Bilingualism: Language and Cognition

cambridge.org/bil

Inside the multilingual and bidialectal mind: an investigation of the cognitive effect on executive function*

Kyriakos Antoniou[1,2,3] and George Spanoudis[2,4]

[1]Department of Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus; [2]Center for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus; [3]School of Humanities, Hellenic Open University, Patras, Greece and [4]Department of Psychology, University of Cyprus, Nicosia, Cyprus

Research Article

Cite this article: Antoniou, K. and Spanoudis, G. (2025). Inside the multilingual and bidialectal mind: an investigation of the cognitive effect on executive function. *Bilingualism: Language and Cognition* **28**, 764–777. https://doi.org/10.1017/S1366728924000506

Received: 15 May 2023
Revised: 22 July 2024
Accepted: 22 July 2024
First published online: 2 December 2024

Keywords:
multilingualism; bilingualism; bidialectalism; dialects; executive function

Corresponding author:
Kyriakos Antoniou;
Email: kyriakos.antoniou@cut.ac.cy

*This publication is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No. 800305.

Abstract

Whether speaking two or more languages (multilingualism) or dialects of one language (bidialectalism) affect executive function (EF) is controversial. Theoretically, these effects may depend on at least two conditions. First, the multilingual and bidialectal characteristics; particularly, (second) language proficiency and the sociolinguistic context of language use (e.g., Green & Abutalebi, 2013). Second, the EF aspects examined; specifically, recent accounts of the locus of the multilingual effect propose a general EF effect rather than an impact on specific processes (Bialystok, 2017). We compared 52 "monolingual" (with limited additional-language/dialect experience), 79 bidialectal and 50 multilingual young adults in the diglossic context of Cyprus, where bidialectalism is widespread and Cypriot and Standard Greek are used in different everyday situations. Three EF processes were examined via seven tasks: inhibition, switching and working memory (Miyake et al., 2000). We found better multilingual and bidialectal performance in overall EF, an effect moderated by high (second) language proficiency.

1. Introduction

Multilingualism—the experience of regularly using more than one language—characterises most of the world's population today (Grosjean & Li, 2013). Undoubtedly, speaking multiple languages confers various professional, cultural and social benefits; for instance, enhancing the opportunities for employment, travel, cross-cultural communication and socialisation in an increasingly interconnected, mobile and globalised world. However, what is currently controversial is the claim that multilingualism leads to further positive effects on cognitive and brain functioning; specifically, on executive function (EF) (e.g., Bialystok, 2017; de Bruin et al., 2021; Lehtonen et al., 2018; Paap, 2019; Paap et al., 2015; Planckaert et al., 2023; Yurtsever et al., 2023; i.a.).

EF is a domain-general, non-verbal neurocognitive system that enables individuals to regulate their thoughts and goal-directed behaviours. The term *EF* is generally used to refer to a collection of neurocognitive processes (e.g., Miyake et al., 2000). Miyake et al. (2000) proposed an influential theoretical model according to which EF consists of various cognitive processes that are moderately related and partly independent. They originally identified three such processes, even though their study was not a comprehensive examination of all possible different EF processes: *inhibition* (the ability to ignore irrelevant information and suppress automatic responses), *shifting* or *task-switching* (the skill of flexibly switching from one task or representation to another) and *updating* or *working memory* (the ability to monitor the contents of working memory and update it based on relevance to a given task). To date, the Miyake et al. (2000) model is still widely used; and the processes of inhibition, switching and working memory remain some of (if not) the most commonly studied neurocognitive functions in the EF literature (see in Karr et al., 2018), including multilingualism research. For example, most meta-analyses on multilingualism and EF evaluate the evidence for at least one or all these three processes because, theoretically, multilingual language use, in some way, draws on and, thus, practices each of these functions (e.g., Gunnerud et al., 2020; Lehtonen et al., 2018; Lowe et al., 2021; Monnier et al., 2022; Paap, 2019; Planckaert et al., 2023).

In general, multilingualism has been argued to positively affect (aspects of) EF because, presumably, multilinguals need to constantly recruit EF during everyday communication to restrict use to the intended language (e.g., Bialystok, 2017). The reason for this is the evidence showing that, when multilinguals speak in one language, the non-relevant language(s) is(are) never completely switched off (e.g., Kroll et al., 2012). Thus, the regular employment of EF for effective language use in multilinguals results in a fortification of executive processes.

Currently, however, evidence for such a multilingual effect is inconsistent. From a theoretical perspective, this effect may depend on at least two conditions. First, the specific multilingual characteristics; particularly, (second) language proficiency level and the sociolinguistic context of language use (e.g., Green & Abutalebi, 2013). Second, the EF aspects examined. For example, recent accounts of the multilingual effect localise it in general EF facets but not in specific EF processes (e.g., Bialystok, 2017). In this context, our research aimed to further investigate the multilingual EF effect through a theoretically driven approach. First, we tested participants from a diglossic sociolinguistic context, where two different linguistic varieties—in our study, Cypriot and Standard Modern Greek—are employed in separate domains of daily life. Crucially, this is a sociolinguistic situation where neurocognitive effects may be expected based on a specific theoretical framework, the adaptive control hypothesis (Green & Abutalebi, 2013). Second, we directly examined the role of (second) language proficiency in the multilingual EF effect, an aspect of multilingual experience that from a theoretical (and empirical) perspective has been linked to EF (e.g., Lehtonen et al., 2018, p. 395). Finally, we used Miyake et al.'s (2000) model to test EF and to contrast different proposals about the cognitive locus of a potential multilingual effect (e.g., Bialystok, 2017; Houtzager et al., 2017; Santillán & Khurana, 2018). We were also interested in examining whether a cognitive effect can be found in bidialectal speakers, who regularly use two very similar, genetically related, and to some degree mutually comprehensible dialects of the same language (e.g., Antoniou et al., 2016; Antoniou & Milaki, 2021). These effects were investigated by administering seven EF tasks to three young-adult groups: participants with relatively limited additional-language/dialect experience ("monolinguals"), bidialectals and multilinguals.

1.1. Multilingualism and EF: the controversy

Early research provided strong evidence for a positive multilingual effect on EF (e.g., Adesope et al., 2010). Moreover, within research supporting the multilingual cognitive effect, there have been different proposals regarding the specific cognitive locus of such impact. These can be summarised into two broad accounts. The first view suggests that multilingualism selectively impacts specific EF components, such as inhibition, switching and working memory (e.g., Houtzager et al., 2017; Santillán & Khurana, 2018). An inhibition effect, for example, may stem, based on Green's (1998) model, from multilinguals' continuous experience in inhibiting the non-relevant language when using the target language. The second view suggests that multilingualism has a broader effect on the EF system, influencing general aspects, such as *executive attention* (Bialystok, 2017). *Executive attention* refers to the ability to focus attention to and maintain in memory information that is relevant to a task at hand in the face of interference (Bialystok, 2017). Bialystok (2017) suggests that it is trained in multilingual adults because of the constant experience of handling conflict between two or more jointly activated languages.

In the last decade, however, many studies failed to find evidence for EF differences favouring multilinguals over monolinguals. This holds for children, young, and older adults (e.g., Antón et al., 2016; Gathercole et al., 2014; Nichols et al., 2020; Studenica et al., 2022). Reflecting this, recent meta-analyses on the multilingual EF impact across the lifespan typically report smaller effect sizes (see also systematic reviews and quantitative analyses of, for example, Degirmenci et al., 2022, for older adults; Grundy, 2020, for young adults; Planckaert et al., 2023 and Yurtsever et al., 2023, for children). These vary from small-to-moderate (e.g., Grundy & Timmer, 2017, for working memory across the lifespan; Gunnerud et al., 2020, for switching in children) to small (e.g., de Bruin et al., 2015, for speakers across the lifespan; Donnelly et al., 2019, for inhibition, across the lifespan; Gunnerud et al., 2020, for overall EF in middle-class children; Monnier et al., 2022, for working memory, across the lifespan) or even null after publication bias correction (e.g., Lehtonen et al., 2018, for adults; Lowe et al., 2021, for children; see also Paap, 2019, for speakers across the lifespan).

Conflicting results are similarly found in studies with multilingual or bidialectal speakers of closely related languages or dialects, across the lifespan. For instance, superior EF has been reported for multilingual or bidialectal speakers of Spanish and Catalan (e.g., Costa et al., 2009, with young adults); Frisian and Dutch (e.g., Blom et al., 2017, with children; Houtzager et al., 2017, with older adults), Sardinian and Italian (e.g., Garraffa et al., 2017, with adults); and Cypriot Greek and Standard Modern Greek (Antoniou et al., 2016, with children). The *Ethnologue* reports that the lexical overlap between Spanish-Catalan and Sardinian-Italian is at 85%, while for Cypriot Greek and Standard Modern Greek is in the range of 84%–92% (Eberhard et al., 2023). It further suggests that 85% is the cut-off point for two linguistic varieties being dialects of the same language. Similarly, Dyen et al. (1992) report that Frisian and Dutch share 82.2% cognates, which is well above the authors' 70% suggested threshold for two linguistic varieties belonging to the same language[1]. However, other studies failed to show better EF performance for multilinguals or bidialectals who spoke very similar languages or dialects, such as English and varieties of Scottish (e.g., Kirk et al., 2014, with older adults; Ross & Melinger, 2017, with children), West Flemish and Dutch (Veenstra et al., 2018, with children); and Venetian and Italian (Scaltritti et al., 2017, with young adults). Relatedly, two meta-analyses conclude that multilinguals' pairs of spoken languages (hence, language proximity) do not moderate the presence (Adesope et al., 2010, for participants across the lifespan) or absence (Lehtonen et al., 2018, for adults) of a multilingual EF effect.

To explain why some studies find a positive multilingual effect while others do not, researchers have discussed various factors (e.g., Antoniou, 2023; Bialystok, 2017; de Bruin et al., 2021; Kroll & Bialystok, 2013; Lehtonen et al., 2018; Luk et al., 2020; Paap, 2019; i.a., for extensive discussion). First, confounding variables, such as culture, socioeconomic and immigration status, are one concern because multilingualism research is often conducted by comparing (naturally occurring) groups of monolinguals and multilinguals. A related issue is over-control. Deliberately matching the comparison groups on (or statistically adjusting for) measures of fluid intelligence or attention, for instance, might mask a multilingual effect on EF because these constructs share a significant amount of variance with and might be considered inherent parts of EF. Working memory capacity, for example, shares 50%–80% variance with fluid intelligence at the latent level (Shipstead & Engle, 2018). To reflect this, fluid intelligence scores have been sometimes used as dependent measures tapping into EF rather than as background variables that need to be controlled for (e.g., de Abreu et al., 2012).

