

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Is Bilingualism Associated with Enhanced Executive Functioning in Adults?
A Meta-Analytic Review

Minna Lehtonen^{1,2}, Anna Soveri^{1,3}, Aini Laine¹, Janica Järvenpää¹, Angela de Bruin⁴ & Jan
Antfolk^{1,3}

¹Department of Psychology, Abo Akademi University, Turku, Finland

²Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki,
Finland

³Turku Brain and Mind Centre, University of Turku, Turku, Finland

⁴Basque Center on Cognition, Brain and Language, Donostia, Spain

Author note

The study was financially supported by Academy of Finland (grant # 288880), Emil Aaltonen Foundation project grant, and University of Helsinki 3-year grants to the first author. The authors wish to thank Benny Salo for statistical consultation, as well as Matti Laine, Jussi Jylkkä, and the rest of the BrainTrain research group for valuable discussions. We are also indebted to all the authors who kindly responded to our queries and who provided additional data to the meta-analysis (for a full list of contributing authors, see Table S2).

The results from this meta-analysis have been presented at the Learning and Plasticity scientific meeting in Äkäslompolo, Finland, in 2017.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Correspondence concerning this article should be addressed to Minna Lehtonen,
2 Department of Psychology, Abo Akademi University, FI-20500 Turku, Finland. Email:
3 minlehto@abo.fi

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

BILINGUALISM AND EXECUTIVE FUNCTIONS

Abstract

1
2 Due to enduring experience of managing two languages, bilinguals have been argued to
3 develop superior executive functioning compared to monolinguals. Despite extensive
4 investigation, there is, however, no consensus regarding the existence of such a bilingual
5 advantage. Here we synthesized comparisons of bilinguals' and monolinguals' performance
6 in six executive domains using 891 effect sizes from 152 studies on adults. We included
7 unpublished data, and considered the potential influence of a number of study-, task-, and
8 participant-related variables. Before correcting estimates for observed publication bias, our
9 analyses revealed a very small bilingual advantage for inhibition, shifting, and working
10 memory, but not for monitoring or attention. No evidence for a bilingual advantage remained
11 after correcting for bias. For verbal fluency, our analyses indicated a small bilingual
12 disadvantage, possibly reflecting less exposure for each individual language when using two
13 languages in a balanced manner. Moreover, moderator analyses did not support theoretical
14 presuppositions concerning the bilingual advantage. We conclude that the available evidence
15 does not provide systematic support for the widely held notion that bilingualism is associated
16 with benefits in cognitive control functions in adults.

17
18 *Keywords:* Bilingualism, executive functions, cognitive control, bilingual advantage, meta-
19 analysis

20
21
22
23
24
25

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 **Significance of the meta-analysis**

2 The idea that bilinguals outperform monolinguals in cognitive control functions seems to have
3 already been accepted by the popular media and educators, due to a number of influential studies
4 reporting a bilingual advantage. Our thorough meta-analysis, however, suggests that healthy
5 bilingual adults do not have such a cognitive control advantage. The synthesis of 152 studies and
6 891 comparisons and several moderator variables do not show systematic advantages across the
7 analyzed cognitive domains, tasks, or bilingual populations.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Bilingualism – acquisition, mastery, and use of two languages – has been associated with superior
2 executive functioning in studies comparing bilinguals and monolinguals (e.g., Bialystok, Craik &
3 Luk, 2012; Bialystok, 2017). Executive functions (EF) is an umbrella term for high-level cognitive
4 control functions that are involved in all complex mental activities and, therefore, are of particular
5 importance to human behavior. Despite a high number of studies addressing the organization of EF,
6 there is still lack of clarity regarding the definition and the subcomponents of EF. The most
7 frequently postulated EF components are, however, working memory, inhibition, and set shifting
8 (for reviews, see e.g., Jurado & Rosselli, 2007; Miyake & Friedman, 2012; Niendam, Laird, Ray,
9 Dean, Glahn & Carter, 2012). The early evidence of better EF performance in bilingual individuals
10 has naturally raised widespread interest among researchers as well as educators and media. Even
11 though the number of studies reporting positive cognitive effects of bilingualism has been high,
12 there have also been several reports of null findings as well as critical claims arguing that
13 convincing evidence for a bilingual advantage is lacking (e.g., Paap & Greenberg, 2013). In fact,
14 despite the wide interest and intense investigation, the field has not reached consensus on the nature
15 and extent of the putative bilingual advantage. The aim of the present meta-analysis is to investigate
16 the suggested bilingual EF advantage in adult samples.

17 Theoretically, the bilingual advantage is assumed to stem from the demands that the
18 use of two languages puts on the cognitive control system. Bilinguals' both languages have
19 been shown to be active even when only one of them is used for communication (e.g., Marian
20 & Spivey, 2003; Wu & Thierry, 2010). Producing a word in one language also activates the
21 word in the other language, eliciting competition between the lexical alternatives. This means
22 that cognitive control functions must work effectively to enable fluent use of the appropriate
23 language and to prevent interference from the other language. Efficient use of two languages
24 is assumed to require inhibition of items of the irrelevant language (Green, 1998) and flexible
25 switching between languages. Further, it requires monitoring the activation levels of the two

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 languages and of the language context in order to choose the appropriate target language.
2 This control of language use is assumed to involve domain-general EF processes, that is,
3 control functions that are also used in other cognitive domains than language, such as
4 monitoring behavior for conflict and inhibition of unwanted mental representations to
5 minimize the conflict. Similarly, language switching and domain-general task switching
6 share many common features, such as switch costs and their asymmetries (Prior & Gollan,
7 2011), as well as partly common neural substrates (De Baene, Duyck, Brass & Carreiras,
8 2015). Frequent language switching has therefore been suggested to train domain-general EF
9 (e.g., Prior & MacWhinney, 2010; Soveri, Rodríguez-Fornells & Laine, 2011a; but see
10 Jylkkä et al., 2017). Furthermore, distinct language use patterns and interactional contexts
11 have been proposed to set differential demands on cognitive control and thus possibly lead to
12 differential “training” gains (Green & Abutalebi, 2013). Due to often lifelong experience of
13 cognitive control in the field of language, bilinguals are believed to have received more
14 practice in domain-general EF processes than monolinguals (for narrative reviews, see, e.g.,
15 Bialystok, 2017; Bialystok et al., 2012).

16 Differences between bilinguals and monolinguals have also been reported in neural
17 measures (for reviews, see, e.g., Abutalebi, 2008; Abutalebi & Green, 2016; Bialystok, 2017;
18 García-Pentón, Fernández García, Costello, Duñabeitia & Carreiras, 2015; Li, Legault &
19 Litcofsky, 2014), most of these in adult samples. These effects have been assumed to reflect
20 similar mechanisms to other kinds of experience-dependent neuroplasticity observed as a
21 result of sustained enriching experiences, such as practicing music (e.g., Elbert, Pantev,
22 Wienbruch, Rockstroh & Taub, 1995) or having extensive experience in spatial navigation
23 (Maguire et al., 2000).

24 **Possible Moderators of the Bilingual Advantage**

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 The first observations of a bilingual EF advantage were reported in the domain of
2 inhibition and interference control (e.g., Bialystok, 2001; Bialystok et al., 2004). Later,
3 advantages have also been reported in the domains of shifting (e.g., Garbin et al., 2010; Prior
4 & MacWhinney, 2010), general conflict monitoring (e.g., Hilchey & Klein, 2011), WM (e.g.,
5 Bialystok, Craik & Luk, 2008a; Luo, Craik, Moreno & Bialystok, 2013), and attentional
6 processes (e.g., Bialystok, 2017; Soveri, Laine, Hämäläinen & Hugdahl, 2011b). While the
7 early findings for domain-general or nonverbal EF tasks have typically been positive,
8 bilingual participants have also been reported to show a disadvantage, that is, inferior
9 performance compared to monolinguals, in verbal tasks (e.g., Bialystok, 2009). These
10 disadvantages have been proposed to be due to less exposure to each individual language
11 compared to monolinguals (e.g. Gollan, Montoya, Cera & Sandoval, 2008) or due to lexical
12 interference from the other competing language (e.g. Kroll & Gollan, 2014).

13 Bilingual advantages in EF have been reported in all age groups, including children,
14 young and middle-aged adults, as well as the elderly. Advantages have, however, been stated
15 to be most consistently observed in older adults who are not at the peak of their cognitive
16 functioning (e.g., Bak, Nissan Allerhand & Deary, 2014; Bialystok et al., 2008a; Luo, Luk &
17 Bialystok, 2010). This could be the case if the normal, age-related decline of EF processes is
18 attenuated in bilingual individuals, as proposed by for example Bialystok et al. (2008a). It has
19 also been suggested that the bilingual advantage decreases with practice during the course of
20 an experiment, reducing differences between groups over time, and these kinds of practice
21 effects are slower in older participants (Hilchey & Klein, 2011).

22 Assuming that the bilingual advantage is due to the training of continuous management
23 of different languages, longer and more intensive bilingual experience should be associated
24 with larger gains. In line with this view, bilingual participants with an early age of acquisition
25 (AoA) of the second language (L2) have been proposed to show larger advantages (e.g., Luk,

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 de Sa & Bialystok, 2011). Similarly, proficiency in the languages has been suggested to
2 modulate training gains: A strong L2 should elicit higher interference to the first language
3 (L1) than a weak L2, and thus lead to higher cognitive control demands. Furthermore, the
4 structural and lexical similarity of bilinguals' two languages has been suggested to influence
5 how much they interfere with each other, with more similar languages assumedly creating
6 more interference and therefore larger training gains. Language similarity could thus affect
7 the intensity of training of cognitive control (Bialystok, 2017; Costa, Santesteban & Ivanova,
8 2006).

9 **The Controversy**

10 As mentioned above, the excitement following early findings supporting the bilingual
11 advantage has recently turned into a strong controversy, with publication of mixed findings
12 and studies with large samples showing null results (e.g., Duñabeitia et al., 2014; Paap,
13 Johnson & Sawi, 2014). Criticism has been raised, in particular, towards the natural groups
14 design, which may allow intervening variables other than language history to affect the
15 observed performance differences between groups. Such variables have been suggested to
16 include socio-economic status (SES; e.g., Morton & Harper, 2007), intelligence, culture
17 (Bialystok & Viswanathan, 2009; Yang, Yang & Lust, 2011), or immigration status (Kousaie
18 & Phillips, 2012; de Bruin, Bak, Della Sala, 2015a), and attempts at controlling for these
19 variables have been made in several studies. Other concerns include the frequent use of small
20 sample sizes decreasing statistical power, as well as questions related to tasks used to
21 measure different aspects of EF (e.g., Paap, Johnson & Sawi, 2015).

