα στο Νηπιαγωγείο.* https://www.minedu.gov.gr/ypapegan/ypour-apof/45416-18-06-20-ypourgikiapofasi-gia-tis-drastiriotites-stin-aggliki-glossa-sto-nipiagogeio

**Hilchey MD and Klein RM** (2011) Are there bilingual advantages on non-linguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review* **18**, 625–658.

**Hilchey MO, Saint-Aubin J and Klein RM** (2015) Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In Schwieter JW (ed), *The Cambridge Handbook of Bilingual Processing*. Cambridge University Press, pp. 586–613. https://doi.org/DOI:10.1017/CBO9781107447257.026

**Huizinga M, Dolan CV and van der Molen MW** (2006) Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia* **44**, 2017–2036.

**Iluz-Cohen P and Armon-Lotem S** (2013) Language proficiency and executive control in bilingual children. *Bilingualism* **16**, 884–899. https://doi.org/10.1017/S1366728912000788

**Jylkkä J, Lehtonen M, Lindholm F, Kuusakoski A and Laine M** (2018) The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition* **21**, 505–522.

**Kapa LL and Colombo J** (2013) Attentional control in early and later bilingual children. *Cognitive Development* **28**, 233–246.

**Kaufman AS and Kaufman NL** (2004) *K-BIT-2: Kaufman Brief Intelligence Test*. 2nd. Circle Pines, Minn: American Guidance Service.

**Kovács ÁM and Mehler J** (2009) Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6556–6560.

**Kubota M, Chevalier N and Sorace A** (2020) Losing access to the second language and its effect on executive function development in childhood: The case of "returnees." *Journal of Neurolinguistics* **55**, 100906. https://doi.org/https://doi.org/10.1016/j.jneuroling.2020.100906

**Kuzyk O, Friend M, Severdija V, Zesiger P and Poulin-Dubois D** (2020) Are there cognitive benefits of code-switching in bilingual children? A longitudinal study. *Bilingualism: Language and Cognition* **23**, 542–553. https://doi.org/DOI:10.1017/S1366728918001207

**Laine M and Lehtonen M** (2018) Cognitive consequences of bilingualism: Where to go from here? *Language, Cognition and Neuroscience*, **33**(9), 1205–1212. https://doi.org/10.1080/23273798.2018.1462498

**Lam KJY and Dijkstra T** (2010) Word repetition, masked orthographic priming, and language switching: bilingual studies and BIA+ simulations. *International Journal of Bilingual Education and Bilingualism* **13**, 487–503. https://doi.org/10.1080/13670050.2010.488283

**Lauchlan F, Parisi M and Fadda R** (2013) Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a 'minority' language. *International Journal of Bilingualism* **17**, 43–56. https://doi.org/10.1177/1367006911429622

**Lehtonen M, Soveri A, Laine A, Järvenpää J, de Bruin A and Antfolk J** (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin* **144**, 394.

**Linver MR, Brooks-Gunn J and Kohen DE** (2002) Family processes as pathways from income to young children's development. *Developmental Psychology* **38**, 719–734. https://doi.org/10.1037/0012-1649.38.5.719

**Luk G and Bialystok E** (2013) Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology* **25**, 605–621.

**Lukács Á, Ladányi E, Fazekas K and Kemény F** (2016) Executive functions and the contribution of short-term memory span in children with specific language impairment. *Neuropsychology* **30**, 296. https://doi.org/10.1037/neu0000232

**Marinis T, Armon-Lotem S and Pontikas G** (2017) Language impairment in bilingual children: state of the art. *Linguistic Approaches to Bilingualism* **7**, 265–276. https://doi.org/10.1075/lab.00001.mar

**Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A and Wager TD** (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.

**Monsell S** (1996) Control of mental processes. In Bruce V (ed), *Unsolved mysteries of the mind: Tutorial essays in cognition*. Hove, UK: Erlbaum, pp. 93–148.

**Morales J, Calvo A and Bialystok E** (2013) Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology* **114**, 187–202.