[1]Mutual intelligibility is the most commonly used and, perhaps, the most complete measure for distinguishing language pairs from dialect pairs (e.g., Tamburelli, 2021). However, a major, practical problem with this criterion is that objective mutual intelligibility measures are simply not available for many pairs of linguistic varieties, including Cypriot and Standard Modern Greek. Thus, the next best available criterion is lexical similarity. Importantly, lexical similarity is a major determinant of mutual intelligibility, with the two measures exhibiting substantial, though not perfect correlations (>.6; e.g., Gooskens & van Heuven, 2021; Tang & van Heuven, 2015; Wichmann, 2020).

Second, the specific characteristics of multilinguals possibly play a role. The adaptive control hypothesis (Green & Abutalebi, 2013), for example, proposes that different sociolinguistic environments place different demands on multilinguals' EF skills. Drawing on this framework, it may be predicted that multilingual EF effects are found only in *single-language* or *dual-language* contexts, where multilinguals use a separate language in different situations of everyday life or with different interlocutors, respectively. This is because multiple EF processes are required for the management of two languages in both a single- (processes such as goal maintenance, monitoring, and interference suppression) and a dual-language context (goal maintenance, monitoring, interference suppression, salient cue detection, selective response inhibition, task engagement and disengagement), even though a dual-language environment places higher and more widespread demands on the EF network.

Language proficiency is another aspect of multilingual experience that is possibly relevant to EF. At the performance level, EF tasks often require subjects to draw on verbal abilities, even though EF is a non-verbal system. This potentially disadvantages (some) multilinguals when performing EF tasks because multilinguals often exhibit lower language performance (e.g., slower lexical access, smaller vocabularies) relative to monolinguals, at least under certain conditions (e.g., when sequential multilinguals are tested in their second language; Bylund et al., 2023). Furthermore, higher proficiency in (the) additional language(s) might pose more demands on and, in turn, more practice of EF during daily communication for multilinguals. For example, higher second-language (L2) proficiency means that individuals have translation equivalent words in two languages for a greater number of lexical concepts. Moreover, high L2 proficiency possibly translates into stronger and more automatic L2 activation compared to low L2 proficiency. In turn, these two considerations suggest more frequent and higher competition between the two languages and, possibly, greater and more frequent EF involvement and training during communication (see Lehtonen et al., 2018, p. 395; Monnier et al., 2022, p. 2232, for similar suggestions at a theoretical level). Finally, an L2 proficiency role may be predicted from neural models that suggest brain adaptations, some of which possibly contribute to domain-general cognition, at high multilingual experience or expertise, including high L2 mastery (e.g., Grundy et al., 2017; Pliatsikas, 2020).

Third, the use of difference scores to measure EF has been criticised for various reasons (e.g., Draheim et al., 2019). The interference effect, for example, is a difference score thought to reflect inhibition and is often calculated in EF tasks, such as the Flanker or Stroop tests (see, e.g., Figures S1 and S2 of Appendix S1, Supplementary Materials Online), by subtracting mean reaction time for congruent or neutral trials from average reaction time for the more difficult, thus, slower incongruent trials. On the one hand, difference scores are useful measures because they isolate EF processing (from other processes such as, for example, processing speed) by subtracting performance in a no or less EF-demanding condition (which, for example, more strongly reflects processing speed) from a (more) EF-demanding condition (which more heavily draws on EF). On the other hand, however, they often show poor reliability. In turn, low reliability decreases statistical power. Relatedly, difference scores from separate tasks which supposedly tap into the same EF process often show weak correlations with each other, indicating low convergent validity. This casts doubt on whether these scores tap into the same process of interest. Moreover, different types of trials in inhibition and switching tasks (e.g., switch and repeat trials) are usually intermixed in the same block. This means that even congruent and repeat trials in mixed blocks involve EF-related processes (e.g., see in Antoniou, 2023; Draheim et al., 2019). Thus, difference scores based on these trials remove EF-related variance. Finally, difference scores are almost always based on reaction times (RTs). However, RT measures (whether mean RT for critical, correct trials or RT difference scores) may be problematic in the presence of speed-accuracy trade-offs (Draheim et al., 2019) – when slow RTs (worse performance) are related to higher accuracy (better performance) or the inverse (faster RTs linked to lower accuracy). For instance, speed-accuracy trade-offs might occur because participants strategically slow their responses to achieve higher accuracy in a task. In this case, however, RT measures are misleading or at least incomplete measures of cognitive ability.

Finally, studies on the multilingual effect often take place in the absence of a clear theory about what EF is and how multilingualism influences this system (e.g., de Bruin et al., 2021). Research that is guided by theory results in increased power to detect a true effect because, for instance, the experimenter will use tasks that are more likely, from a theoretical perspective, to reveal this effect. Also, it leads to a decreased possibility to find a false effect because analyses will be restricted only to theoretically relevant measures.

1.2. The present study

In this context, our study compared multilingual, bidialectal and "monolingual" young adults in the sociolinguistic context of Greek-speaking Cyprus. We focused on a young-adult sample to examine whether past findings with children in the same context (Antoniou et al., 2016) extend to this age group. Cyprus is characterised by diglossia, with most native Greek Cypriots being bidialectal; that is, they typically acquire and use from an early age two Greek dialects in different situations of everyday life. Cypriot Greek (CG) is natively acquired and used for daily oral communication, while Standard Modern Greek (SMG) is mainly learnt through formal education, it is the language of school instruction and literacy; it is used by the media and for formal, public, and official situations (e.g., Antoniou et al., 2016; Antoniou & Katsos, 2017). Thus, theoretically, the diglossic context in Greek-speaking Cyprus resembles a *single-language* context, as described by the adaptive control hypothesis (Green & Abutalebi, 2013)[2].

[2]Given a fair consensus that the sociolinguistic context in Greek-speaking Cyprus is still in some way diglossic (e.g., Arvaniti, 2010; Papapavlou & Pavlou, 1998; Rowe & Grohmann, 2013; Tsiplakou, 2014), our assumption is that, at least theoretically, the two linguistic varieties are used in separate domains of daily life. However, we acknowledge that in reality and practically, the situation may be more complicated. First, the very high similarity of the two linguistic varieties makes it hard to determine what counts as "true" mixing (in the sense of combining, at the performance level, elements from two distinct underlying codes in the same utterance) because, for example, many linguistic elements belong to both varieties (e.g., Tsiplakou, 2009). Relatedly, it has been argued that CG consists of a varietal (or register) continuum that includes less standard-like features at one end and more standard-like characteristics at the other end (e.g., Tsiplakou, 2014; Tsiplakou et al., 2022). The latter CG form (often called a *pancypriot koiné*; e.g., Tsiplakou, 2014) is naturally more mixed, even though it is still distinct from and in diglossic relationship with SMG (as spoken in Cyprus; e.g., Tsiplakou, 2014). Thus, there is a lot of "mixing" when using a higher, formal register of CG (e.g., Tsiplakou, 2014). However, this probably reflects a single, mixed underlying linguistic system (the CG *koiné*) rather than "true" mixing at the performance level (e.g., Tsiplakou et al., 2022). Second, diglossia in Cyprus suggests that, in theory, there is no frequent switching between CG and SMG, from one utterance to another based on the listener, during daily oral

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press

In this research, we predicted better multilingual EF performance that would also be evident in bidialectals. First, bidialectals were recruited from a diglossic sociolinguistic context, where multilingual EF effects might be expected based on the adaptive control hypothesis. The same applies to multilinguals most of who spoke CG and SMG in addition to a non-Greek language. This prediction is further reinforced by a previous study with children in Cyprus showing that both multilinguals and bidialectals had better EF skills than monolinguals (Antoniou et al., 2016). We stress, however, that we did not intend to examine the prediction, based on the adaptive control hypothesis, that different types of contexts (e.g., single-language, dual-language) may have a quantitatively and/or qualitatively different effect on EF. We only predicted that, because multilinguals and bidialectals functioned linguistically in a diglossic, single-language context, they would show (some degree of) better EF performance than "monolinguals". Second, some experimental evidence suggests that there is overlap, at least partial, in the language control processes used by multilingual and bidialectal speakers (e.g., Kirk et al., 2018; but see, e.g., Melinger, 2018). Third, to reiterate, two meta-analyses conclude that multilinguals' pairs of spoken languages (hence, language proximity) do not impact the multilingual cognitive effect (Adesope et al., 2010; Lehtonen et al., 2018). The last two considerations suggest that, if a positive multilingual effect were found, a similar effect should manifest in bidialectals.

We were also interested in adjudicating between the two broad accounts regarding the cognitive locus of the multilingual effect (if found). To achieve this, data from seven cognitive tasks (eight measures) were analysed using the EF model of Miyake et al. (2000)[3]. Thus, if multilingualism has a broader effect on EF, we expected to find group differences in overall EF performance. However, if multilingualism has a selective impact on EF, we expected to find group differences only in one specific EF process.

Finally, we aimed to examine whether certain conditions, particularly (second) language proficiency, might affect the emergence of the multilingual EF effect. We directly examined the effect of SMG vocabulary proficiency because SMG was the language of testing and was also an additional language for bidialectals and for most multilinguals. This impact was investigated in two ways. First, through a statistical analysis on EF that, besides the Group factor (multilinguals, bidialectals and "monolinguals"), included SMG vocabulary proficiency as a predictor variable (see also Section S9 of Appendix S1, Supplementary Materials Online). Second, through a further analysis that compared the EF performance of two groups matched in SMG vocabulary and other background variables: "multilinguals" (comprised of both multilingual and bidialectal participants) and "monolinguals" (see e.g., Papastergiou et al., 2023, for a similar approach). The results of these two analyses in combination could provide evidence that the multilingual and bidialectal EF effect is found in those multilinguals and bidialectals with high, (approximately) similar to monolinguals (second) language proficiency.

communication, at least in conversations among Greek Cypriots (phonetics being the main, reliable indicator of CG-SMG switching, according to Tsiplakou, 2009). However, switching between CG and SMG has been observed in at least some situations. For example, in classroom settings, Greek-Cypriot teachers often use SMG for the "lesson proper" and CG for discourse acts outside this purpose (for instance, for restoring order, telling students off, joking; e.g., Tsiplakou et al., 2022). Another example is conversations in which both Greek-Cypriot and Hellenic Greek interlocutors are involved.

[3]We did not wish to enter the debate on the nature and structure of EF; and to evaluate different models so as to establish which one best represents the organisation of the EF system. Rather, taking an a priori, theoretical approach, we used a specific, influential, and widely employed EF model to analyse our EF data.