22 A wide spectrum of tasks are available to assess EF (see, e.g., Valian, 2015, for a
23 review), leading to variability in task selection in original studies. This variability may be
24 problematic due to issues with task validity and reliability (see Barkley, 2012 for a discussion
25 about issues with validity and reliability of EF tasks). In fact, a widely cited study on EF by

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Miyake and colleagues (2000) shows that the correlations between EF tasks are low (r range
2 = .00 – .41). These correlations indicate that EF tasks share less than 17% of the variance in
3 test performance while at least 83% is accounted for by other factors. Some of this error
4 variance may be related to the fact that tasks assumedly measuring the same EF domain
5 inevitably, to some degree, also engage other cognitive processes (a phenomenon called the
6 “task impurity problem”), whereas some error variance may be related to issues with
7 reliability (e.g., Paap & Sawi, 2016; Soveri et al., 2016). Weak correlations between
8 inhibition tasks have been reported also in a very recent study investigating inhibition as a
9 concept (r range = .00 – .44; Rey-Mermet, Gade, & Oberauer, 2017). Previous studies
10 investigating the reliability of EF tasks report considerable variability in reliability estimates,
11 ranging from low to high (e.g., Paap & Sawi, 2016; Rey-Mermet, et al., 2017; Soveri et al.,
12 2016). The mixed results from previous studies on the bilingual advantage may thus partly be
13 related to problems with task validity and reliability, as group differences will be difficult to
14 detect if the amount of error variance is high.

15 Furthermore, it has been claimed that the field suffers from publication bias. De Bruin
16 and colleagues (de Bruin, Treccani & Della Sala, 2015b) searched for conference abstracts on
17 bilingualism and executive control and looked into which studies were subsequently
18 published. They showed that studies supporting the bilingual advantage hypothesis were most
19 likely to be published, whereas the ones challenging it were less likely to be published. A
20 very recent bibliometric analysis by Sanchez-Azanza, López-Penadés, Buil-Legaz, Aguilar-
21 Mediavilla, and Adrover-Roig (2017), in turn, suggests that publication trends on the
22 bilingual advantage may be changing. Their analysis revealed that from 2014 onwards,
23 published studies challenging the bilingual advantage increased notably, possibly after the
24 influential papers by Hilchey and Klein (2011) and by Paap and Greenberg (2013) that were
25 more critical towards the bilingual advantage hypothesis. Sanchez-Azanza and colleagues

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 (2017) did not find differences in the impact factors of the journals or the accumulating
2 number of citations depending on the kind of effects reported. However, they found that
3 studies from the year 2014 that challenged the advantage had gathered more citations by June
4 2016 than those from the same year supporting the advantage.

5 **Previous Systematic Reviews**

6 Previous meta-analyses and systematic reviews on the relationship between
7 bilingualism and particular aspects of EF have reported somewhat varying results (Adesope
8 et al., 2010; de Bruin et al., 2015b; Donnelly, 2016; Grundy & Timmer, 2016; Hilchey &
9 Klein, 2011; Paap, et al., 2015; Zhou & Krott, 2016). The first meta-analysis in the field was
10 conducted by Adesope and colleagues (2010) based on 63 studies reported in 39 articles
11 investigating the effects of bilingualism in children and adults. Their results showed that
12 bilingual participants outperformed monolinguals on tasks measuring attentional control,
13 problem-solving, symbolic representation and abstract reasoning skills, metalinguistic
14 awareness, metacognitive skills, and WM, with effect sizes ranging from small to large ($g =$
15 $.26$ to $.96$). Adesope et al. (2010) found no clear evidence of publication bias by using a
16 classic fail-safe N and Orwin's fail-safe test.

17 In a systematic review including 13 articles, Hilchey and Klein (2011) investigated
18 the effect of bilingualism on nonverbal inhibitory control tasks in children and adults. They
19 found a bilingual advantage in the interference effect only in middle-aged or elderly adults,
20 not in young adults or children. They also found a clear bilingual advantage in all age groups
21 for global RTs (i.e., a measure including both incongruent trials with conflict present and
22 congruent trials without conflict) that assumedly reflect conflict monitoring processes.

23 Using a vote-count method, Paap et al. (2015) summarized the results of all studies
24 published after the review by Hilchey and Klein (2011) investigating differences between
25 bilingual and monolingual participants in nonverbal inhibition and set shifting. Their

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 summary showed a bilingual advantage only in a small proportion of the included studies
2 (proportions ranging from .125 to .217). Furthermore, their review showed that a bilingual
3 advantage was typically reported in studies with small samples, while null results were only
4 reported in studies with larger samples ($n > 50$). Based on this information, Paap et al. (2015)
5 concluded that it is unlikely that a bilingual advantage in EF exists. Similarly, in an updated
6 review, Hilchey, Saint-Aubin, and Klein (2015) concluded that, contrary to their 2011 review
7 (Hilchey & Klein, 2011) showing a bilingual advantage on global RTs, there is little support
8 for this claim in more recent publications.

9 De Bruin and colleagues (2015b) performed a meta-analysis on the published data (41
10 studies) of tasks from various EF domains included in their publication bias analysis. The
11 results showed a small but significant positive effect (Cohen's $d = .3$) of bilingualism on EF
12 – an outcome that likely overestimated the bilingual advantage, given the presence of a
13 publication bias in the selection of reports. The studies with different result types (i.e.,
14 supporting or challenging the advantage) did not differ significantly in sample size, type of
15 tasks used, power to detect an effect, or the year of the conference abstract. The only
16 difference they found was the number of tasks reported, which was typically lower for studies
17 with positive results. In a further analysis of the same data (de Bruin, Treccani & Della Sala,
18 2015c), the results remained the same when excluding verbal tasks that could be
19 hypothesized to show smaller effects (Bialystok, Kroll, Green, MacWhinney & Craik, 2015).

20 Donnelly (2016) investigated the effects of bilingualism on interference control and
21 set shifting in healthy children and adults. The meta-analysis on interference control included
22 168 effect sizes from 43 studies and showed a small overall effect of bilingualism ($d = .29$).
23 The effect was significantly moderated by which research group had conducted the original
24 studies. There was, however, no effect of task (e.g., Flanker task and Simon task) or type of
25 measure (i.e., global RTs vs. interference cost) and the significant moderator effects of AoA

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 and age were interpreted to be due to publication bias. The meta-analysis on set shifting was
2 based on 30 effect sizes from 10 studies. The results showed no effect of bilingualism on set
3 shifting ($d = -.03$) and there was no effect of research group.

4 Grundy and Timmer (2016) studied the bilingual advantage in WM in children as well
5 as young and older adults and found a small to moderate positive effect size (Pearson's $r =$
6 $.20$) for the difference in WM performance between bilinguals and monolinguals. This meta-
7 analysis included 88 effect sizes from 27 studies, and based on fail-safe N , the authors
8 concluded that their population estimate is likely safe from publication bias. They also
9 reported that the largest advantage was observed in children and that the effect sizes were
10 moderated by the language in which the verbal tasks were performed, that is, the L1 or L2 of
11 the bilingual participants. The advantages were smaller when the bilinguals had performed
12 the WM tasks in their L2.¹

13 To summarize, despite extensive efforts and previous systematic reviews, the
14 evidence regarding the bilingual advantage is inconclusive and controversial. Adesope and
15 colleagues (2010) reported positive effects of bilingualism in several cognitive domains.
16 Hilchey and Klein (2011) also found a robust bilingual advantage in conflict monitoring and
17 some evidence for an advantage in inhibitory control, but later found little support for this
18 (Hilchey et al., 2015). Paap et al. (2015) and Donnelly (2016) reported small or no bilingual
19 advantages in inhibitory control or set shifting, and de Bruin et al. (2015b) a small effect in a
20 set of various EF tasks. Donnelly (2016) and de Bruin et al. (2015) also, however, observed a
21 publication bias, calling these effects into question. Some evidence for an advantage in WM
22 was observed in the analyses of Grundy and Timmer (2006) and Adesope et al. (2010), which

¹ In one meta-analysis, it has furthermore been suggested that aspects of data analysis such as data trimming can affect the outcome. Untrimmed studies with longer RTs were found to be more likely to report a bilingual advantage (Zhou & Krott, 2016).

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 included both children and adults. None of the systematic reviews on the bilingual advantage
2 which observed a publication bias attempted to correct for it in the analyses.

3 The inconsistencies in the previous systematic reviews are probably mainly related to
4 differences in inclusion criteria, domains studied, and statistical methods employed. The
5 inconsistencies in previous original studies in the field, on the other hand, may be due to the
6 limits of relatively small experimental groups, varying methods, and unclear theory behind
7 the EF tasks and the functions they measure. To be able to conclude that bilinguals show an
8 executive advantage over monolinguals, studies should demonstrate that there is a component
9 or components of EF in which bilinguals are consistently showing an advantage compared to
10 monolinguals. A bilingual advantage seen in only one task does not necessarily mean that
11 there is an advantage in the cognitive domain the task in question is assumed to measure.
12 This is because correlations between tasks that are assumed to measure the same domain
13 have turned out to be surprisingly low e.g., Miyake et al., 2000; Paap & Greenberg, 2013;
14 Paap & Sawi, 2014; Waris et al., 2017). Also, for many EF tasks, validity information is
15 completely lacking.

16 The specific characteristics of the participant groups studied deserve particular
17 attention. As pointed out, for example, by Bialystok (2001, as cited in Klein, 2016), and
18 Hilchey and Klein (2011), cognitive development throughout the lifespan is a complex and
19 multidimensional process with several hidden factors influencing information processing
20 abilities. Additionally, the previous original studies have been conducted on very different
21 bilingual and monolingual populations in different countries and regions with unique socio-
22 cultural characteristics, which likely contributes to the mixed results. These issues highlight
23 the complexity of the research question and the need for increasingly extensive, yet
24 sufficiently detailed systematic investigations.

25 **The Present Study**

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 In this meta-analysis, we review the currently available extensive literature of
2 bilingualism and EF in adults. Compared to previous systematic reviews, our meta-analysis is
3 considerably more wide-ranging in the number of included studies and the domains, tasks,
4 and background variables investigated. As an attempt to reduce the effect of publication bias
5 (de Bruin et al., 2015b), we also include unpublished studies, primarily doctoral dissertations
6 and Master's theses, along with peer-reviewed journal articles. Most previous meta-analyses
7 on this topic have also not explicitly taken into account the fact that effect sizes extracted
8 from the same studies and participant samples are not independent of each other. Here,
9 however, we employ a multi-level meta-analytic approach that allows us to include all
10 observations of interest from the original studies without violating assumptions of
11 independence. The dependence between observations is empirically estimated, and estimates
12 and confidence intervals are appropriately adjusted for this dependency.