**Moreno-Stokoe CM and Damian MF** (2020) Employing Natural Control for Confounding Factors in the Hunt for the Bilingual Advantage in Attention: Evidence from School Children in Gibraltar. *Journal of Cognition* **3**, 5. DOI: http://doi.org/10.5334/joc.94

**Morris N and Jones DM** (1990) Memory updating in working memory: The role of the central executive. *British Journal of Psychology* **81**, 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

**Morton JB and Harper SN** (2007) What did Simon say? Revisiting the bilingual advantage. *Developmental science* **10**, 719–726. https://doi.org/10.1111/j.1467-7687.2007.00623.x

**Mouzaki A, Ralli A, Antoniou F, Diamanti V and Papaioannou S** (2017) *Logometro*. Intelearn. https://www.logometro.gr/start

**Namazi M and Thordardottir E** (2010) A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism* **13**, 597–616.

**Nicolay AC and Poncelet M** (2015) Cognitive benefits in children enrolled in an early bilingual immersion school: A follow up study. *Bilingualism: Language and Cognition* **18**, 789–795. https://doi.org/DOI:10.1017/S1366728914000868

**Noble KG, Norman MF and Farah MJ** (2005) Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science* **8**, 74–87. https://doi.org/10.1111/j.1467-7687.2005.00394.x

**Paap K** (2018) Bilingualism in cognitive science: the characteristics and consequences of bilingual language control. In De Houwer A and Ortega L (eds), *Handbook of bilingualism*. Cambridge: Cambridge University Press.

**Paap KR, Anders-Jefferson R, Mason L, Alvarado K and Zimiga B** (2018) Bilingual advantages in inhibition or selective attention: More challenges. *Frontiers in Psychology* **9**, 1–10.

**Paap KR and Greenberg ZI** (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* **66**, 232–258.

**Paap KR, Johnson HA and Sawi O** (2015) Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* **69**, 265–278.

**Paap KR, Johnson HA and Sawi O** (2016) Should the search for bilingual advantages in executive functioning continue? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior* **74**, 305–314. https://doi.org/10.1016/j.cortex.2015.09.010

**Paap KR and 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**, 962. https://doi.org/10.3389/fpsyg.2014.00962

**Papastergiou A, Pappas V and Sanoudaki E** (2021) The executive function of bilingual and monolingual children: A technical efficiency approach. *Behavior research methods*, 10.3758/s13428-021-01658-7. Advance online publication. https://doi.org/10.3758/s13428-021-01658-7

**Papastergiou A and Sanoudaki E** (2021) Language skills in Greek–English bilingual children attending Greek supplementary schools in England. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2021.1980496

**Papastergiou A, Sanoudaki E and Collins B** (2019) Does minimal exposure to Welsh influence executive functions? Evidence from Welsh–English children. Oral presentation at the *2nd International Conference on Bilingualism*, University of Malta, Malta, 25–27 March 2019. https://www.um.edu.mt/__data/assets/pdf_file/0007/386719/BookofAbstracts_2019.pdf

**Pelham SD and Abrams L** (2014) Cognitive advantages and disadvantages in early and late bilinguals. *Journal of experimental psychology. Learning, memory, and cognition* **40**, 313–325. https://doi.org/10.1037/a0035224

**Poarch GJ and van Hell JG** (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.

**Posner MI and Petersen SE** (1990) The attention system of the human brain. *Annual Review of Neuroscience* **13**, 25–42.

**Poulin-Dubois D, Blaye A, Coutya J and Bialystok E** (2011) The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology* **108**, 567–579.

**Purić D, Vuksanović J and Chondrogianni V** (2017) Cognitive advantages of immersion education after 1 year: Effects of amount of exposure. *Journal of Experimental Child Psychology* **159**, 296–309.

**Renfrew CE** (1995) *The Renfrew language scales*. Winslow.

**Ross J and Melinger A** (2017) Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science* 20, e12405. https://doi.org/10.1111/desc.12405

**Rubin DB** (1973) Matching to remove bias in observational studies. *Biometrics* 29, 159–184.

**Schneider W, Eschman A and Zuccolotto A** (2002) *E-Prime: User's guide.* Psychology Software Incorporated.

**Semel E, Wiig EH and Secord WA** (2006) *Child Evaluation of Language Fundamentals-Preschool Second UK Edition.* Oxford: Pearson Assessment.