2. Method

2.1. Participants

Participants were 50 multilinguals (35 females; aged 18–35 years, mean age 22.5, *SD* 3.9 years), 79 bidialectals (54 females; aged 17–37 years, mean age 21.6, *SD* 3.6 years), and 52 "monolinguals" (42 females; aged 18–38 years, mean age 22.6, *SD* 4 years). All participants lived in Cyprus at the time of testing and were speakers of SMG. Participants' language experience (with CG, SMG, English, and any other language specified by the participants) was further examined through a Language Background and Socioeconomic Status Questionnaire (see in Appendix S2, Supplementary Materials Online)[4]. This inquired (among other topics) about (1) age of onset of exposure to/use of each language/dialect (one item for each language/dialect); (2) general use of each language/dialect, at the production and exposure level, on a scale from *0* – never to *4* – always, in various situations of past and present-day life; specifically, with mother, father, first two siblings, among parents, at school and university education (preschool, primary, junior and senior high school, undergraduate, master's and doctoral level), with friends, for reading, for computer and internet use, and for watching television (21 total items for each language/dialect); and (3) language/dialect proficiency on a scale from *0* – no proficiency to *4* – very good proficiency (one item for each language/dialect). General use for each language/dialect was calculated by taking the average of each participant's responses from the 21 items (those applicable to each participant) asking for the information in (2). This score reflected both the participants' history of overall use and current general use of each language/dialect in production and exposure. Detailed information on participants' language characteristics can be found in Section S1 and Table S1 of Appendix S1, Supplementary Materials Online (SMO).

Multilinguals reported experience with various non-Greek languages including English, Turkish, Russian, Romanian, Swedish, German, French, Spanish, Arabic, Polish, Tagalog, Armenian, Albanian, Bulgarian, Georgian, Serbo-Croatian, Dutch and Filipino. Bidialectals were native Greek Cypriots who spoke CG and SMG, while "monolinguals" were Hellenic Greeks who came to Cyprus for work or studies.

At the group level, bidialectals and "monolinguals" indicated some experience with non-Greek languages. "Monolinguals", as a group, further had some limited CG experience given that they lived in Cyprus at the time of testing (see Section S1 of Appendix S1, SMO). In general, however, multilinguals and bidialectals had, on average, substantially higher and earlier additional-language or

[4]A very short pre-screening questionnaire was also given to participants before testing. This asked for the following information: age, gender, place of birth (for participants, their mother and father), mother's and father's native language, language of instruction at high school, whether participants were studying or had studied a subject related to languages (e.g., English literature), whether they generally used a language other than Greek (CG and SMG) during everyday communication, whether they had colour blindness (or difficulty to see differences in colour), vision difficulty (e.g., wear glasses), hearing difficulty, difficulty in reading or writing, record of epileptic seizure, or other difficulties (e.g., attention deficit hyperactivity disorder, autism spectrum disorder).

dialect experience than "monolinguals" who had relatively low experience with additional languages or dialects (Table S1). Overall, self-reported proficiency in SMG (language of testing) was, on average, very high for all groups. Moreover, multilinguals self-reported a significantly earlier age of onset of exposure to/use of, more overall use of, and higher proficiency in non-Greek languages compared to both bidialectals and "monolinguals" (who did not differ from each other). Multilinguals also self-reported a significantly earlier age of starting to be exposed to/use, and higher overall use of and proficiency in CG than "monolinguals". This suggests that multilinguals, as a group, were also bidialectal to some degree. Finally, bidialectals had significantly higher CG proficiency and use, and an earlier age of starting to be exposed to/use CG than multilinguals and "monolinguals", attesting to their bidialectalism.

2.2. Materials and procedure

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. Participants were administered the following tasks (in SMG): the Flanker (e.g., Fan et al., 2002) and Stroop tasks (see e.g., Unsworth et al., 2012) for inhibition; the Colour-Shape (e.g., Friedman et al., 2008) and Number-Letter tasks (e.g., Karayanidis et al., 2010) for switching; the forward and backward Corsi Blocks (Mueller & Piper, 2014), N-Back (e.g., Jaeggi et al., 2010), and Rotation Span (Foster et al., 2015) tasks for working memory; the Matrix Reasoning sub-test of the Wechsler Abbreviated Scale of Intelligence for non-verbal fluid intelligence (WASI; Wechsler, 1999); and the Mill Hill vocabulary test (Raven et al., 1997). The Flanker, Stroop, Colour-Shape, Number-Letter, and N-Back tasks were designed with E-Prime software (Schneider et al., 2012). Schematic illustrations of these tasks are presented in Figures S1–S5 in Section S2 of Appendix S1, SMO.

The tests were administered in two sessions. One session included the Stroop, Colour-Shape, Corsi Blocks, Rotation Span tasks and the WASI test in five orders, each starting with a different task. The Mill Hill vocabulary test was always administered last in this session. The other session included the Flanker, Number-Letter and N-Back tasks in three orders, each starting with a different task. In this session, we also recorded participants' electrophysiological responses (EEG), but these measures are not reported here.

2.2.1. Language background and socioeconomic status questionnaire

An English version of this questionnaire is given in Appendix S2, SMO. Besides asking for information on participants' language experience (see Section 2.1), the questionnaire also included items regarding the participants' age, gender and education level (from *0* = no university studies to *3* = awarded or pursuing a doctorate degree), among other questions. Moreover, it included three measures of socioeconomic status (SES): Family Affluence Scale (FAS; highest score = *9*; Boyce et al., 2006) and maternal and paternal education levels (from 1 = primary education to *5* = doctoral-level education).

2.2.2. Flanker task

Participants saw a single or five arrows. They were instructed to focus on the direction (left or right) of the (centre) arrow and respond (as fast and accurately as possible) by pressing one of two buttons on a game controller. In the congruent condition, the Flanker arrows pointed in the same direction as the centre arrow. For incongruent trials, the Flanker arrows pointed in the opposite direction relative to the centre arrow. Finally, under the neutral condition, only one arrow appeared. There were four blocks of trials: a practice block with 24 trials (eight for each condition) and three test blocks (presented in random order) with 96 trials each (32 trials for each condition, presented randomly).

2.2.3. Stroop task

Participants saw a colour word ("BLUE", "RED" and "GREEN" in SMG) or a string of asterisks (e.g., *****) in blue, red or green font colour. They had to respond (as fast and accurately as possible) by pressing one of three buttons on a response box depending on the print colour of the word stimulus. There were three test versions, each with different button-colour associations. In the congruent condition, the colour word matched its print colour (e.g., "GREEN" in green colour), while, for incongruent trials, the colour word was incompatible with its font colour (e.g., "BLUE" in red colour). Finally, in the neutral condition, a string of asterisks was presented. The task included a practice block with 18 trials (six for each condition) and two test blocks (presented randomly) with 108 trials each (36 trials for each condition, presented randomly).

2.2.4. Colour-Shape task

Participants initially saw a cue (the letter "Z" or "Y") and, with the cue on display, a triangle or circle in a green or red square then appeared. Based on the cue, they had to decide about the shape (whether the shape was a triangle or a circle; shape task) or the colour (whether the colour was red or green; colour task) of the target stimulus by pressing (as fast and accurately as possible) one of two buttons on a game controller. There were two test versions, each with a different cue-task, button-shape and button-colour association.

In pure blocks, participants had to focus either on the colour or shape only. Mixed blocks included repeat (repeat the task of the previous trial) and switch trials (switch to the other task compared to the previous trial). There were three practice and four test blocks. Practice blocks included two pure blocks (pure colour and pure shape, with 24 trials each) and one mixed block (with 11 repeat and 12 switch trials or the other way around). Test blocks consisted of a pure colour, a pure shape, and two mixed blocks presented in random order. There were 24 trials in each pure block and 24 switch and 23 repeat trials (or the reverse) in each mixed block. Trials within pure blocks appeared in random order. Mixed block trials appeared in a pseudo-randomised order such that there were no more than three consecutive repeat or switch trials.

2.2.5. Number-Letter task

Participants were initially presented with a cue (green, blue, red or orange square) and, with the cue on display, a target number-letter (e.g., "6A") or letter-number pair (e.g., "T7") then appeared. We used the letters "E", "I", "A", "K", "M" and "T" and the numbers "3", "5", "7", "2", "4", and "6". Based on the cue, participants had to decide whether the number was odd or even (number task) or whether the letter was a vowel or a consonant (letter task) by pressing (as fast and accurately as possible) one of two buttons on a game controller. There were two test versions, each with a different cue-task, button-letter and button-number association.

In pure blocks, participants had to focus either on the letter or on the number only. Mixed blocks consisted of switch and repeat trials. There were three practice blocks and six test blocks. Practice blocks included two pure blocks (pure letter and pure number, with 36 trials each) and a mixed block (with 35 repeat and 36 switch trials or the other way around). Test blocks included a pure letter, a pure number, and four mixed blocks presented in random order. Test pure blocks included 72 trials each (presented randomly), while test mixed blocks included 35 switch and 36 repeat trials (or the reverse). Trials within mixed blocks appeared in a pseudo-randomised order such that there were no more than three consecutive repeat or switch trials.

2.2.6. Corsi Blocks task

We used the (forward and backward versions) of the Corsi Blocks task from the Psychology Experiment Building Language (PEBL) battery but adapted for use with Greek-speaking participants. Each sub-test (forward, backward), started with a trial of two and could continue up to trials of nine to-be-remembered boxes. There could be 18 total trials, two for each sequence length (e.g., two trials for the sequence length of two to-be-recalled boxes). Each sub-test was discontinued after two consecutive erroneous trials on the same sequence length.

2.2.7. N-Back task

Participants were presented with a series of (difficult-to-describe) visual images. Under each condition (2-back, 3-back and 4-back), they had to decide whether the image currently on display was identical or not to an image presented *n* (2, 3 or 4) positions before, by pressing (as fast and accurately as possible) one of two buttons on a game controller. There were two test versions, each with a different button-task (identical or not) association. Participants performed three practice blocks (2-back, 3-back and 4-back) and 12 test blocks (four for each *n* level). The 2-back blocks always appeared first and the 4-back condition last. Blocks within each condition were randomly presented.

Target trials were those where, depending on the condition (2-back, 3-back and 4-back), the current stimulus matched a stimulus that appeared *n* positions earlier. Blocks included sequences of 22, 23, or 24 stimuli (under the 2-back, 3-back and 4-back conditions, respectively). Test blocks included a total of 72 target trials, six within each block.

2.2.8. Rotation Span task

We adapted (for use with Greek-speaking participants) the shortened version of the Rotation Span task. For each trial, participants had to keep in mind and then recall a sequence of arrows that were either long or short and pointed in one of eight directions. Before each arrow in the sequence, they had to decide whether a (rotated) letter (the Greek letters "β", "η", "ι", "μ") was in a mirror image or not (distractor task). The number of arrows to remember ranged from two to five. There were 12 trials, three for each sequence length, in three blocks. Each block included one trial from each sequence length, appearing in random order. There was no time limit for responding in the recall phase of each trial. However, in the instructions, participants' mean RT was calculated while practicing the distractor part of the test. Participants were then instructed that, during the main task, they had to keep a high accuracy in the distractor task (≥ 85%); and that if their response for a distractor letter was slower than their practice mean RT, the letter would automatically disappear and their response for that letter would be coded as erroneous. In fact, for the distractor letters, the test allowed responses within 2.5 standard deviations of participants' practice mean RT.