13 While previous meta-analyses have primarily studied one or two domains of EF, we
14 include a whole spectrum of EF domains: inhibitory control, monitoring, shifting, WM,
15 attention, and verbal fluency. Furthermore, due to reported low convergent validity of EF
16 tasks, it is not self-evident that similar effects are observed in different tasks even if they
17 assumedly measure the same function. Therefore, we pay particular attention to the specific
18 task paradigms used in the original studies.

19 Furthermore, we analyze whether the stimulus material used in the tasks is verbal or
20 nonverbal in nature (see also de Bruin et al., 2015c; Grundy & Timmer, 2016). As bilingual
21 participants have been reported to be at a disadvantage in verbal tasks (e.g., Bialystok et al.,
22 2008b, Bialystok, 2009; Bialystok, Barac, Blaye & Poulin-Dubois, 2010; Bialystok & Luk,
23 2012), larger bilingual advantages may be observed for nonverbal than verbal tasks. For
24 verbal fluency, which includes a strong language component, smaller bilingual advantages
25 could be observed than for other EF tasks. This would be the case especially for category

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 fluency, in which the demands on EF may be lower than for letter fluency. Letter fluency has
2 been suggested to be more effortful because phonemic generation is not a task one commonly
3 performs and it does not reflect the organization of words in the mental lexicon. With
4 category fluency, participants can use existing links and practiced strategies to activate
5 relevant representations (see, e.g., Luo et al., 2010). We also consider the language in which
6 the EF tasks are performed, that is, whether the testing language is bilinguals participants' L1
7 or L2, in order to ensure that the group comparisons in verbal EF tasks are fair (Grundy &
8 Timmer, 2016).

9 The problems with matching participant groups have been widely acknowledged in
10 the field, but no previous meta-analyses has explicitly studied its effect. We thus examined
11 the extent to which the participant groups have been matched, for example, for SES or age.
12 Several studies have also matched the groups for IQ. Because IQ has been shown to correlate
13 highly with WM in healthy young adults (e.g., Friedman, Miyake, Corley, Young, DeFries,
14 Hewitt, 2006; Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süss,
15 2005), matching participants according to IQ may be problematic as it might lead to groups
16 that are matched according to WM ability as well, and thus conceal a possible bilingual
17 advantage. In contrast, matching for vocabulary size could artificially augment group
18 differences (Bialystok, Craik & Luk, 2008b): Assuming that bilinguals typically suffer from a
19 disadvantage in verbal tasks, matching for vocabulary might lead to including unusually well-
20 performing individuals in the bilingual group. We therefore analyze whether such matching
21 practices have had an influence on the reported effects.

22 We also consider several participant-related variables: age group, AoA of L2,
23 language proficiency, and immigrant status. First, we study whether older adults show a
24 larger bilingual advantage than younger adults do (e.g., Hilchey & Klein, 2011; Bialystok,
25 2017). Second, we test the hypothesis that bilingual participants with an early AoA of L2

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 show a larger advantage than late bilinguals, due to the assumedly longer amount of training
2 received (e.g., Luk, De Sa & Bialystok, 2011). Third, we analyze the effect of proficiency
3 level in L2, with the assumption that high-proficiency bilingual participants have faced
4 stronger demands and more training for cognitive control than lower-proficiency bilinguals.
5 Fourth, we test whether possible immigrant status of bilinguals, a variable often discussed but
6 not systematically analyzed in previous reviews, moderates the effects. Our focus is on
7 adults: There was a vast amount of studies available even with the present focus. Also, while
8 bilingual advantages have been reported in children as well (e.g., Grundy & Timmer, 2016;
9 Hilchey et al., 2015; but see Antón et al., 2014; Duñabeitia et al. 2014), we would expect the
10 advantage to be better observed in adults due to an assumedly longer “training period” of EF,
11 at least in early bilinguals (who have decades of bilingual language control experience vs. a
12 few years in children). Moreover, the significance of the phenomenon would naturally be
13 limited if the positive effects were only observed in children and not in adults².

14 In addition, we study whether an EF advantage is better observed in bilingual
15 participants with particular language pairs, for example those that have a great deal of
16 structural and lexical overlap (e.g., Spanish and Catalan). Lastly, the country in which the
17 study is conducted may moderate the effects as it is related to not only the cultural and
18 sociolinguistic environment of the participants but also to that of the researchers. For
19 example, it has been suggested that the general societal atmosphere regarding bilingualism in
20 different countries (e.g., Canada vs. USA) may be associated with a tendency of reporting
21 either positive or negative findings (Bak & Alladi, 2016; Bak, 2016). While we acknowledge
22 that it is difficult to isolate the exact contributing effects of language similarity and country
23 from, for example, the intertwined cultural factors or the typical language use patterns in the

² It should be noted that the present study only uses cross-sectional data, and longitudinal studies following the same bilinguals and monolinguals from childhood to adulthood would provide more conclusive evidence of the persistence of bilingual advantages possibly observed in children. Unfortunately, such studies are largely lacking (but see Bak et al., 2014).

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 bilingual communities (e.g., language switching, see, e.g., Green, 2011), these variables
2 should at least give us an idea of whether advantages are consistently observed in particular
3 bilingual populations or environments. Such findings could, in turn, give us directions for
4 further research with regard to what kinds of socio-cultural aspects may potentially be
5 associated with EF gains in bilingual individuals. These variables have not been studied in
6 the previous meta-analyses.

7 **Primary Research Questions of the Present Study**

- 8 1. In which EF domain (if any) do we observe a bilingual advantage? What are
9 estimates for the advantage in each cognitive domain when correcting for possible
10 publication bias?
- 11 2. Are possible advantages specific to particular task paradigms?
- 12 3. Are possible advantages of different magnitude in verbal than nonverbal tasks? In
13 verbal tasks, have the tasks been performed in bilinguals' L1?
- 14 4. Are observed advantages affected by how participant groups have been matched for
15 age, SES, vocabulary knowledge, or IQ?
- 16 5. Is there a larger advantage in older than younger bilingual adults?
- 17 6. Does AoA or proficiency in L2 moderate the advantages? Is the advantage related
18 to possible immigration background of the bilingual participants?
- 19 7. Does the country in which the study was conducted or language pair of the
20 bilinguals moderate the effects?

23 **Method**

24 **Literature Search**

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 We searched the electronic databases PsycINFO (ProQuest), PubMed, Google
2 Scholar, ProQuest Dissertations & Theses, Networked Digital Library of Thesis and
3 Dissertations, and WorldCat. Additionally, we used the data from the published studies
4 included in the meta-analysis by de Bruin et al. (2015b). The main search included the terms
5 *bilingual* and *monolingual* and terms referring to both EF in general (e.g., "executive
6 function", "executive control", "attentional control", "cognitive control") and the chosen
7 cognitive domains ("inhibition", "shifting", "monitoring", "WM", and "attention"). The
8 search strings were adjusted for each database depending on the size of the corpus, functional
9 differences of Boolean operators, and advanced search functions (for exact search strings,
10 see Table S1 in Supplementary Materials).

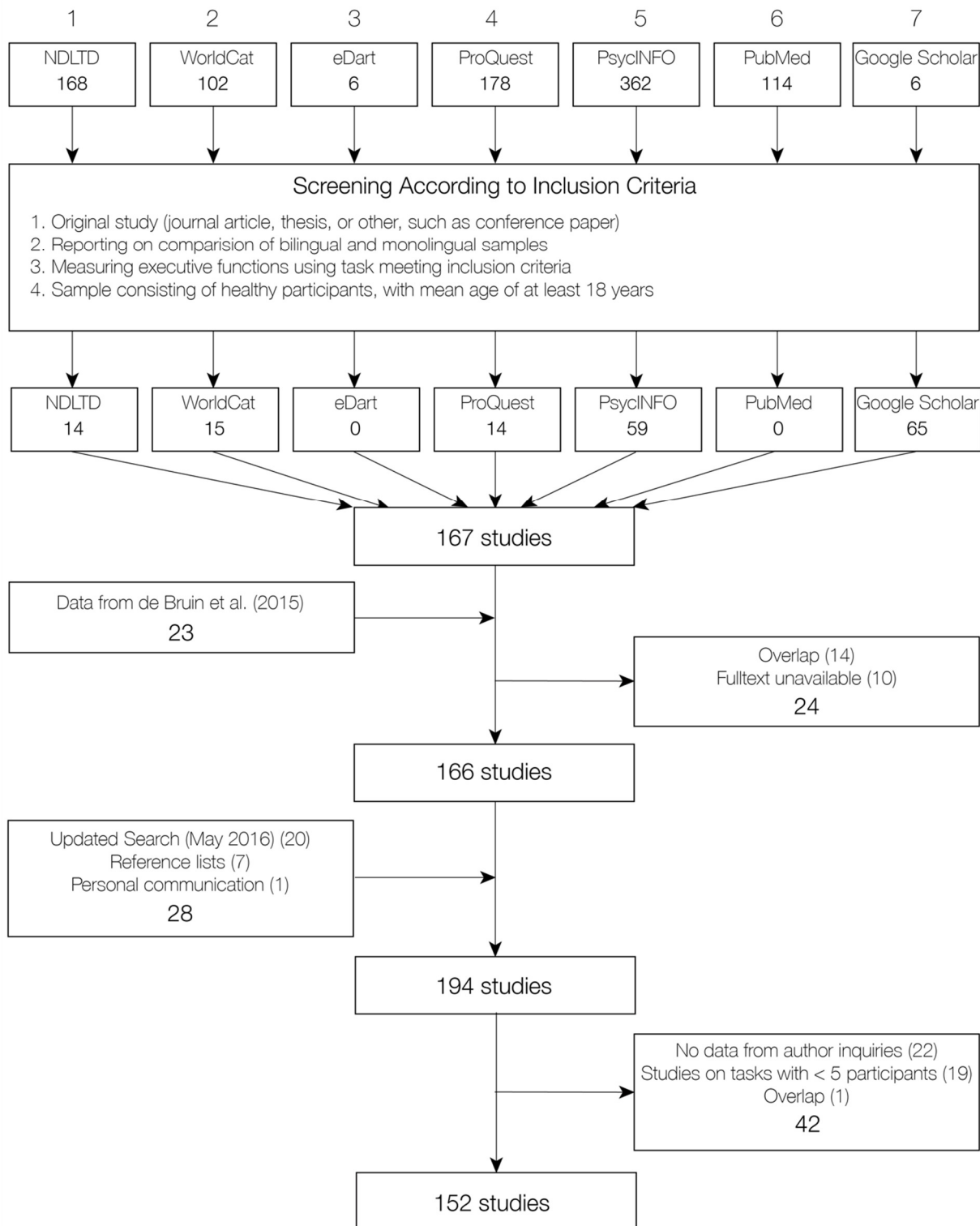
11 Prior to the search, we tested the sensitivity (i.e., the amount of the existing relevant
12 studies that would be found) of our search string in PsycINFO and Google Scholar. To do
13 this, we first randomly picked 10 studies matching our inclusion criteria from de Bruin et al.
14 (2015b) meta-analysis. Because all of these studies also appeared in our searches, the search
15 string was deemed sufficiently sensitive.