**Simos PG, Sideridis GD, Protopapas A and Mouzaki A** (2011) Psychometric evaluation of a receptive vocabulary test for Greek elementary students. *Assessment for Effective Intervention* 37, 34–49.

**Soveri A, Lehtonen M, Karlsson LC, Lukasik K, Antfolk J and Laine M** (2018) Test-retest reliability of five frequently used executive tasks in healthy adults. Applied neuropsychology. *Adult*, **25**(2), 155–165. https://doi.org/10.1080/23279095.2016.1263795

**Stavrakaki S and Tsimpli IM** (2000) Diagnostic verbal IQ test for Greek pre-school and school age children: Standardization, statistical analysis, psychometric properties. In *Proceedings of the 8th Symposium of the Panhellenic Association of Logopedists* (pp. 95–106).

**Torregrossa J, Andreou M, Bongartz C and Tsimpli I** (2021) Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition* 24, 694–706. https://10.1017/S1366728920000826

**Valian V** (2015) Bilingualism and cognition. *Bilingualism: Language and Cognition* 18, 3–24.

**Vestberg T, Gustafson R, Maurex L, Ingvar M and Petrovic P** (2012) Executive functions predict the success of top-soccer players. *PLoS ONE* **7**, 1–5.

**Vogindroukas I, Protopapas A and Sideridis G** (2009a) *Experiment on the Expressive Vocabulary (Greek version of Renfrew Word Finding Vocabulary Test).* Chania: Glafki.

**Vogindroukas I, Protopapas A and Stavrakaki S** (2009b) *Action Picture Test for Greek.* Chania: Glafki.

**Yang S, Yang H and Lust B** (2011) Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition* 14, 412–422.

**Zelazo PD, Frye D and Rapus T** (1996) An age-related dissociation between knowing rules and using them. *Cognitive Development* **11**, 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

## Technical Appendix

### *Matching estimators*

Matching estimators are used in evaluating the impact of a treatment effect upon an outcome of interest. Let $W_i$ indicate whether the individual $i$ ($i = 1, …, N$) is exposed to treatment, with treatment denoted as $W_i = 1$, thus $W_i = 0$ represents the control group. For simplicity we assume that $W_i \in \{0, 1\}$ (i.e., only a treated and a control group are present), but extension to multiple treatments is possible. The number of treated individuals is denoted as $N_1 = \sum_{i=1}^{N} W_i$, and thus the control group includes $N_0 = N - N_1$ individuals.

The outcome of interest may be represented as $Y_i$, and we can denote as $Y_i(0)$ and $Y_i(1)$ the outcomes without and with treatment, respectively. The treatment effect upon the outcome of interest for individual $i$ is given as $Y_i(1) - Y_i(0)$. Thus, in a fictional setting of parallel universes we would evaluate the average treatment effect as $\tau = N^{-1} \sum_{i=1}^{N} Y_i(1) - Y_i(0)$. However, for a given individual $i$, only one of the two quantities is observed:

$$Y_i = \begin{cases} Y_i(1) & \text{if } W_i = 1 \\ Y_i(0) & \text{if } W_i = 0 \end{cases} \tag{1}$$

Hence, for each individual $i$ that has participated in a treatment, we need a counterfactual equivalent of the *same* participant that would not have participated in the treatment, and *vice versa*. One way to achieve this is via randomization of the treatment, but this is not always possible and/or arguably unattainable in practice. Another way is to estimate the average expected outcome of a counterfactual participant on each occasion, which leads to regression estimators. Or phrased differently, we can identify a counterfactual participant and estimate the outcome, which leads to matching estimators. Under both occasions a set of $k$ ($k = 1, …, K$) observed characteristics (i.e., covariates), we can denote these as $X_{i,k}$, are used to identify the $i$ counterfactual individual.[3]

### Matching using regression estimators

In the case of regression, we assume a single covariate for simplicity and $\hat{\mu}_w(X)$ is a consistent estimator of $\mu_w(X)$, thus we have:

$$\bar{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \hat{\mu}_0(X_i) & \text{if } W_i = 1 \end{cases} \tag{2}$$

And

$$\bar{Y}_i(1) = \begin{cases} \hat{\mu}_1(X_i) & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 0 \end{cases} \tag{3}$$