2.2.9. WASI Matrix Reasoning test

This was used as a test of non-verbal fluid intelligence (Tourva & Spanoudis, 2020; Wechsler, 1999). For each item (35 test items), participants saw a matrix from which a piece was missing. They had to find the missing piece from five possible choices. We measured participants' accuracy out of a maximum possible score of 35.

2.2.10. Mill Hill vocabulary scale

In the Definitions sub-test, participants had to write down, using SMG, the meaning of 44 SMG words. Responses were coded as correct if they (1) roughly described the correct meaning of the target word; and/or (2) correctly used the target word in a phrase or sentence in a way that indicated knowledge of its meaning; and/or (3) included a synonym or a related word which showed knowledge of the target word's meaning (e.g., "neighbour" for the target word "near"). Responses were coded as incorrect if knowledge of the target word's meaning was shown by using CG. All responses were scored by two independent raters. Participants' score in this sub-test was calculated as the average of the two coder's scores (each coder score reflecting participants' accuracy out of a maximum of 44).

In the multiple-choice sub-test, participants were given another set of 44 SMG target words and were required to choose the word (out of six options) whose meaning was more related to the target word. We measured participants' accuracy out of a maximum possible score of 44.

3. Results

3.1. Preliminary analyses

3.1.1. Measures

For EF, the main analyses included the following eight target measures: mean RT for correct responses for incongruent trials in the Flanker and Stroop tasks (for inhibition), mean RT for correct responses for switch trials in the Colour-Shape and Number-Letter tasks (for switching), the partial score from each of the forward and backward Corsi Blocks task and the Rotation Span task (e.g., Foster et al., 2015); and the proportion of correct target trials in the N-Back test (for working memory). The partial score is the sum of items (boxes in the Corsi and arrows in the Rotation Span task) recalled correctly and in the correct order across trials (maximum score of 88 for 18 total possible trials in each Corsi sub-test and 42 for 12 trials in the Rotation Span task). For the Mill Hill Definitions sub-test, the two independent coders' scoring showed an intraclass correlation coefficient of 0.94.

For the inhibition and switching tasks, we used mean RTs in the critical incongruent and switch conditions (henceforth, target RT measures) because of the various issues surrounding RT difference scores. Moreover, as illustrated in Table S2 (Appendix S1, SMO), these target RT measures from each task showed negative correlations with accuracy in the same condition and task, suggesting no speed-accuracy trade-offs. Additionally, and more importantly, even though mean RTs for incongruent and switch trials reflect both EF processing and basic processing speed, our experimental design could test and exclude the possibility that a multilingual/bidialectal effect, if found, were due to the latter cognitive component rather than to EF; particularly, if a Group effect on overall EF or in a single specific component (e.g., inhibition) were found, then this would, simultaneously, provide evidence against the alternative explanation that the group difference lies in basic processing speed. This is because processing speed is a common aspect of the RT

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press

Table 1. Bivariate correlations between the target executive function measures

	Stroop RT	Flanker RT	NL RT	CS RT	N-Back	Corsi For.	Corsi Back.	RS
Flanker RT	0.37**							
NL RT	0.56**	0.40**						
CS RT	0.41**	0.22**	0.56**					
N-Back	−0.23**	−0.20*	−0.29**	−0.19*				
Corsi For.	−0.30**	−0.25**	−0.23**	−0.18**	0.24**			
Corsi Back.	−0.31**	−0.25**	−0.28**	−0.21**	0.32**	0.38**		
RS	−0.25**	−0.16*	−0.32**	−0.32**	0.26**	0.34**	0.46**	

*Correlation is significant at the 0.05 level (two-tailed).
**Correlation is significant at the 0.01 level (two-tailed).
Note: RT = mean reaction times for correct trials; Stroop RT = RT for incongruent trials in the Stroop task; Flanker RT = RT for incongruent trials in the Flanker task; NL RT = RT for switch trials in the Number-Letter task; CS RT = RT for switch trials in the Colour-Shape task; Corsi For. = partial score in the forward Corsi Blocks task; Corsi Back. = partial score in the backward Corsi Blocks task; RS = partial score in the Rotation Span task.

measures for both the inhibition and switching component scores. Thus, the alternative explanation would predict a Group by Component interaction in the direction of a Group effect in both RT components but not in only one RT component or in overall EF. Relatedly, overall EF performance reflects the common variance between the target RT and the accuracy EF measures. Thus, it cannot reflect simple processing speed (McVay & Kane, 2012, p. 308)[5].

In any case, in Appendix S1 of the SMO (Sections S5, S6, and S8), we further take up the issue of RT difference scores. We investigate more closely if and the conditions under which RT EF measures – target RT and, especially, difference scores – show higher and/or adequate reliability and convergent validity. We stress, however, that this was only an exploratory investigation in this study. We had strong, a priori reasons to avoid the use of RT difference scores (Draheim et al., 2019). However, we examine and compare the psychometric properties (reliability and convergent validity) for all RT EF measures – target RT and RT difference scores – because we wanted to further empirically motivate why we used the target RT EF measures (mean RT in critical EF conditions) over (some) RT difference scores; and because these properties have been suggested to limit past research on EF in general and on the cognitive effects of multilingualism in particular (e.g., Draheim et al., 2019). Also, this exploratory investigation could provide preliminary results to guide further, dedicated research on the topic. Finally, based on these reliability and convergent validity results for the difference scores, we report additional (group and other) analyses in Appendix S1 of the SMO, with EF composite scores that used difference scores. These analyses provide a further, complementary test of the alternative explanation that a potential multilingual/bidialectal effect lies in processing speed. This is because those difference scores that do not suffer from the limitations outlined in the *Introduction* better isolate and reflect EF processes (see Sections S5, S6, and S8 of Appendix S1, SMO).

[5]Of course, processing speed is known to correlate with and, indeed, may be considered a component of EF; for instance, it may reflect the common aspect of EF in the Miyake et al. (2000) model (see e.g., in Friedman et al., 2008), even though not all researchers and empirical evidence support this hypothesis (e.g., Friedman et al., 2008). In this study, we wanted to exclude the possibility that results may reflect an effect on processing speed because, to our knowledge, different accounts regarding the cognitive locus of a multilingual effect typically do not consider simple processing speed as a cognitive component that is impacted by multilingualism.

Data were missing for various reasons, including participants completing only one of two sessions (five bidialectals, five "monolinguals"), experiment error (e.g., a test was not administered) or, for the RT measures, the participant had (close to) 0% (< 5%) accuracy in the relevant task condition (one multilingual in the Stroop, two multilinguals in the Colour-Shape, and one "monolingual" in the Flanker task). These data were treated as "unavailable" (="non-responses") for that particular test for each participant. However, for each participant, we retained and used their data from the rest of their completed tasks. No other data removal or trimming procedure was applied to the data (Zhou & Krott, 2016).

Descriptive statistics for accuracy and/or RT by EF task and condition, across the three groups, can be found in Table S3 of Appendix S1 in the SMO. Table 1 shows the correlations between the target EF measures and Table S4 (Appendix S1, SMO) reports their reliabilities. Overall, the inhibition and switching tests showed the expected condition effect for RT, in that incongruent and switch trials were significantly slower than the other types of trials within the same task. Moreover, all target measures showed reliabilities close to or above Cronbach's α = 0.8. Correlations between target measures tapping into the same cognitive function were in the small-to-moderate range and all statistically significant.

3.1.2. Confirmatory factor analysis

A Confirmatory factor analysis (CFA) was conducted on z-transformed (and, for switching and inhibition, reverse-scored) measures from each EF task using R version 4.0.3 (R Core Team, 2020) and the package *lavaan* (Rosseel, 2012). Results from this CFA are displayed in Figure 1. This CFA indicates that the three-factor structure provided an excellent fit to the EF data (Hu & Bentler, 1999): $\chi^2(17, n = 165) = 17.17$, $p > 0.05$, $\chi^2/df = 1.01$. Fit indices were as follows: CFI = 0.999, NNFI = 0.999 and RMSEA value = 0.008 (90% confidence interval 0.00–0.07). Moreover, there were significant moderate-to-large correlations between the three factors.

3.1.3. Composite scores

Composite scores were calculated as follows. First, each separate target score from the Flanker, Stroop, Colour-Shape and Number-Letter tasks was reverse-scored by multiplying with −1 so that for all target measures a higher value indicated better performance. Then, for each individual variable, we transformed

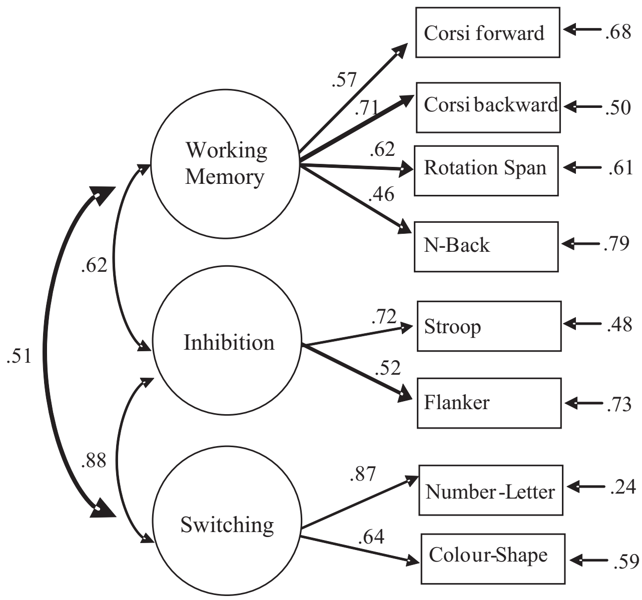

Figure 1. The fitted three-factor structure of executive function.
Note. Numbers at the centre of large, single-headed arrows show factor loadings. Numbers at the back ends of the small arrows indicate error terms. Numbers at the centre of double-headed arrows are correlation coefficients. All numbers are standardised values. Paths are all statistically significant, p <0.01. The Flanker, Stroop, Colour-Shape and Number-Letter target measures were reverse-scored (before the CFA) by multiplying with −1.

participants' scores for the entire sample into z scores. Finally, we averaged the relevant z-transformed measures for each construct. The following composite scores were created: working memory (based on Corsi forward and backward, N-Back and Rotation Span target measures), inhibition (from Flanker and Stroop target RT scores), switching (from Colour-Shape and Number-Letter target RT measures), vocabulary in SMG (from the definitions and multiple-choice parts of the Mill Hill test) and SES (based on FAS score, maternal and paternal education level).

3.1.4. Background variables

Descriptive statistics by group and group comparisons on background variables are presented in Table 2. There were significant group differences only in SMG vocabulary (as expected) in that "monolinguals" had a higher score than the other two groups; and bidialectals exhibited higher performance than multilinguals.