16 The first search was conducted in October–November 2015 and covered the years
17 from 1999 to present. We screened all search hits to identify studies potentially relevant for
18 the present meta-analysis and then screened abstracts and method sections for the inclusion
19 criteria. After this, we employed a snowballing procedure and reviewed the reference lists of
20 44 randomly selected studies (i.e., 1/4 of the potentially relevant studies) found in the first
21 search. We conducted a second search in June 2016. In case the study was relevant for our
22 meta-analysis but necessary data for calculating effect sizes were not reported, we contacted
23 authors via email to obtain more data. We also asked the authors for information regarding
24 the immigration status of the bilingual participants if it had not been provided in the study.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 We got a response to 60% of the emails sent out, and 36% of the responses led to acquisition
- 2 of data (see Table S2 for authors providing additional data).

BILINGUALISM AND EXECUTIVE FUNCTIONS



1

2 *Figure 1.* Flowchart of the screening process. Numbers 1-7 at the top of the figure refer to the

3 order in which the results were screened. Other values refer to unique inclusions at each

4 stage. Inclusions reported to the left and exclusions reported to the right. NDLTD =

5 Networked Digital Library of Thesis and Dissertations.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Inclusion and Exclusion Criteria

2 For a study to be included, it had to report a comparison of bilingual and monolingual
3 participants in at least one measure of EF. We included published journal articles,
4 unpublished doctoral dissertations, Master's and Bachelor's theses, as well as other types of
5 reports (e.g., posters). If we noted clear similarity between a thesis and a later published
6 article, we asked the corresponding authors to verify the overlap. In cases of verified or
7 deemed overlap, we only included the peer-reviewed article data. However, if not all tasks in
8 the thesis had been reported in the published article, we included data for these tasks. If no
9 relevant data were available even after author inquiries, the study was excluded. No studies
10 were excluded based on the language they were written in. The included studies were written
11 in English, French, or Portuguese. We only included behavioral data, excluding
12 neuroimaging data. Behavioral data from relevant tasks from brain imaging or
13 electrophysiological studies were, however, included.

14 In the following, we introduce the inclusion criteria related to participants, task
15 paradigms, and stimuli of the original studies.

16 **Inclusion related to participants.** We only included samples of healthy adult
17 participants (mean age at least 18 years). Clinical groups, such as those including participants
18 with deafness or neurological illnesses such as dementia, were excluded. We relied on the
19 original studies' grouping of participants to bilinguals and monolinguals even though there
20 was large variation in the operational definitions of bilingualism. However, we excluded
21 studies that explicitly reported having blended bilingual and monolingual participants in a
22 sample (e.g., Deslauries, 2008; Roth, 2003). Despite the variability in the definitions, the
23 majority of studies included monolinguals with limited experience of a second language, and
24 in all studies, monolingual participants had markedly less L2 experience than the bilinguals.
25 Because we implemented rather liberal initial inclusion criteria for the participant groups, we

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 also coded AoA and language proficiency in L2 and used this information to narrow down
2 the groups in further analyses (see Participant characteristics under section Data Coding).

3 **Inclusion related to task paradigms.** In order to follow a paradigm-based approach
4 in the analysis, we focused on tasks that were relatively common in the whole set of studies.
5 For a task paradigm to be included, it had to be used in at least five assumedly different
6 samples (i.e., monolingual vs. bilingual comparisons).

7 To ensure the included tasks were sufficiently homogeneous, we excluded modified
8 tasks clearly measuring also another functional dimension of EF. Examples of such
9 modifications include switching between Simon and Flanker -type tasks (Kovelman, 2006) or
10 adjoining a parallel task to a typical EF task (e.g., N-back conducted while driving; Chong de
11 la Cruz, 2015). With the aim of not mixing two separate task paradigms, we only included
12 measures presented in pure, separate blocks for paradigms that included elements from
13 different EF tasks (e.g., from Go-NoGo Flanker, we only included Flanker effects from
14 blocks presented separately from the Go-NoGo trials). Similarly, from the Antisaccade tasks,
15 we only included conditions with a pure antisaccade block. For the sake of homogeneity, we
16 also excluded data from studies in which major changes had been made to the typical setup.
17 Examples of such changes include using ANT Flanker blocks with an unusual proportion of
18 incongruent and congruent trials (Costa, Hernández, Costa-Faidella & Sebastián-Gallés,
19 2009), introducing a rule change in the middle of a Simon task (Samuel, 2015), or presenting
20 emotionally arousing pictures as distractors in a Flanker task (Pelham, 2014). In interference
21 control tasks, we did not include sentence tasks with conflict resolution due to the large
22 variation in how the conflict or interference control demands were operationalized, and none
23 of the individual tasks fulfilled the five samples minimum criterion.

24 **Inclusion related to stimuli.** We included switching tasks consisting of both
25 nonverbal and verbal stimuli (e.g., words, letters), but excluded language switching tasks, as

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 they are not relevant for monolinguals. We also excluded tasks where bilinguals' testing
2 language was switched within a block (e.g., Tabares, 2012). We excluded tasks involving
3 learning of new language items (e.g., Meuter & Ehrich, 2012). The nature or modality of the
4 stimulus material was not a criterion for exclusion: For example, we included Stroop tasks
5 with written words, auditory words and sounds, as well as visuospatial stimuli.

6 Screening all the potentially relevant studies for the abovementioned criteria as well
7 as responses for inquiries for missing data from authors resulted in a final dataset of 152
8 studies with 891 effect sizes. (See Figure 1 for a flowchart of the screening process).

9 **Data Coding**

10 In the following, we introduce the principles of coding of our study-, participant-, and
11 task-related variables and measures, as well as estimates for interrater reliability.

12 **Study characteristics.** We extracted the following study characteristics: list of
13 authors, year of publication or submission, country in which the study was conducted (or the
14 authors' country if other information was not explicitly available), and peer-review status of
15 study (peer-reviewed journal article or other study, i.e., thesis or poster).

16 **Participant characteristics.** We extracted a description of participants (e.g.,
17 university students) and the languages of the bilingual and monolingual samples. We then
18 extracted the participants' mean age and *SD* in each group and also coded age as a
19 dichotomous age group variable ("younger", mean 18–59 years; "older", mean 60 years or
20 older). This division was chosen as it reflected the distribution of the included studies well
21 and divided the studied samples naturally to the groups with as little overlap as possible. If
22 group-specific means for age were missing, we noted the combined age of mono- and
23 bilingual participants. Furthermore, we coded the most commonly occurring language pairs
24 of the bilinguals in the dataset (i.e., present in at least five bilingual samples).

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 We coded AoA of L2 as two variables according to the group mean AoA added with
2 one *SD*. For the first variable, we dichotomously grouped the bilingual participants according
3 to whether they had started learning their L2 before or after puberty (cut-off at 12 years of
4 age). We included a sample in the pre-pubertal category only if the mean AoA added with
5 one *SD* was below the cut-off age, or if all participants were reported to have acquired the L2
6 in kindergarten or in childhood. With the second variable, we similarly formed a group with
7 very early onset bilingualism (cut-off at 6 years). We further coded L2 proficiency as “high”
8 or “other” according to the description provided in the study³.

9 In addition, we coded the immigration status of the bilingual participants according to
10 whether more than half, less than half, or none of the bilingual participants in a sample were
11 first-generation immigrants (i.e., living in a country other than their country of birth).

12 **Matching of groups.** We coded whether the bilingual and monolingual groups were
13 matched for the following variables: age (mean and *SD*), income and education (as measures
14 of SES), IQ (e.g., WAIS score), and measures of vocabulary size (expressive, receptive or
15 both). Matching of groups for age was analyzed via calculating the effect size for the reported
16 difference between the groups; if this effect size was between $g = -0.3$ and $g = 0.3$, the groups
17 were considered matched⁴. For other variables, we relied on authors’ own statements
18 regarding the matching or checked the relevant statistics when reported. With regard to
19 education, a common situation was one where both groups consisted of university students.
20 In that case, the groups were considered matched.

21 **Task-related variables.** We coded the task paradigms and their measures and
22 grouped the measures into six domains: 1) inhibitory control, 2) monitoring, 3) set shifting, 4)

³ We chose this operationalization as the varying proficiency criteria used in the original studies make reliable comparisons between different proficiency levels (e.g., low, medium, high) impossible. Also, some original studies report bilingual samples with large within-group differences in proficiency. Such samples were coded as ‘other’.

⁴ Note that for some samples, this could not be calculated as relevant statistics were not reported. Such samples were treated as non-matched.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 WM, 5) attention, and 6) verbal fluency. For example, the interference effect from ANT
2 Flanker was categorized as a measure of inhibitory control and the orienting effect from the
3 same task as an attention measure. We utilized published factor analyses (e.g., Miyake et al.,
4 2000; Friedman & Miyake, 2004) to motivate our classification of task paradigms into
5 domains; however, as these are lacking for several tasks, we grouped such task paradigms
6 according to the functions they are typically considered to measure in the previous literature.
7 In addition to grouping measures into domains, we used the task paradigm variable in further
8 moderator analyses within each domain. For the full list of task paradigms, domain
9 groupings, and measures included, see Table 1.

10 Based on the nature of the stimulus material or produced output, a task was
11 dichotomously coded as “verbal” when including words, digits, or letters as stimuli (or
12 output), or as “nonverbal” when including other kinds of stimuli (such as pictures,
13 nonlinguistic sounds, or shapes). We also coded whether a verbal task was reported to be
14 performed in the bilinguals’ L1 or L2, as results of verbal tasks are likely to be influenced by
15 bilingual individuals’ skills in that language.