Therefore, we use $\hat{\mu}_0(X_i)$ and $\hat{\mu}_1(X_i)$ to estimate the counterfactual outcomes. Thus, the treatment effect may be estimated as:

$$\hat{\tau}_{\text{reg}} = N^{-1} \sum_{i=1}^{N} (\bar{Y}_i(1) - \bar{Y}_i(0)) \tag{4}$$

### Matching using k-means nearest neighbour estimators

In the case of k-means nearest neighbour matching estimators, we have:

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ M^{-1} \sum_{j \in \mathcal{J}_m(i)} Y_j & \text{if } W_i = 1 \end{cases} \tag{5}$$

and

$$\hat{Y}_i(1) = \begin{cases} M^{-1} \sum_{j \in \mathcal{J}_m(i)} Y_j & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1 \end{cases} \tag{6}$$

where $M$ ($m = 1, …, M$) denotes the number of matches to individual $i$. If $m = 1$ then $M^{-1} \sum_{j \in \mathcal{J}_m(i)} Y_j \equiv Y_j$, that is only the closest match is used. Thus, the treatment effect using matching estimators may be estimated as:

$$\hat{\tau}_{NN}^m = N^{-1} \sum_{i=1}^{N} (\hat{Y}_i(1) - \hat{Y}_i(0)) \tag{7}$$

Contrary to regression, matching estimators utilize the observed characteristics ($X_i$) to identify candidate matches for each individual $i$ that has participated in the treatment ($W_i = 1$). The k-means nearest neighbour matching relies on a distance function to measure the distance (i.e., the similarity) between two individuals $i$, $l$ where the $i$ individual is part of the treatment group (i.e., bilingual), and the $l$ individual is not. The distance between these observations may be given as:

$$|X_i - X_l| = \frac{(X_i - X_l)(X_i - X_l)}{Cov(X, X)} \tag{8}$$

We can generalise this formula for when we have $k$ number of covariates using matrix algebra. Assume that $X = \{X_1, X_2, …, X_k\}$ and that each individual $i$ has the following set of observed characteristics $\mathbf{X}_i = \{X_{i,1}, X_{i,2}, …, X_{i,k}\}$. The

---

[3]A set of assumptions is required here, most notably that of "unconfoundness" between the outcome variable ($Y$) and the covariates ($X$), which states that conditional on the covariates, the treatment $W$ is as good as randomised.

distance between individuals $i$, $l$ is now given as:

$$||\mathbf{X}_i - \mathbf{X}_l|| = ((\mathbf{X}_i - \mathbf{X}_l)'\mathbf{S}^{-1}(\mathbf{X}_i - \mathbf{X}_l))^{0.5} \quad (9)$$

where $\mathbf{S}$ is the variance-covariance matrix of the observed characteristics. The structure of the variance-covariance matrix can be one of Euclidean, Mahalanobis, or invese variance, formally:

$$\mathbf{S} = \begin{cases} \mathbf{I}_k & \text{for the Euclidean case} \\ \dfrac{(\mathbf{X} - \bar{\mathbf{X}}'1_N)'\mathbf{W}(\mathbf{X} - \bar{\mathbf{X}}'1_N)}{\sum_{i=1}^{N} w_i - 1} & \text{for the Mahalanobis case} \\ \text{diag}\left\{\dfrac{(\mathbf{X} - \bar{\mathbf{X}}'1_N)'\mathbf{W}(\mathbf{X} - \bar{\mathbf{X}}'1_N)}{\sum_{i=1}^{N} w_i - 1}\right\} & \text{for the inverse variance case} \end{cases} \quad (10)$$

where $1_N$ is a $N \times 1$ vector of ones, $\mathbf{I}_k$ is the identity matrix of order $k$, same as the number of observed characteristics used, $w_i$ is the frequency weight for the $i$ observation, $\bar{\mathbf{X}} = \sum_{i=1}^{N} w_i X_i / \sum_{i=1}^{N} w_i$ which denotes a weighted mean, and $\mathbf{W}$ is an $N \times N$ diagonal matrix containgn the frequency weights.