3.2. Main analyses

Table 3 presents descriptive statistics for the target EF measures by Group, EF component and EF task.

To examine the multilingual and bidialectal effect on EF, we performed linear mixed-effects model analyses in RStudio (R Core Team, 2020) using the *lme4* package (Bates et al., 2015). We used linear mixed-effect model analyses to account for the sample size imbalance between groups. For all analyses, we attempted to specify the maximal random-effects structure, but random slopes were eventually dropped because models including them were unidentifiable (Barr et al., 2013).

In all analyses, the first model included Group (*1* = "monolinguals", *2* = bidialectals, *3* = multilinguals) as a between-subjects factor, EF as a within-subjects factor (*1* = inhibition, *2* = switching, *3* = working memory), their interaction, and by-subject random intercepts. The dependent measure was the composite score, with each participant having three data points: one score for each of inhibition, switching and working memory. Statistical significance for the Group effect and the Group by Component interaction that were of interest in this study was determined through model comparison (using likelihood ratio tests), whereby two models that differed in only one aspect (e.g., fixed effect of interest) were compared (Barr et al., 2013). In all analyses, the Group by Component interaction was dropped from the initial comparison model if it was non-significant, so that to examine the Group effect. Any continuous variables that were not already z-transformed because of the way they were constructed (SMG vocabulary, SES) were also z-standardised (Age)[6]. In the case that model comparisons indicated statistically significant results for the Group factor, we performed post-hoc comparisons with the *multcomp* package (Hothorn et al., 2016) to establish which groups (multilinguals, bidialectals) differed from "monolinguals"; specifically, we used Dunnett's test to compare multilinguals and bidialectals to "monolinguals"; and the default single-step method to correct for multiple comparisons.

The first analysis (with Group, Component, and the Group by Component interaction) showed no significant results (Group by Component interaction: $\chi^2(4, n = 181) = 3.75$, $p > 0.05$; Group effect: $\chi^2(2, n = 181) = 3.95$, $p > 0.05$). To examine whether the Group effect depends on SMG vocabulary proficiency, two further analyses were conducted. First, we included SMG vocabulary in the comparison models. From an interpretation perspective, this analysis examines the Group effect when SMG vocabulary is held constant at a fixed value for all participants; that is, it examines the Group effect when, because of covarying SMG vocabulary, neither Group nor EF is related to SMG vocabulary (e.g., Cohen et al., 2002; see also Miller & Chapman, 2001, for a discussion in the context of Analysis of Covariance). In this respect, results can be interpreted as indicating the effect of Group when the three comparison groups have (approximately) similar SMG vocabulary proficiency. Section S9 of Appendix S1 (SMO) further discusses regression analysis and the interpretation of its results when a covariate is related to the Group factor.

This analysis showed a non-significant Group by Component interaction ($\chi^2(4, n = 179) = 3.61$, $p > 0.05$) but a significant Group effect ($\chi^2(2, n = 179) = 9.03$, $p = 0.011$), in the direction of higher multilingual and bidialectal EF performance (estimate = 0.39, $SE = 0.14$, z value = 2.75, $p = 0.011$ and estimate = 0.31, $SE = 0.12$, z value = 2.61, $p = 0.017$, respectively)[7]. Table S7 (Appendix S1, SMO) further presents the results from the initial comparison model in the analysis for the Group effect (with only SMG vocabulary covaried).

To further examine the role of SMG proficiency, we used propensity score matching to form a group of "multilinguals" who were matched to the "monolingual" group in SMG vocabulary and the other background variables (gender, SES, age, education). We did not match for fluid intelligence, because, to reiterate, EF and fluid intelligence are inherently related at a conceptual and empirical level (see *Introduction*). Thus, when examining the Group effect on EF, matching our groups on fluid intelligence or statistically

[6]We note that a z-standardised variable is also mean-centred (e.g., Field, 2013).

[7]Results were similar, in terms of statistical significance, when the other background variables (gender, SES, age and education) were covaried in the analyses.

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press

Table 2. Descriptive statistics (means and standard deviations) for background variables by language group

	"Monolinguals" (*n* = 52)	Bidialectals (*n* = 79)	Multilinguals (*n* = 50)	Statistic
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	
Gender	42 female	54 female	35 female	χ^2(2, *n* = 181) = 2.61, *p* >0.05
Age	22.6 (4)	21.6 (3.6)	22.5 (3.9)	*F*(2, 163) = 1.47, *p* >0.05
Education level	9 postgraduate	9 postgraduate	11 postgraduate	χ^2(2, *n* = 181) = 2.65, *p* >0.05
	0 no degree	0 no degree	2 no degree	Fisher's exact test, *p* >0.05
	43 undergraduate	70 undergraduate	37 undergraduate	
	5 master's	7 master's	10 master's	
	4 doctoral	2 doctoral	1 doctoral	
SES	−0.08 (0.7)	−0.03 (0.7)	0.12 (0.8)	*F*(2, 169) = 1.02, *p* >0.05
SMG vocabulary	0.54 (0.51)	0.004 (0.6)	−0.54 (1.3)	*F*(2, 176) = 21.5, *p* <0.001 Mult. < Monol. Mult. < Bid. Monol. > Bid.
WASI	27.02 (4.3)	28.1 (3.07)	27.6 (3.6)	*F*(2, 174) = 1.37, *p* >0.05

Note: SD = Standard Deviation; SES = Socioeconomic Status composite score; SMG vocabulary = Standard Modern Greek vocabulary composite score; WASI = Matrix Reasoning test from the Wechsler Abbreviated Scale of Intelligence; Mult. = Multilinguals; Bid. = Bidialectals; Monol. = "Monolinguals"; > = group on the left has a significantly higher score than the group on the right (Bonferroni correction applied); < = group on the left has a significantly lower score than the group on the right (Bonferroni correction applied).

adjusting for fluid intelligence by including it as a predictor in a regression analysis would amount to partly matching the groups on EF aspects; and removing part of the EF measure that we were actually interested in explaining (see also Section S9 of Appendix S1, SMO).

Matching was performed using the *Matchit* package (Ho et al., 2011) in RStudio. Multilinguals and bidialectals were treated as one group to ensure that the matched samples had the maximum number of participants (for increased statistical power) and because both groups exhibited a statistically significant EF effect in the previous analysis. Also, descriptively, the size of the raw multilingual and bidialectal effects on overall EF were approximately similar: standardised mean difference (Cohen's *d*) = 0.35 and 0.32, respectively (Table 3; see also Table S9 of Appendix S1, SMO). We used the optimal matching algorithm and 1:1 matching (with "multilinguals" as control and "monolinguals" as the treatment group). To match the two groups for missing data, we used the missing indicator approach (Rosenbaum, 2010).

The resulting matched samples included 52 "multilinguals" (34 bidialectals; 39 female; aged 17–36 years, mean age 22.2, *SD* 4.3 years) and 52 "monolinguals"[8] (42 female; aged 18–38 years, mean age 22.6, *SD* 3.96 years). Overall, there were no significant differences between the matched samples in any background factor, including SMG vocabulary and fluid intelligence (Table 4).

Table 5 presents descriptive statistics for the composite EF scores for the matched samples. All matching variables (SMG vocabulary, gender, SES, age, education) were covaried in the group analyses to account for potential residual imbalance after matching (e.g., Stuart & Rubin, 2008). Comparison models showed a non-significant Group by Component interaction (χ^2(2, *n* = 97) = 0.011, *p* > 0.05) but a significant Group effect (χ^2(1, *n* = 97) = 6.75, *p* = 0.009) in the direction of higher "multilingual" EF performance. The "multilingual" effect in this analysis remained significant even when not covarying SMG vocabulary (χ^2(1, *n* = 97) = 5.38, *p* = 0.02). Table S8 (Appendix S1, SMO) further presents the results from the initial comparison model in the analysis with the matched samples for the Group effect (with SMG vocabulary and all other background variables covaried).

Finally, Tables S12 and S13 (Appendix S1, SMO) report further correlational analyses between the EF measures and various multilingual and bidialectal experiences measured continuously across groups. Table S12 shows the correlations with English- and other-language experience (age of onset of exposure/use, overall use, and proficiency, as measured with the self-report questionnaire) across all participants, while Table S13 presents the correlations with SMG experience (age of onset of exposure/use and overall use, as measured with the questionnaire; and SMG proficiency measured with the vocabulary test) across multilinguals and bidialectals. These analyses, however, were only exploratory in nature and complementary to the group comparisons because our original and main aim was to examine the multilingual and bidialectal impact by comparing groups; and to investigate whether such potential group effect depends specifically on SMG vocabulary proficiency.

There were significant small correlations only between self-reported English proficiency and the Inhibition scores (Spearman's ρ = 0.22, *p*(two-tailed) < 0.05, for Inhibition based on target RT measures; Spearman's ρ = 0.22, *p*(two-tailed) < 0.05, for Inhibition based on a target RT and a difference score); and between age of onset of exposure to/use of English and the Inhibition score based on the target RT measures (Spearman's ρ = − 0.17, *p*(two-tailed) = 0.049). Significant correlations indicated better inhibition with increased English experience.

4. Discussion

This study investigated whether multilingualism and bidialectalism have a positive effect on EF. EF was examined through eight measures from seven tasks (approximately 463 critical trials per

[8]The group samples in the following analyses, however, were slightly smaller (e.g., 49 "multilinguals", 48 "monolinguals", for the group effect on EF) because of missing values for some covariates.