16 **Measures.** In order to calculate effect sizes, we extracted group sizes, means and *SDs*.
17 For most task paradigms, we preferred reaction times (RTs) over accuracy if available (see
18 Table 1). We excluded data where *SD* was reported as zero and thus not permitting effect size
19 calculation. If data from different blocks were reported separately, we used the first block, as
20 the differences between groups and demands for controlled processing have been shown to
21 diminish with practice (e.g., Bialystok et al., 2004).

22 The decision on which measures to include followed a priority order that was
23 primarily based on the measures most typically used for these tasks in bilingual EF studies. In
24 general, we prioritized measures that controlled for baseline performance such as cost effects
25 in inhibitory control tasks or set shifting tasks. In these measures, we preferred a neutral

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 baseline (if available) rather than a congruent one which can show facilitatory effects (e.g.,
2 Coderre, van Heuven & Conklin, 2013). The decision of inclusion was a tradeoff between
3 reducing noise in data (using as “clean” and homogenous measures as possible) while
4 including as much data as possible. In case the preferred domain-specific task measure (e.g.,
5 the Simon effect) was not available, we included the most difficult task condition (e.g., the
6 incongruent condition). If the most difficult condition was also not available, we excluded the
7 measure in question. In the following, we will outline the chosen measures, grouped by
8 cognitive domain (see also Table 1).

9 *Inhibitory control.* Inhibitory control refers to the ability to deliberately inhibit
10 dominant responses (Miyake et al., 2000) or competing representations (Stahl et al., 2014).
11 Task paradigms with inhibitory control measures included Simon⁵ (Simon & Rudell, 1967),
12 (ANT) Flanker (Eriksen & Eriksen, 1974; Fan, McCandliss, Sommer, Raz, & Posner, 2002),
13 Stroop (Stroop, 1935), Go-NoGo, and the Antisaccade task (Hallett, 1978). For interference
14 control tasks (Flanker, Simon, Stroop), we used 1) the interference effect (incongruent vs.
15 neutral trials; see, e.g., Bialystok, Craik & Luk, 2008a; Coderre et al., 2013); When that was
16 not available, the following measures were used in the following order of preference 2)
17 conflict effect (incongruent vs. congruent trials); or 3) incongruent condition. However, for
18 Go-NoGo tasks, we only included accuracy measures (i.e., failure to inhibit responses), as
19 Go-trial RTs may also measure monitoring (Donkers & Van Boxtel, 2004). For antisaccade

⁵ The Simon Arrows task (also sometimes called Stroop Arrows or Spatial Stroop; Hilchey & Klein, 2011; Blumenfeld & Marian, 2014) was categorized as a Stroop task (version “Spatial Stroop”). This was done as the conflict in the incongruent condition arises from two conflicting *stimulus* dimensions in both Stroop color-word (color and meaning of the word) and Simon Arrows tasks (direction and location of the arrow) instead of conflicting stimulus-response locations (as in standard Simon). The bilingual advantage has been suggested to be larger in the former case, i.e., in tasks requiring Stimulus-Stimulus inhibition (see Blumenfeld & Marian, 2014).

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 tasks, we included the antisaccade interference effect, antisaccade trial RTs, or antisaccade
2 error rates in that preference order.

3 *Set shifting.* Set shifting refers to the ability to switch between tasks or mental sets
4 (Miyake et al., 2000). For the set shifting domain, we grouped together the tasks from the
5 typical alternating runs paradigm (e.g., Rogers & Monsell, 1995). These included, for
6 example, the color-shape and number-letter tasks, here coined as “Task Switching”. For a
7 task to be included in this category, it had to have a mixed stimulus block with alternating
8 switch and non-switch (i.e., repetition) trials. As measures of set shifting in this paradigm, we
9 included, in preference order: 1) switching costs (switch minus repetition trials in a mixed
10 block); or 2) switching trials.

11 Shifting tasks also included the Trail Making Test (TMT; Reitan, 1958), Wisconsin
12 Card Sorting Test (WCST; Grant & Berg, 1948), and the switching measure of the Test of
13 Everyday Attention (TEA; Robertson, Ward, Ridgeway, Nimmo-Smith & McAnespie, 1994).
14 In case of TMT, we preferred measures controlling for baseline processing speed (e.g., Trail
15 B minus Trail A). If not available, Trail B (or another switching measure of a TMT version,
16 such as Trails 5; Duncan, Segalowitz & Phillips, 2016) was used. For WCST, we preferred
17 “perseverative errors”, as it is shown in Miyake et al.’s analysis (2000) to load on shifting
18 factor. If not available, we used “number of completed categories” (Miyake et al., 2000).
19 Lastly, we used “elevator counting with reversal” (a subtest of auditory switching) as the
20 shifting measure of TEA.

21 *Monitoring.* Monitoring refers to the ability to monitor conflict in information processing
22 and evaluate the need for cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001).
23 Task paradigms with monitoring measures included Flanker, Go-NoGo, Simon, Stroop⁶, and Task

⁶ In Stroop, the blocks of different conditions are often administered separately (especially in the paper versions).

However, high monitoring demands can be assumed to be present only in the incongruent block and thus be equivalent

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Switching. For the inhibitory control tasks, we preferred global RTs, as they have been more
2 commonly associated with general monitoring demands, but if not available, we included either the
3 neutral or congruent condition from blocks where they were presented together with incongruent
4 trials. In Task Switching, we preferred mixing costs (i.e., repetition trials from mixed blocks minus
5 trials from single-task blocks), and if not available, we used global RTs or repetition trials from the
6 mixed block.

7 *Working memory.* WM refers to a capacity-limited, multicomponent system
8 responsible for maintaining and manipulating information in the face of ongoing processing
9 (Baddeley, 2000). WM task paradigms consisted of N-back and span tasks. We grouped the
10 WM spans to the following standard categories: a) *simple spans* (e.g., forward or backward
11 digit span, Corsi block); b) *transformational WM tasks* that require re-ordering of items (here
12 called “LNS” tasks after the Letter-Number-Sequencing task (e.g., Wechsler Adult
13 Intelligence Scale (WAIS) 3rd ed.), also including Number Sequencing Span, Alpha Span,
14 and Matrix Span; Feng, 2008); c) *complex spans* in which another task is added to retrieving
15 or re-ordering items (e.g., Reading Span, Operational Span). In the N-back tasks, we did not
16 separate between lure and non-lure trials, as not all studies explicitly reported the difference
17 or had controlled for this. In the coding, we collapsed dual n-back tasks with standard simple
18 n-back tasks.

19 *Attention.* In the present meta-analysis, the Attention domain refers to the ability to
20 selectively direct and maintain attention to stimuli. Task paradigms with attention measures
21 included Sustained Attention to Response Task (SART; Robertson, Manly, Andrade,
22 Baddeley & Yiend, 1997), TEA and Flanker. For SART, we included error rates. For TEA,
23 we used “elevator counting” and “elevator counting with distraction” associated with

to the interference control measure. Therefore we did not include global RTs from Stroop in case a version was used that included separate blocks for the different conditions.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 sustained and selective attention, respectively. For ANT, we included orienting and alerting
2 measures, and if not available, center cue condition for orienting and no cue condition for
3 alerting.

4 *Verbal fluency.* Verbal fluency tasks are commonly used tools to assess both verbal
5 ability and executive control (e.g., Shao et al., 2014). Verbal fluency included the number of
6 produced words in the letter fluency or category fluency tasks. The latter was included for
7 comparison, with assumedly a smaller EF load and more emphasis on lexical competence.

8 **Interrater reliability.** The studies were coded by two raters with earlier experience in
9 meta-analyses. Interrater reliability was addressed via the following process: First, both raters
10 coded the same ten studies and checked that their coding was uniform. Disagreements were
11 resolved through discussion. Then both raters independently coded approximately half of the
12 remaining studies each. In addition, we randomly selected twenty⁷ studies from the whole set,
13 which both raters coded close to the end of the process. For these studies, the interrater
14 reliability was calculated. The interrater reliability (Cohen's Kappa) for the different
15 variables ranged from strong $\kappa = .834, p < .001$ to perfect agreement $\kappa = 1.000, p < 0.001$.

16

⁷ Seven of these randomly selected studies were excluded in the screening process after coding, and interrater reliability was analyzed only for the ones that were included in the final analyses.

BILINGUALISM AND EXECUTIVE FUNCTIONS

Table 1

Overview of the Included Domains, Task paradigms, Task versions, and Measures

Domain (<i>k</i>) ¹	Task Paradigm (<i>k</i>)	Task Version	Measure (type, <i>k</i>) ²
Inhibitory control (220)	Antisaccade (6)	Antisaccade Letters;	Interference Effect (RT, 2);
		Antisaccade Faces	Antisaccade Trials (Acc, 4)
	Flanker (56)	ANT;	Interference /
		Flanker Task;	Conflict Effect (RT, 39);
		Go-No/Go Flanker;	Incongruent Trials (RT, 17)
		LANT;	
	Go-No/Go (15)	Go-No/Go;	No/Go (Acc, 15)
		Go-No/Go Flanker	
	Simon (59)	Auditory Simon;	Interference /
		Simon 2-colors;	Conflict Effect (RT, 50);
		Simon Letters	Incongruent Trials (RT, 9)
	Stroop (84)	Auditory Stroop;	Interference /
Color-Word Stroop;		Conflict Effect (RT, 60; Acc, 6);	
Numerical Stroop;		Incongruent Trials (RT, 17; Acc, 1)	

BILINGUALISM AND EXECUTIVE FUNCTIONS

Domain (<i>k</i>) ¹	Task Paradigm (<i>k</i>)	Task Version	Measure (type, <i>k</i>) ²
		Spatial Stroop	
Set shifting (79)	Task Switching (45)	Color-Shape; Digits (Parity-Size); Quantity-Identity; Social Category; Word-Object; Words (Relational-Semantic)	Switching Cost (RT, 40); Switching Trials (RT, 5)
	TEA (7)	Elevator Counting with Reversal	Total Score
	TMT (12)	TMT	Effect (RT, 2; O, 2); Trail B (RT, 6; O, 2)
	WCST (15)	WCST	Perseverative Errors (6); Completed Categories (9)

BILINGUALISM AND EXECUTIVE FUNCTIONS

Domain (<i>k</i>) ¹	Task Paradigm (<i>k</i>)	Task Version	Measure (type, <i>k</i>) ²
Monitoring (188)	Flanker (52)	ANT;	Global RT (RT, 30);
		Flanker Task;	Congruent Trials (RT, 22)
		Go-No/Go Flanker;	
		LANT;	
	Linguistic Flanker		
	Simon (46)	Auditory Simon;	Global RT (RT, 21);
		Simon 2-colors;	Congruent Trials (RT, 25)
		Simon Letters	
	Stroop (44)	Auditory Stroop;	Global RT (RT, 21; Acc, 1); Congruent Trials (RT, 18; Acc, 4)
Color-Word Stroop;			
Numerical Stroop;			
Spatial Stroop			