Coming back to individual $i$, we can define the following set of nearest-neighbour index:

$$\mathcal{J}(i) = \{l | W_l = 1 - W_i, \ ||\mathbf{X}_i - \mathbf{X}_l|| < ||\mathbf{X}_i - \mathbf{X}_q||_\mathbf{S}, \ W_q = 1 - W_i \forall l \neq q\} \quad (11)$$

where $i$ is the treated individual (i.e., bilingual) and for whom we want to find a matching control (i.e., monolingual); $l$, $q$ denote two candidate matching monolinguals; $W$ denotes the treatment effects and takes the value of 1 for bilinguals, zero otherwise; $||\mathbf{X}_i - \mathbf{X}_l|| < ||\mathbf{X}_i - \mathbf{X}_q||_\mathbf{S}$ denotes the distance between $i$, $l$ and $i$, $q$ respectively and in the formula above we require that the distance between $i$, $l$ is smaller than $i$, $q$ (since we select the matching $l$ individual as our match). The notation $W_l = 1 - W_i$ and $W_q = 1 - W_i$ implies that our $i$ individual, who is bilingual (hence $W_i = 1$), needs to be matched with some monolingual individual for whom $W_l = 0$ or $W_q = 0$.

The above can be generalised for $m$ matching individuals, as:

$$\mathcal{J}_m(i) = \{l_1, l_2, \ldots, l_m | W_{l_m} = 1 - W_i, \ ||\mathbf{X}_i - \mathbf{X}_{l_m}|| < ||\mathbf{X}_i - \mathbf{X}_{q_m}||_\mathbf{S}, \ W_q = 1 - W_i \forall l_m \neq q_m\} \quad (12)$$

Hence, for the prediction of outcomes using the k-means nearest neighbour and assuming $m$ matches we have:

$$\hat{Y}_i(W_i) = \begin{cases} Y_i \text{ if } W_i = w \text{ for } w \in \{0, 1\} \\ \dfrac{\sum_{j \in \mathcal{J}_m(i)} m_j Y_i}{\sum_{j \in \mathcal{J}_m(i)} m_j} \end{cases} \quad (13)$$

where the first is the case where the outcome is observed whether the individual is bilingual ($W_i = 1$) or monolingual ($W_i = 0$). The second case is the counterfactual outcome which does not exist and is estimated as the weighted average outcome of the $m$-closest matches.

## Matching using bias-adjusted k-means nearest neighbour estimators

The two estimators thus far are not asymptotically equivalent; the $\hat{Y}_i(W)$ is not a consistent estimator of $\mu_w(X)$ due to the specific choice of matches for the former, see Abadie and Imbens (2006) for more details. Bias-adjusted k-means nearest neighbour matching estimators for continuous distributed characteristics use a regression correction term to ensure consistency of the matching estimator. Assuming the regression used is equal to the true regression function (i.e., no misspecification) this bias-adjustment adds only noise to the matching estimator, without however affecting its unbiasedness. Nevertheless, under the presence of misspecification in the regression, which may arise due to omitted variables and/or imprecise measurement, the bias-adjustment ensures that the quantity $M^{-1} \sum_{j \in \mathcal{J}_m(i)} Y_j(1)$ converges to $\mu_w(X)$; thus ensuring consistency of the estimator. The bias-adjusted k-means nearest matching estimator is given as:

$$\check{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ M^{-1} \sum_{j \in \mathcal{J}_m(i)} (Y_j + \hat{\mu}_0(\mathrm{X}_i) - \hat{\mu}_0(\mathrm{X}_j)) & \text{if } W_i = 1 \end{cases} \quad (14)$$

and

$$\check{Y}_i(1) = \begin{cases} M^{-1} \sum_{j \in \mathcal{J}_m(i)} (Y_j + \hat{\mu}_1(\mathrm{X}_i) - \hat{\mu}_1(\mathrm{X}_j)) & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1 \end{cases} \quad (15)$$

Thus, the treatment effect using matching estimators may be estimated as:

$$\hat{\tau}_{B-NN}^m = N^{-1} \sum_{i=1}^{N} (\check{Y}_i(1) - \check{Y}_i(0)) \quad (16)$$

In our case we compare the executive function accuracy and response time of a monolingual with those of a bilingual child, matching on observed characteristics related to: i) Age in months; ii) Sex; ii) K-BIT-2; iv) SES; v) English proficiency factor; vi) Greek proficiency factor.