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press

Table 3. Descriptive statistics (raw values) for target executive function measures by Group, executive function component, and executive function task

		"Monolinguals"	Bidialectals	Multilinguals	Mean difference: Unstandardised [95% CIs], Standardised [95% CIs]
Task	Measure	Mean [95% CIs]	Mean [95% CIs]	Mean [95% CIs]	
	Inhibition[1]	−0.17 [−0.41, 0.06]	0.02 [−0.16, 0.20]	0.11 [−0.13, 0.35]	Mult.-Monol.:0.29 [−0.05, 0.62], 0.33 [−0.06, 0.73] Mult.-Bid.:0.09 [−0.20, 0.38], 0.11 [−0.25, 0.47] Bid.-Monol.:0.19 [−0.097, 0.49], 0.23 [−0.12, 0.59]
	Switching[1]	−0.16 [−0.35, 0.04]	0.13 [−0.06, 0.31]	−0.03 [−0.33, 0.27]	Mult-Monol.:0.12 [−0.23, 0.48], 0.13 [−0.26, 0.53] Mult.-Bid.: −0.16 [−0.49, 0.17], −0.17 [−0.53, 0.19] Bid.-Monol.:0.28 [0.002, 0.56], 0.35 [0.00, 0.72]
	Working Memory	−0.13 [−0.33, 0.07]	0.01 [−0.15, 0.17]	0.10 [−0.08, 0.29]	Mult.-Monol.:0.24 [−0.04, 0.51], 0.34 [−0.05, 0.74] Mult.-Bid.:0.09 [−0.16, 0.34], 0.13 [−0.22, 0.49] Bid.-Monol.:0.14 [−0.11, 0.40], 0.20 [−0.15, 0.56]
	Executive Function[1]	−0.15 [−0.32, 0.02]	0.04 [−0.10, 0.17]	0.07 [−0.11, 0.26]	Mult.-Monol.:0.22 [−0.03, 0.47], 0.35 [−0.04, 0.75] Mult.-Bid.:0.03 [−0.19, 0.26], 0.05 [−0.31, 0.41] Bid.-Monol.:0.19 [−0.02, 0.41], 0.32 [−0.03, 0.68]
Stroop	Incongruent RT	842 [790, 894]	814 [774, 854]	794 [737, 850]	
	Incongruent Acc.	0.95 [0.92, 0.99]	0.96 [0.95, 0.97]	0.95 [0.92, 0.98]	
	Congruent RT	716 [675, 757]	692 [663, 721]	671 [634, 709]	
	Congruent Acc.	0.98 [0.97, 0.99]	0.98 [0.98, 0.99]	0.97 [0.94, 0.997]	
	Neutral RT	719 [683, 756]	695 [667, 722]	686 [649, 724]	
	Neutral Acc.	0.98 [0.97, 0.98]	0.98 [0.97, 0.98]	0.96 [0.93, 0.99]	
Flanker	Incongruent RT	623 [595, 651]	601 [578, 623]	597 [569, 624]	
	Incongruent Acc.	0.98 [0.97, 0.99]	0.97 [0.96, 0.98]	0.97 [0.97, 0.98]	
	Congruent RT	559 [535, 583]	545 [526, 565]	537 [516, 559]	
	Congruent Acc.	0.99 [0.99, 0.996]	0.99 [0.99, 0.995]	0.99 [0.99, 0.99]	
	Neutral RT	545 [525, 565]	531 [513, 548]	524 [504, 544]	
	Neutral Acc.	0.99 [0.98, 0.99]	0.99 [0.98, 0.99]	0.99 [0.99, 0.99]	
Colour-Shape	Switch RT	1882 [1778, 1985]	1645 [1556, 1735]	1838 [1681, 1994]	
	Switch Acc.	0.87 [0.82, 0.92]	0.88 [0.85, 0.91]	0.87 [0.83, 0.91]	
	Repeat RT	1365 [1268, 1463]	1201 [1126, 1276]	1196 [1098, 1293]	
	Repeat Acc.	0.899 [0.85, 0.95]	0.91 [0.88, 0.95]	0.91 [0.87, 0.95]	
	Pure RT	1008 [925, 1090]	932 [850, 1014]	869 [779, 959]	
	Pure Acc.	0.92 [0.87, 0.97]	0.92 [0.88, 0.96]	0.91 [0.86, 0.97]	
Number-Letter	Switch RT	1253 [1151, 1356]	1237 [1147, 1327]	1202 [1069, 1335]	
	Switch Acc.	0.93 [0.896, 0.96]	0.92 [0.899, 0.93]	0.91 [0.86, 0.94]	
	Repeat RT	1089 [1000, 1177]	1012 [941, 1083]	993 [875, 1110]	
	Repeat Acc.	0.94 [0.91, 0.97]	0.94 [0.92, 0.95]	0.94 [0.91, 0.96]	
	Pure RT	775 [728, 822]	733 [700, 767]	729 [668, 790]	
	Pure Acc.	0.98 [0.97, 0.98]	0.95 [0.93, 0.97]	0.95 [0.92, 0.98]	
Corsi Forward	Partial score	41.9 [38, 45.6]	44 [41.2, 46.8]	47.1 [43.5, 50.7]	
Corsi Backward	Partial score	45.2 [42.2, 48.2]	46.7 [43.9, 49.4]	48 [44.6, 51.4]	
Rotation Span	Partial score	24.8 [22.5, 27.1]	25.9 [24.4, 27.4]	26.3 [24.3, 28.3]	
	Distractor Acc.	0.91 [0.88, 0.94]	0.92 [0.90, 0.94]	0.92 [0.89, 0.94]	
N-Back	Target Acc.	0.52 [0.48, 0.56]	0.55 [0.52, 0.59]	0.54 [0.496, 0.59]	
	Lure Acc.	0.71 [0.68, 0.74]	0.73 [0.71, 0.76]	0.697 [0.66, 0.74]	
	Non-target Acc.	0.84 [0.82, 0.87]	0.87 [0.85, 0.89]	0.83 [0.798, 0.87]	

Note: CIs = Confidence Intervals; Acc. = Proportion correct; Mult. = Multilinguals; Bid. = Bidialectals; Monol. = "Monolinguals".
[1]Composite scores were calculated from the reverse-scored (multiplied with −1) and *z*-transformed individual target measures (from the inhibition and switching tasks) so that a higher value indicates better performance. The executive function composite score was calculated by averaging all individual *z*-transformed (and reverse-scored where appropriate) target executive function measures (see also section *Composite Scores*). The latter score was not used in the main analyses. It is reported here for descriptive purposes because group differences between multilinguals, bidialectals, and "monolinguals" were found across the three executive function components (=in overall executive function); and this helps quantify the magnitude of the raw group differences across executive function processes (=in overall executive function).

Table 4. Descriptive statistics (means and standard deviations) for background variables for the matched samples

	"Monolinguals"	"Multilinguals"	Statistic
	Mean (*SD*)	Mean (*SD*)	
Gender	42 Female	39 Female	$\chi^2(1, n = 104) = 0.50, p > 0.05$
Age	22.6 (3.96)	22.2 (4.3)	$F(1, 101) = 0.31, p > 0.05)$
Education level	9 Postgraduate	9 Postgraduate	$\chi^2(1, n = 104) = 0, p > 0.05$
	0 no degree	1 No degree	Fisher's exact test, $p > 0.05$
	43 Undergraduate	42 Undergraduate	
	5 Master's	7 Master's	
	4 Doctoral	2 Doctoral	
SES	−0.08 (0.7)	0.06 (0.6)	$F(1, 95) = 1.11, p > 0.05$
SMG vocabulary	0.54 (0.51)	0.44 (0.45)	$F(1, 101) = 1.15, p > 0.05$
WASI	27.02 (4.3)	27.8 (3.5)	$F(1, 99) = 1.06, p > 0.05$

Note: SD = Standard Deviation; SES = Socioeconomic Status composite score; SMG vocabulary = Standard Modern Greek vocabulary composite score; WASI = Matrix Reasoning test from the Wechsler Abbreviated Scale of Intelligence.

Table 5. Descriptive statistics for the composite executive function scores for the matched samples

	I"Monolinguals" (*n* = 52)	I"Multilinguals" (*n* = 52)	IMean difference: Unstandardised [95% CIs], Standardised [95% CIs]
	Mean [95% CIs]	Mean [95% CIs]	
Inhibition[1]	−0.17 [−0.41, 0.06]	0.11 [−0.11, 0.33]	Mult.-Monol.:0.28 [−0.04, 0.60], 0.34 [−0.04, 0.74]
Switching[1]	−0.16 [−0.35, 0.04]	0.10 [−0.13, 0.33]	Mult.-Monol.:0.26 [−0.04, 0.56], 0.33 [−0.05, 0.73]
Working Memory	−0.13 [−0.33, 0.07]	0.11 [−0.09, 0.30]	Mult.-Monol.:0.24 [−0.04, 0.52], 0.33 [−0.05, 0.73]
Executive Function[1]	−0.15 [−0.32, 0.02]	0.10 [−0.06, 0.27]	Mult.-Monol.:0.26 [0.02, 0.50], 0.42 [0.03, 0.82]

Note: CIs = Confidence Intervals; Mult. = "Multilinguals" (multilinguals and bidialectals); Monol. = "Monolinguals".

[1]Composite scores were calculated from the reverse-scored (multiplied with −1) and *z*-transformed individual target measures (from the inhibition and switching tasks) so that a higher value indicates better performance. The executive function composite score was calculated by averaging all individual *z*-transformed (and reverse-scored as appropriate) target executive function measures (see also section *Composite Scores*). The latter score was not used in the main analyses. It is reported here for descriptive purposes because group differences between "multilinguals" and "monolinguals" were found across the three executive function components (=in overall executive function); and this helps quantify the magnitude of the raw group differences across executive function processes (=in overall executive function).

participant[9]). Moreover, we analysed EF with three processes (inhibition, switching and working memory) using the theoretical model of Miyake et al. (2000) according to which executive processes are moderately related but partly separable.

First, we found some evidence for better EF in multilinguals and bidialectals; specifically, in the analysis that covaried SMG vocabulary proficiency. Second, in the same analysis, both the multilingual and bidialectal effects were evident in overall EF. Equally, an effect on overall EF was found in the analysis with the matched "multilingual" (including both multilingual and bidialectal participants) and "monolingual" groups. These findings support accounts that suggest that multilingualism affects general aspects of the EF system rather than specific EF processes (e.g., Bialystok, 2017). Furthermore, the raw multilingual and bidialectal effects on overall EF were descriptively similar in size (Table 3 and Table S9 of Appendix S1, SMO). These results suggest that the EF effect is quantitatively (in size) and qualitatively (in terms of EF aspects impacted) similar for multilingual speakers of completely different languages and bidialectal speakers of very similar dialects.

Additionally, group differences in EF were evident only when covarying in the group comparison analysis or when matching groups of "multilinguals" (multilinguals and bidialectals) and "monolinguals" on SMG vocabulary proficiency. Our interpretation is that better EF is evident in multilinguals and bidialectals with high vocabulary proficiency. EF tasks often rely on verbal abilities, despite EF being a non-verbal cognitive system. Hence, multilinguals and bidialectals with high proficiency in the language of testing might better cope with these verbal demands at the performance level and exhibit superior EF. Moreover, multilinguals and bidialectals with high vocabulary in an additional-language/dialect possibly experience more competition from the non-relevant language/dialect when communicating. This greater interference possibly occurs because high L2 multilinguals and bidialectals have more translation equivalent words in their two languages/dialects and/or because their L2 is more automatically and strongly activated. In turn, this might lead to more frequent, greater demands on and, consequently, more training of EF (see Lehtonen et al., 2018, p. 395; Monnier et al., 2022, p. 2232, for similar suggestions at a theoretical level).

The group differences in our study could not be explained by the most discussed confounding variables in the literature. First, gender, SES, age and education were directly measured and excluded as potential confounds. For immigration status, both the multilingual and "monolingual" groups included participants with an immigrant background, yet multilinguals outperformed "monolinguals". Finally, cultural differences between bidialectals and "monolinguals" were minimal: Cyprus and Greece are in very close geographic proximity and have strong historic, ethnic, religious, political, economic, and cultural ties. Yet, again, a bidialectal cognitive effect emerged.