BILINGUALISM AND EXECUTIVE FUNCTIONS

Domain (<i>k</i>) ¹	Task Paradigm (<i>k</i>)	Task Version	Measure (type, <i>k</i>) ²
	Task Switching (46)	Color-Shape; Digits (Parity-Size); Picture-Shape; Quantity-Identity; Social Category; Word-Object; Words (Relational-Semantic)	Mixing Cost (RT, 26); Global RT (RT,10); Repetition Trials (RT, 10)
Working Memory (251)	Complex Span (37)	Listening Span; Minus 2 Span; Operation Span; Reading Span; Stroop Span; Symmetry Span	Accuracy
	LNS (33)	Alpha Span; Matrix Span; Number Sequencing Span	Accuracy
	N-back (5)	Dual N-back;	N-bck effect (2-back minus 1-back; Acc)

BILINGUALISM AND EXECUTIVE FUNCTIONS

Domain (<i>k</i>) ¹	Task Paradigm (<i>k</i>)	Task Version	Measure (type, <i>k</i>) ²
		N-back	
	Simple Span (176)	Digit Span (FW; BW); Corsi Span (FW; BW); Spatial Span (FW; BW); Word Span (FW)	Accuracy
Attention (53)	ANT Alerting (16)	ANT;	Alerting Effect (RT, 11);
		LANT	No Cue Trials (RT, 5)
	ANT Orienting (20)	ANT;	Orienting Effect (RT, 14);
		LANT	Center Cue Trials (RT, 6)
	SART (7)	SART	SART Accuracy
	TEA Selective (7)	Elevator Counting with Distraction	Total Score
TEA Sustained (3)	Elevator Counting	Total Score	
Verbal fluency (100)	Category Fluency (53)	Category Fluency (all)	Total Score
	Letter Fluency (47)	Letter Fluency (all)	Total Score

1 Note. SART = The Sustained Attention to Response Task; TEA = Test of Everyday Attention; TMT = Trail-Making Test; WCST = Wisconsin Card Sorting Test; LNS =

2 Letter-Number Sequencing Task; ANT = Attention Network Task; LANT = Lateralized Attention Network Task; FW = forward; BW = backward.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 ¹ *k* refers to the number of effect sizes used in our final analyses (i.e., after pooling and before outlier exclusion).
- 2 ² Measures are presented in order of preference. Acc = accuracy; RT = reaction time; O = other.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 **Statistical Analyses**

2 For statistical analyses, we used metafor (Viechtbauer, 2010) for R (version 3.2.3; R
3 Core Team, 2015). The R script including all reported analyses, an output of all the analyses,
4 and the data file used in the analyses are available at (links omitted).

5 **Calculation of effect sizes.** To obtain an effect size for the difference between groups,
6 we calculated the standardized mean difference (SMD) using the escalc function. The
7 function documentation describes this argument as producing a Hedges' g by adjusting the
8 positive bias in the calculation for standardized mean differences. To obtain an unbiased
9 estimate of the sampling variances, we also set the vtype argument to "UB" (Viechtbauer,
10 2010).

11 In most tasks a lower value (of, e.g., Simon effect) indicated better performance.
12 However, because in some cases a higher value indicated better performance, the values for
13 group mean, SD , and sample size for the monolingual and bilingual group were first reversed,
14 that is, the values for the monolingual group were replaced with the corresponding values for
15 the bilingual group, and vice versa. This procedure allowed us to interpret positive effect-size
16 values as corresponding to a bilingual advantage, and negative effect size values as
17 corresponding to a bilingual disadvantage.

18 **Pooling effect sizes within comparisons.** In 35 instances, we pooled effect sizes
19 across highly similar outcome measures (e.g., verbal fluency scores for different letters
20 within the same task). To pool effect sizes, we replaced the rows for these measures with a
21 single row that included the average effect size and the average variance.

22 **Multi-level modelling.** In our data, effect sizes could not be considered entirely
23 independent. Compared to independent effect sizes, dependent effect sizes are not as
24 informative. When effect sizes are correlated, the information obtained from one estimate
25 overlaps with information obtained from another estimate. Unless this overlap is taken into

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 consideration, the amount of information is overestimated, and standard errors and
2 confidence intervals are underestimated, leading to a high number of Type I errors (e.g.,
3 Becker, 2000). To consider the dependency between effect sizes, we used a multi-level meta-
4 analysis (e.g., Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013)
5 considering dependency of the following forms:

6 First, a unique *pair* of groups (a bilingual group vs. a monolingual group) could be
7 repeatedly compared on more than one outcome measure (e.g., Simon and Flanker effects).
8 To consider this form of dependency, we coded for repeated comparisons within pairs.

9 Second, groups could also be repeatedly used in more than one pair. For example, two or
10 more monolingual groups could be compared to one bilingual group, or vice versa.⁸ To take
11 this latter form of dependency into consideration, we also coded for *clusters* of pairs within
12 which either a monolingual or a bilingual group was repeated. In this case, we made the
13 reasonable assumption that there would be no systematic difference depending on the group
14 type (monolingual or bilingual) that was repeated within the pairs.

15 These two forms of dependency were accounted for in a model with four levels of
16 variance. The variance within effect sizes, which is accounted for in a fixed effects meta-
17 analysis model, constitutes the first level. The second level is the variance between outcome
18 measures, which is accounted for in a random effects model. The third level is the variance
19 between different pairs. This level models the dependency of repeated comparisons within a

⁸ We used combined data for such two groups of bilingual or of monolingual participants if the groups did not differ regarding a moderator of interest (such as AoA) and such data were available. For example, combined data were used for two monolingual groups who only differed in their native language (e.g., Gutierrez, 2009), a variable that we were not interested in regarding monolinguals. In case the same group of bilingual participants were analyzed in the original study according to several different dimensions (e.g., early vs. late AoA and language dominance; Bennett, 2012), we chose the AoA division results for our analysis.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 pair. The fourth level is the variance between clusters of pairs. This level assumes that pairs
 2 within a cluster are more similar than pairs from other clusters and thus models the
 3 dependency within clusters.

4 We tested all three levels of variance by comparing the fit of the one-, two-, three-,
 5 and four-level models through likelihood-ratio tests using the `anova.rma`-function in `metafor`
 6 (Viechtbauer, 2010). In these comparisons, we used data that was trimmed from outliers (see
 7 below). All tests were statistically significant (Table 2). This indicates that the four-level
 8 model represents our data more adequately than any of the reduced models.

9
 10 Table 2

11 *Model Fit Indices, Model Comparison Statistics, and Variance Components*

Model Levels	Added Higher Level	Model Fit Indices		Model Comparison		Variance Components		
		AIC	LogLik	Models	LRT	σ^2_1	σ^2_2	σ^2_3
1. One		1526.84	-762.42					
2. Two	Measures	1064.47	-530.23	1 vs. 2	464.37***	0.11		
3. Three	Pairs	970.03	-482.02	2 vs. 3	96.43***	0.07	0.05	
4. Four	Clusters	950.98	-471.49	3 vs. 4	21.05***	0.05	0.02	0.05

12 *Note:* AIC = Akaike Information Criterion; LogLik = Log-Likelihood; LRT = Likelihood-
 13 Ratio Test. The Likelihood-ratio test statistic is tested against a chi-square distribution with 1
 14 degree of freedom. * $p < .05$; ** $p < .01$; *** $p < .001$

15
 16 The magnitude of the dependency of outcome measures within comparisons can be
 17 estimated with the intraclass correlation coefficient (*ICC*). The *ICC* is calculated by dividing
 18 the variance between comparisons by the sum of the variance between and within
 19 comparisons (i.e., $\sigma^2_1 / [\sigma^2_1 + \sigma^2_2]$). Hence, the *ICC* value also considers variance in the effect
 20 sizes that is attributed to differences between comparisons. When the variance within

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 comparisons is small in relation to the variance between comparisons, the *ICC* value is high.
2 If outcome measures within comparisons vary greatly so that each measure could equally
3 well belong to any one of the included comparisons, the correlation will drop towards zero. In
4 our final four-level model, the *ICC* for outcome measures within pairs was .129 and the *ICC*
5 for pairs within clusters was .464.

6 **Publication bias.** One of the most common methods to assess publication bias is the
7 trim and fill method (Duval & Tweedie, 2000). The trim and fill method is commonly
8 considered problematic (Peters, Sutton, Jones, Abrams, & Rushton, 2007). Because of this,
9 new methods including the p-curve (Simonsohn, Nelson, & Simmons, 2014) and different
10 types of regression-based models (Egger, Smith, Schneider, & Minder, 1997; Moreno et al.,
11 2009) have been developed. However, the application of these methods is complicated by our
12 multi-level approach. Common p-curve methods require independent effect sizes and perform
13 poorly if studies include so-called ghost variables (i.e., outcome measures that might
14 systematically be underreported due to non-significant findings; Bishop & Thompson, 2016;
15 Simonsohn et al., 2014). To test for asymmetry in the distribution of effect sizes, while
16 maintaining our four-level model, we therefore added the standard error (SE; or variance) for
17 each effect size as a predictor in our two main analyses (overall estimate of differences
18 between monolinguals and bilinguals without considering cognitive domain as a possible
19 moderator, and an analysis adding cognitive domain as a moderator). This should be
20 considered a close equivalent of the PET-PEESE method (Stanley & Doucouliagos, 2014).

21 In the precision-effect test (PET), the effect sizes are first regressed on their standard
22 errors in a weighted least-squares regression. If there is a significant and positive association
23 between effect sizes and their standard errors, this indicates a bias where studies with low
24 precision tend to report larger effect sizes (or, equivalently, that studies with low precision
25 and small effect sizes are underreported). The intercept (variance = 0) of the weighted least-

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 squares regression is taken as an estimate of an unbiased effect size in a hypothetical study
2 with perfect statistical power. In a simulation study (Stanley & Doucouliagos, 2014), the PET
3 method performed well when the true, unbiased effect was zero. When the true, unbiased
4 effect differed from zero, a better performance was observed when the standard error was
5 replaced with the variance. This test is called precision-effect test with standard error
6 (PEESE). The authors suggested that a PET test that reveals a significant association between
7 the effect sizes and their SE is followed up by a PEESE test.

8 The performance of the PET-PEESE in multi-level models was not evaluated by
9 Stanley and Doucouliagos (2014), but we consider it the best available method to correct
10 estimates in the presence of bias. This method also allows us to adjust for pertinent
11 moderators in the same model.