That said, because of the cross-sectional nature of the present study, our conclusion that multilingualism and bidialectalism have a positive effect on EF hinges entirely on the assumption that there are no other hidden confounds. Our results cannot be explained by the most-cited possible confounds but this list may not be exhaustive (e.g., Paap, 2019). Thus, we consider our conclusion as the best available interpretation of our results, given our hypotheses and the background factors examined. However, directly controlling for a

[9]We say *approximately* because, for the Corsi Blocks test, there was a discontinue rule. Thus, not all items were presented to all participants.

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press

larger number of background variables or employing alternative experimental methods, such as longitudinal research, would arguably make a stronger case for a positive multilingual and bidialectal EF effect (e.g., Antoniou, 2023).

A second limitation of our study is that, due to its focus on young adults, unavoidably, "monolinguals" as a group had some additional-language and dialect experience; and, similarly, bidialectals, on average, had some experience with languages other than the two Greek dialects. Relatedly, this research mainly aimed to examine the multilingual and bidialectal effects by comparing groups of multilinguals, bidialectals and "monolinguals". As a result, certain aspects of multilingual/bidialectal experience, particularly proficiency in additional languages (apart from SMG), were only crudely and unreliably measured through a single, self-report questionnaire item. Thus, the correlations between these specific multilingual/bidialectal experiences (measured continuously, across groups) and EF were only exploratory in nature and results should be taken with caution. That said, it is important to note that all self-report questionnaire scores (age of onset of exposure/use, overall use, proficiency) were sensitive enough to reveal group differences in multilingual/bidialectal experience in the expected direction; specifically, multilinguals had more additional-language experience than both bidialectals and "monolinguals"; and bidialectals had more CG experience than the other two groups (and equal to "monolinguals" additional-language experience). The only exception was self-reported SMG proficiency for which (almost) ceiling responses were given by the vast majority of participants. Crucially, however, SMG proficiency was further measured through a direct vocabulary test; and group differences were obtained with this measure, again, in an expected direction. Additionally, the self-report English proficiency score (along with age of onset of exposure to/use of English) did show a positive correlation with the Inhibition composite score, in line with a multilingual effect.

A third limitation is that two of our EF tasks (Stroop and Number-Letter) required, to some degree, verbal processing. Thus, it is ambiguous from our results whether the role of SMG vocabulary in the emergence of the multilingual and bidialectal effects is due to the verbal demands of some EF tasks, given that multilinguals and bidialectals exhibited lower performance in the language of testing (SMG); or whether it reflects a positive modulatory effect of increased L2 (SMG) proficiency. Finally, it is unclear whether our findings for the multilingual/bidialectal (group or continuous) effects would be different, particularly, in terms of effect size and/or cognitive locus, had we (1) used different participant samples with no (if at all possible) or more limited additional-language experience for bidialectals; or with no or more limited additional-language and additional-dialect experience for "monolinguals"; (2) used more reliable and fine-grained measures of additional-language proficiency (for languages besides SMG); and (3) employed exclusively non-verbal EF tasks. It is reassuring, however, that despite these limitations our results revealed multilingual and bidialectal effects in the hypothesised directions.

Many studies failed to find evidence for better EF in multilinguals (e.g., Paap, 2019). One interpretation of these null results is that the multilingual cognitive effect does not generalise to all multilinguals and/or under all conditions. In this study, we identified vocabulary proficiency as one specific aspect of multilingual experience that moderates the multilingual effect. Moreover, the sociolinguistic context of language use possibly makes a difference. Our study took place in the diglossic, single-language context of Greek-speaking Cyprus, where CG and SMG are widely used, each in a separate domain of daily life. This sociolinguistic situation, as well as dual-language environments, in which multilinguals use a separate language and frequently switch but do not mix languages based on the interlocutor, are two contexts that, theoretically, might lead to multilingual effects on EF (drawing on Green & Abutalebi, 2013). Cyprus is also characterised by widespread bidialectalism, given that the vast majority of the native Greek-speaking population learns from a young age and speaks both CG and SMG. Thus, while it is not possible from this study to generalise results beyond the Cypriot context, a hypothesis for future research is that the multilingual effect might be evident in environments characterised by widespread multilingualism and separation of language use but might not manifest in other sociolinguistic situations (e.g., Costa et al., 2009; Green & Abutalebi, 2013).

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/S1366728924000506.

Data availability statement. Data for this study are available by emailing the first author.

Competing interest. The authors declare none.

References

Antoniou, K. (2023). The ups and downs of bilingualism: A review of the literature on executive control using event-related potentials. *Psychonomic Bulletin & Review*, **30**, 1187–1226. https://doi.org/10.3758/s13423-023-02245-x

Antoniou, K., & **Katsos, N.** (2017). The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics*, **38**, 787–833. http://doi.org/10.1017/S014271641600045X

Antoniou, K., & **Milaki, E.** (2021). Irony comprehension in bidialectal speakers. *The Modern Language Journal*, **105**, 697–719. https://doi.org/10.1111/modl.12724

Antoniou, K., **Grohmann, K. K.**, **Kambanaros, M.**, & **Katsos, N.** (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*, **149**, 18–30. http://doi.org/10.1016/j.cognition.2015.12.002

Adesope, O. O., **Lavin, T.**, **Thompson, T.**, & **Ungerleider, C.** (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, **80**, 207–245. https://doi.org/10.3102/0034654310368803

Antón, E., **García, Y. F.**, **Carreiras, M.**, & **Duñabeitia, J. A.** (2016). Does bilingualism shape inhibitory control in the elderly? *Journal of Memory and Language*, **90**, 147–160. https://doi.org/10.1016/j.jml.2016.04.007

Arvaniti, A. (2010). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Mediterranean Language Review*, **17**, 15–45.

Barr, D. J., **Levy, R.**, **Scheepers, C.**, & **Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, **143**, 233. https://doi.org/10.1037/bul0000099

Blom, E., **Boerma, T.**, **Bosma, E.**, **Cornips, L.**, & **Everaert, E.** (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology*, **8**, 552. https://doi.org/10.3389/fpsyg.2017.00552

Boyce, W., **Torsheim, T.**, **Currie, C.** & **Zambon A** (2006). The family affluence scale as a measure of national wealth: validation of an adolescent self-report measure. *Social Indicators Research*, **78**, 473–487. https://doi.org/10.1007/s11205-005-1607-6

Bylund, E., **Antfolk, J.**, **Abrahamsson, N.**, **Olstad, A. M. H.**, **Norrman, G.**, & **Lehtonen, M.** (2023). Does bilingualism come with linguistic costs? A meta-analytic review of the bilingual lexical deficit. *Psychonomic Bulletin & Review*, **30**, 897–913. http://doi.org/10.3758/s13423-022-02136-7

Cohen, J., **Cohen, P.**, **West, S.G.**, & **Aiken, L.S.** (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Routledge.

Costa, A., **Hernández, M.**, **Costa-Faidella, J.**, & **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

De Abreu, E. P. M., **Cruz-Santos, A.**, **Tourinho, C. J.**, **Martin, R.**, & **Bialystok, E.** (2012). Bilingualism enriches the poor: Enhanced cognitive control in low-income minority children. *Psychological Science*, **23**, 1364–1371. https://doi.org/10.1177/0956797612443836

De Bruin, A., **Dick, A. S.**, & **Carreiras, M.** (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language*, **2**, 433–451. https://doi.org/10.1162/nol_a_00038

De Bruin, A., **Treccani, B.**, & **Della Sala, S.** (2015). Cognitive advantage in bilingualism: An example of publication bias?. *Psychological Science*, **26**, 99–107. https://doi.org/10.1177/0956797614557866

Degirmenci, M. G., **Grossmann, J. A.**, **Meyer, P.**, & **Teichmann, B.** (2022). The role of bilingualism in executive functions in healthy older adults: A systematic review. *International Journal of Bilingualism*, **26**, 426–449. http://doi.org/10.1177/13670069211051291

Donnelly, S., **Brooks, P. J.**, & **Homer, B. D.** (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, **26**, 1122–1147. https://doi.org/10.3758/s13423-019-01567-z

Draheim, C., **Mashburn, C. A.**, **Martin, J. D.**, & **Engle, R. W.** (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, **145**, 508. https://doi.org/10.1037/bul0000192

Dyen, I., **Kruskal, J. B.**, & **Black, P.** (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society*, **82**, 132. https://doi.org/10.2307/1006517

Eberhard, D. M., **Simons, G.F.** & **Fennig, C.D.** (2023). *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Fan, J., **McCandliss, B. D.**, **Sommer, T.**, **Raz, A.**, & **Posner, M. I.** (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience*, **14**, 340–347. https://doi.org/10.1162/089892902317361886

Field, A. (2013). *Discovering Statistics Using SPSS* (4^{th} ed.). London: Sage.

Foster, J. L., **Shipstead, Z.**, **Harrison, T. L.**, **Hicks, K. L.**, **Redick, T. S.**, & **Engle, R. W.** (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, **43**, 226–236. https://doi.org/10.3758/s13421-014-0461-7

Friedman, N. P., **Miyake, A.**, **Young, S. E.**, **DeFries, J. C.**, **Corley, R. P.**, & **Hewitt, J. K.** (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, **137**, 201. https://doi.org/10.1037/0096-3445.137.2.201

Garraffa, M., **Obregon, M.**, & **Sorace, A.** (2017). Linguistic and cognitive effects of bilingualism with regional minority languages: a study of sardinian–italian adult speakers. *Frontiers in Psychology*, **8**, 1907. https://doi.org/10.3389/fpsyg.2017.01907

Gathercole, V. C. M., **Thomas, E. M.**, **Kennedy, I.**, **Prys, C.**, **Young, N.**, **Viñas Guasch, N.**, **Roberts, E. J.**, **Hughes, E. K.**, & **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**, 11. http://doi.org/10.3389/fpsyg.2014.00011

Gooskens, C., & **Van Heuven, V. J.** (2021). Mutual intelligibility. In M. Zampieri & P. Nakov (Eds.), *Similar Languages, Varieties, and Dialects: A Computational Perspective* (pp. 51–95). New York: Cambridge University Press.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**, 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & **Abutalebi, J.** (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**, 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F., & **Li, P.** (2013). *The Psycholinguistics of Bilingualism*. New York: Wiley-Blackwell.

Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, **4**, 177–199. http://doi.org/10.1007/s41809-020-00062-5

Grundy, J. G., **Anderson, J. A.**, & **Bialystok, E.** (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Annals of the New York Academy of Sciences*, **1396**, 183–201. http://doi.org/10.1111/nyas.13333

Grundy, J. G., & **Timmer, K.** (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, **33**, 325–340. https://doi.org/10.1177/0267658316678286

Gunnerud, H. L., **Ten Braak, D.**, **Reikerås, E. K. L.**, **Donolato, E.**, & **Melby-Lervåg, M.** (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, **146**, 1059. https://doi.org/10.1037/bul0000301

Ho, D. E., **Imai, K.**, **King, G.**, & **Stuart, E. A.** (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, **42**. https://doi.org/10.18637/jss.v042.i08

Hothorn, T., **Bretz, F.**, **Westfall, P.**, **Heiberger, R. M.**, **Schuetzenmeister, A.**, **Scheibe, S.**, & **Hothorn, M. T.** (2016). *Package 'multcomp'. Simultaneous inference in general parametric models*. Project for Statistical Computing: Vienna, Austria.