12 Prior to the PET-PEESE, we also conducted a visual inspection of two types of funnel
13 plots. In the first one, a contour-enhanced funnel plot (Peters, Sutton, Jobes, Abrams &
14 Rushton, 2008), each effect size is plotted against the inverse of its standard error. A vertical
15 reference line represents Hedges' $g = 0$, and the contours change shade at different levels of
16 two-tailed p-values. In the absence of publication bias, effect sizes will be distributed
17 symmetrically around the estimated overall effect, so that when precision increases, the
18 distribution of effects sizes becomes smaller. In the presence of publication bias, effects sizes
19 are expected to be asymmetrically distributed, with the distribution of studies in the bottom of
20 the funnel skewed towards the right.

21 As pointed out by Egger and colleagues (1997), asymmetry in a funnel plot can also
22 be explained by moderators. Because of this, we also used a method suggested in Soveri,
23 Antfolk, Karlsson, Salo and Laine (2017). To consider moderators, we plotted the residuals in
24 each cognitive domain against the SE (with lower SE higher on the y-axis). In this case, the

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 asymmetry can be evaluated in relation to the expected value. Contours can also be added to
2 this funnel plot.

3 Because an observed association between the effect sizes is not necessarily the result
4 of publication bias, we also investigated peer-review status in a moderator analysis.

5 **Moderator Analyses.** After the overall analysis of possible bilingual and
6 monolingual EF differences including all domains, we analyzed the effects in each cognitive
7 domain separately. Further moderator variables included peer-review status of the study
8 (peer-reviewed or other), task paradigm, nature of the task (verbal or nonverbal task),
9 whether language of the task (testing language) was bilinguals' L1 or L2, matching of the
10 groups (for age, education⁹, IQ, and vocabulary size), age group, AoA of L2, proficiency in
11 L2, immigrant status of the bilinguals, country in which the study was conducted (we only
12 included countries with at least five samples, and studies conducted in more than one country
13 were excluded from this analysis), and language pair of the bilinguals (similarly, only
14 language pairs with at least five samples were included).

15

16

Results

17

Descriptive Results

18 The final dataset included 152 studies, of which 106 were journal articles, 29 doctoral
19 dissertations, 13 other theses, and 4 other non-peer-reviewed studies. For descriptive
20 information about the participant- and task-related characteristics of the studies, as well as the
21 results, see Tables S3 and S4.

22

23

In most of the comparisons between monolingual and bilingual samples there was
more than one outcome measure, and our meta-analysis included 891 effect sizes in total. Of

⁹ Matching the groups for income was reported in very few studies and we therefore focused our SES analyses only on matching of education.

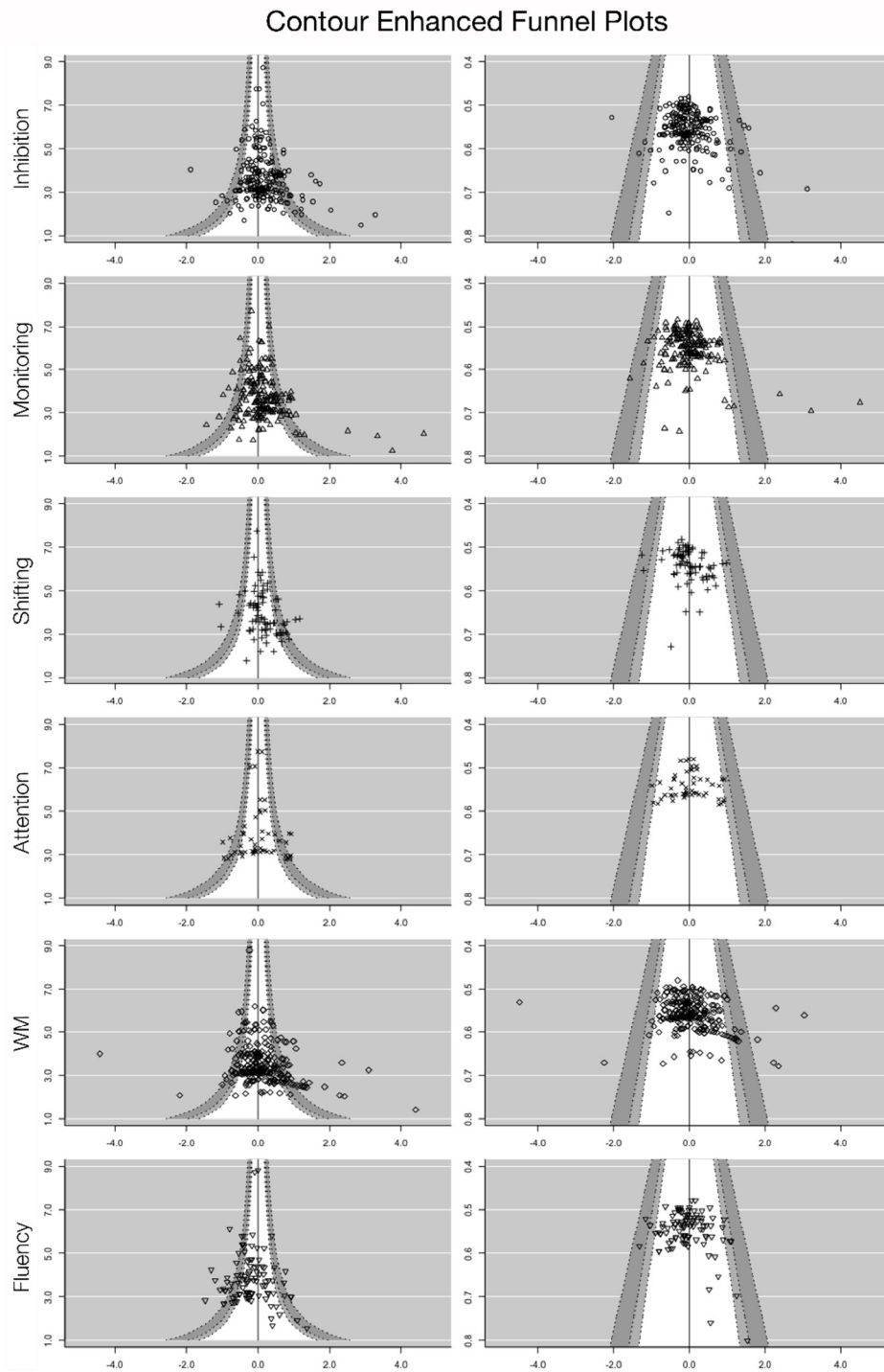
BILINGUALISM AND EXECUTIVE FUNCTIONS

1 these effect sizes 220 represented inhibition, 188 monitoring, 79 shifting, 53 attention, 251
2 WM, and 100 verbal fluency.

3 **Assessment of Bias and Data Screening**

4 To investigate possible reporting or publication bias, we first investigated the data
5 with regard to the distribution of study outcomes. We created six contour-enhanced funnel
6 plots, each representing the distribution of effect sizes within a chosen domain. We also
7 generated six plots, in which the residuals, after accounting for cognitive domain as a
8 moderator, were plotted. Here effect sizes are plotted in relation to the expected value. Thus,
9 a value of 0 means that the observed effect size is the same as the average effect size for the
10 entire domain (See Figure 2).

BILINGUALISM AND EXECUTIVE FUNCTIONS



1

2 *Figure 2.* Contour enhanced funnel plots for each cognitive domain (by row). Contours
 3 change shades at p -levels .1 (white), .05 (light grey), and .01 (dark grey). In the left column,
 4 effect sizes are plotted against their precision ($1/SE$) and the reference line is set at Hedge's g
 5 = 0. In the right column, effect sizes are plotted against the SE , and the reference line, against
 6 which residuals are plotted, indicates the synthesized effect within each domain.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 For inhibition, monitoring, and WM, the funnel plots showed a clear asymmetry, such
2 that effect sizes with high *SE* (or low precision) were more likely to show a bilingual
3 advantage than a bilingual disadvantage. For shifting, attention and verbal fluency, the funnel
4 plots suggest less bias. Moreover, when considering all domains together, some studies
5 appear as outliers. Either their effect size is very large or their *SE* is unexpectedly high in
6 relation to others. Note that the *SE* includes the variance from each level of the three-level
7 model, and not only the variance estimated in the original studies. Before proceeding to
8 further analysis we excluded potential outliers. The reason for this is twofold. First, this
9 would reduce asymmetry, which, in turn, would increase precision in subsequent analyses.
10 Second, because these were outliers also as to their *SE*, they could have an unduly strong
11 effect on PET-PEESE analyses leading to the corrected effect sizes being underestimates.