Hu, L., & **Bentler, P. M.** (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1–55. https://doi.org/10.1080/10705519909540118

Houtzager, N., **Lowie, W.**, **Sprenger, S.**, & **De Bot, K.** (2017). A bilingual advantage in task switching? Age-related differences between German monolinguals and Dutch-Frisian bilinguals. *Bilingualism: Language and Cognition*, **20**, 69–79. https://doi.org/10.1017/S1366728915000498

Jaeggi, S. M., **Buschkuehl, M.**, **Perrig, W. J.**, & **Meier, B.** (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, **18**, 394–412. https://doi.org/10.1080/09658211003702171

Karayanidis, F., **Jamadar, S.**, **Ruge, H.**, **Phillips, N.**, **Heathcote, A.**, & **Forstmann, B. U.** (2010). Advance preparation in task-switching: converging evidence from behavioral, brain activation, and model-based approaches. *Frontiers in Psychology*, **1**, 25. https://doi.org/10.3389/fpsyg.2010.00025

Karr, J. E., **Areshenkoff, C. N.**, **Rast, P.**, **Hofer, S. M.**, **Iverson, G. L.**, & **Garcia-Barrera, M. A.** (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, **144**, 1147. https://doi.org/10.1037/bul0000160

Kirk, N. W., **Fiala, L.**, **Scott-Brown, K. C.**, & **Kempe, V.** (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, **26**, 640–648. https://doi.org/10.1080/20445911.2014.929580

Kirk, N. W., **Kempe, V.**, **Scott-Brown, K. C.**, **Philipp, A.**, & **Declerck, M.** (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition*, **170**, 164–178. https://doi.org/10.1016/j.cognition.2017.10.001

Kroll, J. F., & **Bialystok, E.** (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, **25**, 497–514. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., **Dussias, P. E.**, **Bogulski, C. A.**, & **Valdes Kroff, J. R.** (2012). Juggling Two Languages in One Mind: What Bilinguals Tell Us About Language Processing and its Consequences for Cognition. In B. Ross (Ed.), *Psychology of Learning and Motivation-Advances in Research and Theory* (Vol. **56**, pp. 229–262). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Lehtonen, M., **Soveri, A.**, **Laine, A.**, **Järvenpää, J.**, **De Bruin, A.**, & **Antfolk, J.** (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*, **144**, 394. https://doi.org/10.1037/bul0000142

Lowe, C. J., **Cho, I.**, **Goldsmith, S. F.**, & **Morton, J. B.** (2021). The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological Science*, **32**, 1115–1146. https://doi.org/10.1177/0956797621993108

Luk, G., **Anderson, J. A.**, **Grundy, J. G.**, **Pliatsikas, C.**, & **Rothman, J.** (2020). Letter to the Editors of Psychological Science: What Have We Learned About Bilingualism? Regarding Nichols et al.(2020). *Psychological Science*. 10.25384/SAGE.15180460.v1

Miller, G. A., & **Chapman, J. P.** (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, **110**, 40. https://doi.org/10.1037/0021-843X.110.1.40

Melinger, A. (2018). Distinguishing languages from dialects: a litmus test using the picture-word interference task. *Cognition*, **172**, 73–88. http://doi.org/10.1016/j.cognition.2017.12.006

McVay, J. C., & **Kane, M. J.** (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, **141**, 302. https://doi.org/10.1037/a0025250

Miyake, A., **Friedman, N. P.**, **Emerson, M. J.**, **Witzki, A. H.**, **Howerter, A.**, & **Wager, T. D.** (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. https://doi.org/10.1006/cogp.1999.0734

Monnier, C., **Boiché, J.**, **Armandon, P.**, **Baudoin, S.**, & **Bellocchi, S.** (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism*, **25**, 2229–2255. https://doi.org/10.1080/13670050.2021.1908220

Mueller, S. T., & **Piper, B. J.** (2014). The Psychology Experiment Building Language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, **222**, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Nichols, E. S., **Wild, C. J.**, **Stojanoski, B.**, **Battista, M. E.**, & **Owen, A. M.** (2020). Bilingualism affords no general cognitive advantages: A population study of executive function in 11,000 people. *Psychological Science*, **31**, 548–567. http://doi.org/10.1177/0956797620903113

Paap, K. R. (2019). The bilingual advantage debate: Quantity and quality of the evidence. In J. W. Schwieter (Ed.). *The handbook of the neuroscience of multilingualism* (pp. 701–735). West Sussex, UK: Wiley-Blackwell. https://doi.org/10.1002/9781119387725.ch34

Paap, K. R., **Johnson, H. A.**, & **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. http://doi.org/10.1016/j.cortex.2015.04.014

Papapavlou, A. N., & **Pavlou, P.** (1998). A review of the sociolinguistic aspects of the Greek Cypriot dialect. *Journal of Multilingual and Multicultural Development*, **19**, 212–220. http://doi.org/10.1080/01434639808666353

Papastergiou, A., **Sanoudaki, E.**, **Tamburelli, M.**, & **Chondrogianni, V.** (2023). A study on the executive functioning skills of Greek–English bilingual children–a nearest neighbour approach. *Bilingualism: Language and Cognition*, **26**, 78–94. http://doi.org/10.1017/S1366728922000335

Planckaert, N., **Duyck, W.**, & **Woumans, E.** (2023). Is there a cognitive advantage in inhibition and switching for bilingual children? A systematic review. *Frontiers in Psychology*, **14**, 1191816. http://doi.org/10.3389/fpsyg.2023.1191816

Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The dynamic restructuring model. *Bilingualism: Language and Cognition*, **23**, 459–471. http://doi.org/10.1017/S1366728919000130

Raven, J., **Raven, J. C.**, & **Court, J. H.** (1997). *Raven Manual: Section 5*. San Antonio, TX: Harcourt Assessment.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer. https://doi.org/10.1007/978-1-4419-1213-8

Rowe, C., & **Grohmann, K. K.** (2013). Discrete bilectalism: Towards co-overt prestige and diglossic shift in Cyprus. *International Journal of the Sociology of Language*, **224**, 119–142. http://doi.org/10.1515/ijsl-2013-0058

Ross, J., & **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

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, **48**, 1–36. https://doi.org/10.18637/jss.v048.i02

Santillán, J., & **Khurana, A.** (2018). Developmental associations between bilingual experience and inhibitory control trajectories in Head Start children. *Developmental Science*, **21**, e12624. https://doi.org/10.1111/desc.12624

Scaltritti, M., **Peressotti, F.**, & **Miozzo, M.** (2017). Bilingual advantage and language switch: What's the linkage?. *Bilingualism: Language and Cognition*, **20**, 80–97. https://doi.org/10.1017/S1366728915000565

Schneider, W., **Eschman, A.**, & **Zuccolotto, A.** (2012). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.

Shipstead, Z., & **Engle, R.** (2018). Mechanisms of working memory capacity and fluid intelligence and their common dependence on executive attention. In R. Sternberg (Ed.), *The Nature of Human Intelligence* (pp. 287–307). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316817049.019

Stuart, E. A., & **Rubin, D. B.** (2008). Best practices in quasi-experimental designs. Best practices in quantitative methods. In J. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 155–176). New York: Sage. https://doi.org/10.4135/9781412995627.d14

Studenica, A., **Laketa, A.**, **Chrysochoou, E.**, **Blakey, E.**, & **Vivas, A. B.** (2022). The influence of bilingualism on adolescent cognition: The roles of biculturalism, the bilingual profile, and linguistic similarity. *Cognitive Development*, **63**, 101203. http://doi.org/10.1016/j.cogdev.2022.101203

Tamburelli, M. (2021). Taking taxonomy seriously in linguistics: Intelligibility as a criterion of demarcation between languages and dialects. *Lingua*, **256**, 103068. http://doi.org/10.1016/j.lingua.2021.103068

Tang, C., & **Van Heuven, V. J.** (2015). Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures. *Linguistics*, **53**, 285–312. http://doi.org/10.1515/ling-2015-0005

Tourva, A., & **Spanoudis, G.** (2020). Speed of processing, control of processing, working memory and crystallized and fluid intelligence: Evidence for a developmental cascade. *Intelligence*, **83**, 101503. http://doi.org/10.1016/j.intell.2020.101503

Tsiplakou, S. (2009). Code-switching and code-mixing between related varieties: establishing the blueprint. *The International Journal of Humanities*, **6**, 49–66. http://doi.org/10.18848/1447-9508/CGP/v06i12/42576

Tsiplakou, S. (2014). How mixed is a 'mixed' system? The case of the Cypriot Greek koiné. *Linguistic Variation*, **14**, 161–178. https://doi.org/10.1075/lv.14.1.07tsi

Tsiplakou, S., **Kambanaros, M.**, & **Grohmann, K. K.** (2022). Literacy Development in Cyprus: Exploring the Effects of Diglossia and Bilectalism. In *Handbook of Literacy in Diglossia and in Dialectal Contexts: Psycholinguistic, Neurolinguistic, and Educational Perspectives* (pp. 51–75). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-030-80072-7_4

Unsworth, N., **Redick, T. S.**, **Spillers, G. J.**, & **Brewer, G. A.** (2012). Variation in working memory capacity and cognitive control: Goal maintenance and microadjustments of control. *Quarterly Journal of Experimental Psychology*, **65**, 326–355. https://doi.org/10.1080/17470218.2011.597865

Veenstra, A., **Antoniou, K.**, **Katsos, N.**, & **Kissine, M.** (2018). Resisting attraction: Individual differences in executive control are associated with subject–verb agreement errors in production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **44**, 1242. http://doi.org/10.1037/xlm0000516

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation. https://doi.org/10.1037/t15170-000

Wichmann, S. (2020). How to distinguish languages and dialects. *Computational Linguistics*, **45**, 823–831. http://doi.org/10.1162/coli_a_00366

Yurtsever, A., **Anderson, J. A.**, & **Grundy, J. G.** (2023). Bilingual children outperform monolingual children on executive function tasks far more often than chance: An updated quantitative analysis. *Developmental Review*, **69**, 101084. http://doi.org/10.1016/j.dr.2023.101084

Zhou, B., & **Krott, A.** (2016). Data trimming procedure can eliminate bilingual cognitive advantage. *Psychonomic Bulletin & Review*, **23**, 1221–1230. https://doi.org/10.3758/s13423-015-0981-6

https://doi.org/10.1017/S1366728924000506 Published online by Cambridge University Press