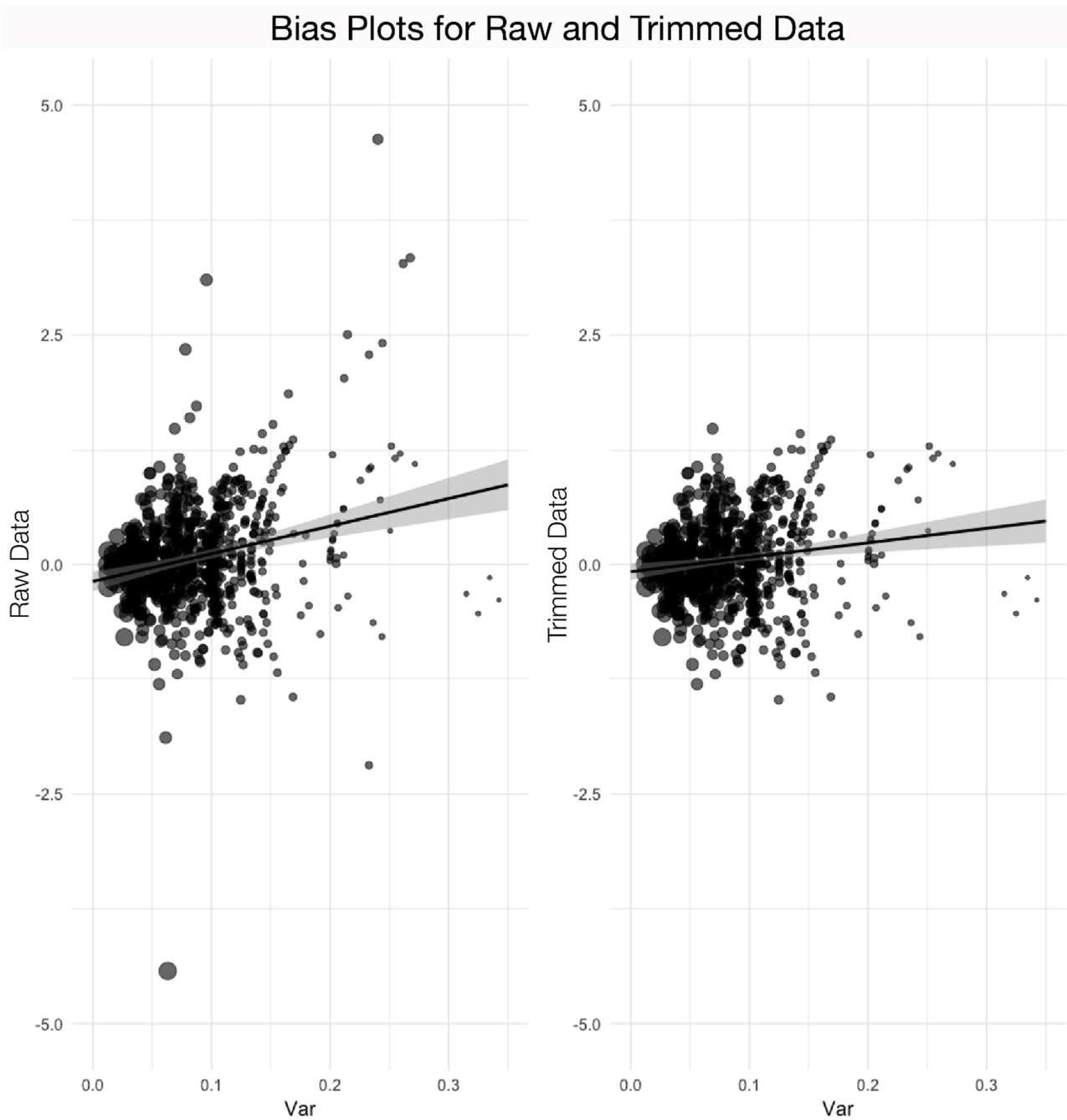
12 A visual examination of the funnel plots revealed a natural cut-off point at $SE = 0.6$.
13 After this, a few effect sizes remained outside a range of $g = -1.5$ to $g = 1.5$. These were also
14 removed. A total of 22 effect sizes (2.5%) were removed in this procedure. (See Table S4 for
15 the excluded effect sizes).

16 Bilingual Advantage

17 After trimming the data, we first investigated the bilingual advantage across all
18 included EF domains. We found a very small positive effect size in favor of bilingual groups,
19 $g = 0.06$ [0.01, 0.10], $p < .05$, Q_E [868] = 2139.79. Because of the asymmetry of effect sizes
20 observed in the contour-enhanced funnel plots, we used the PET-PEESE method to obtain a
21 corrected, unbiased effect size. Both the PET analysis and the PEESE analysis showed
22 significant negative associations between the effect sizes and their *SE* and variance ($p < .001$
23 and $p < .001$, respectively). The PET-PEESE corrected effect size was negative, $g = -0.08$ [-
24 0.17, 0.02], $p = .099$, but not statistically significant. We then investigated whether the
25 obtained results would be different if they were based on analyses conducted without

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 trimming the data. With outliers included, the estimated effect size was $g = 0.08$ [0.03, 0.14],
 2 $p < .01$, $Q_E [890] = 3173.75$. Again, a PET-PEESE correction yielded a statistically
 3 significant negative effect size, -0.29 [-0.38, -0.19], $p < .001$. This suggests that trimming
 4 data led to a less biased distribution of effect sizes and likely more reliable corrected and
 5 uncorrected estimates (See Figure 3). Because of the remaining bias, we decided to perform
 6 PET-PEESE analyses for estimates above $g = 0.2$ in the subsequent analyses.



7
 8 *Figure 3.* Plots visualizing the PEESE-corrected effect size. The line depicts the regression

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 slope for the association between the variance (Var; x-axis) and the effect size (y-axis). The
- 2 shaded area gives the 95% confidence intervals. The effect size of a hypothetical study with
- 3 perfect precision is estimated at $\text{Var} = 0$. All raw data are displayed in the left panel; data
- 4 trimmed for outliers are displayed in the right panel.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 **Bilingual Advantage by Cognitive Domain**

2 Because we expected the difference between monolinguals and bilinguals to be of
3 different magnitude in different EF domains, we investigated whether cognitive domain
4 moderated the outcome. We found that cognitive domain moderated the outcomes, $Q_M[5] =$
5 $53.37, p < .001$. The test for residual heterogeneity remained significant, $Q_E[863] = 2025.32,$
6 $p < .001$.

7 The moderator analysis yielded statistically significant positive outcomes indicating a
8 very small bilingual advantage for inhibition, shifting, and WM. The analysis also indicated a
9 small bilingual disadvantage for verbal fluency. For monitoring and attention, the analysis
10 indicated neither an advantage nor a disadvantage. To correct the estimates for the already
11 observed bias, we again used a PET-PEESE method. Adding the *SE* of each effect size as a
12 predictor to the model revealed a significant association between the size and direction of the
13 effects and the *SE* and variance ($p < .01$ and $p < .01$, respectively). After this correction,
14 statistically significant negative outcomes were found for attention and verbal fluency. Other
15 outcomes were not statistically significant. (See Figure 4).

16

17

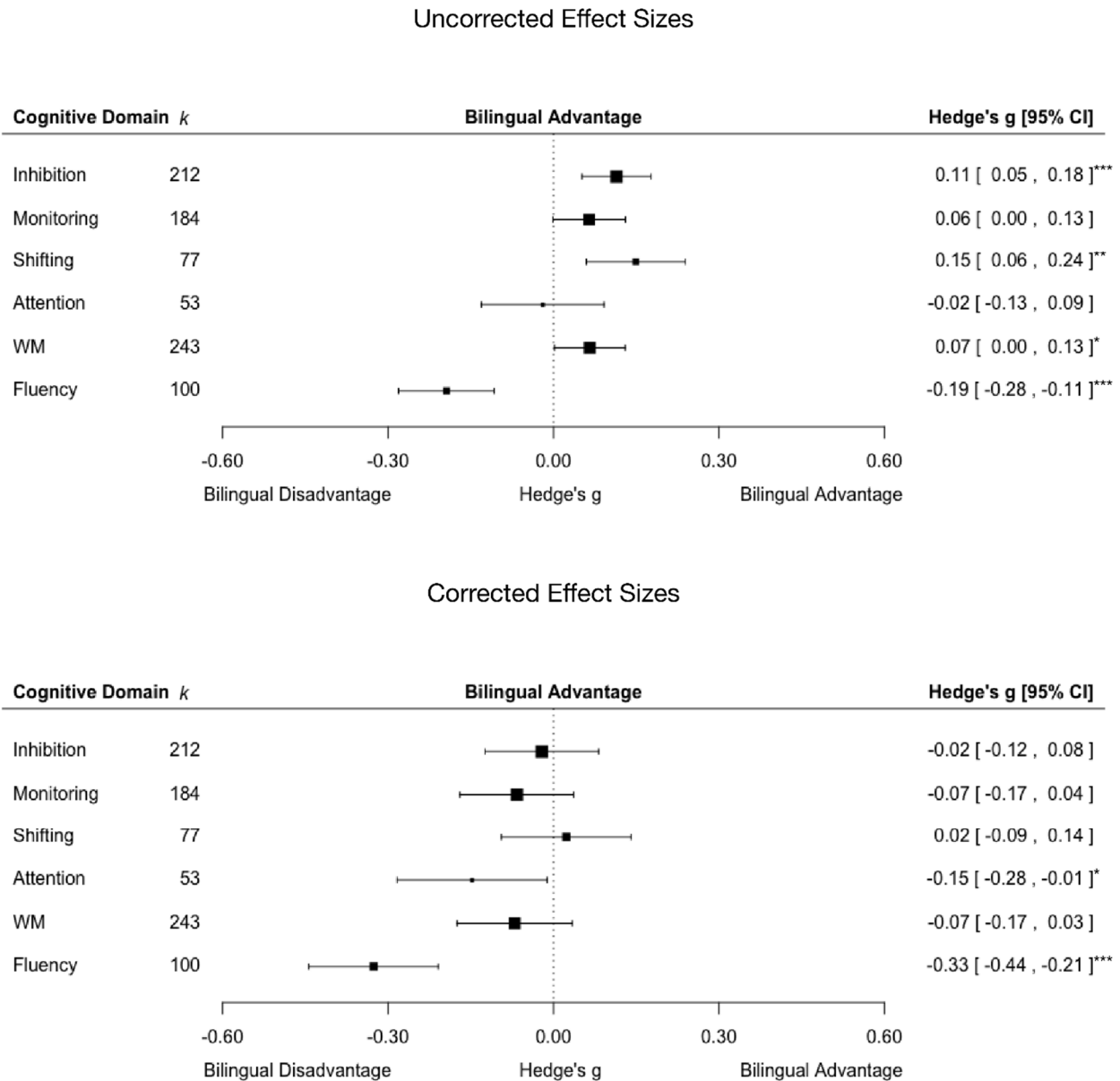
18

19

20

21

BILINGUALISM AND EXECUTIVE FUNCTIONS



1

2 *Figure 4.* For each cognitive domain, the figure displays synthesized effect sizes and 95%

3 confidence intervals (CI) for the comparison between monolinguals and bilinguals. Positive

4 values indicate a bilingual advantage and negative values indicate a bilingual disadvantage. k

5 = number of effect sizes. Uncorrected effect sizes are displayed in the upper panel and

6 corrected effect sizes are displayed in the lower panel. * $p < .05$; ** $p < .01$; *** $p < .001$

7

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Because the bias might be different in different cognitive domains and therefore lead
2 to either under or over-correction in individual domains, we also included the variance as an
3 interaction term together with cognitive domain in the PEESE-analysis. The slope was
4 relatively steep for shifting, WM, and verbal fluency; for inhibition, monitoring, and
5 attention, the slope was more horizontal. After correction, there was no evidence of a
6 bilingual advantage for inhibition, $g = 0.01$, $[-0.12, 0.14]$, $p = .867$, monitoring, $g = -0.04$, $[-$
7 $0.18, 0.09]$, $p = .520$, shifting, $g = -0.03$, $[-0.21, 0.16]$, $p = .782$, attention, $g = -0.06$, $[-0.32,$
8 $0.20]$, $p = .667$, or WM, $g = -0.14$, $[-0.29, 0.01]$, $p = .065$. The corrected estimates suggested
9 a statistically significant bilingual disadvantage for verbal fluency $g = -0.28$, $[-0.46, -0.10]$, p
10 $< .01$ (See Figure 5).

BILINGUALISM AND EXECUTIVE FUNCTIONS

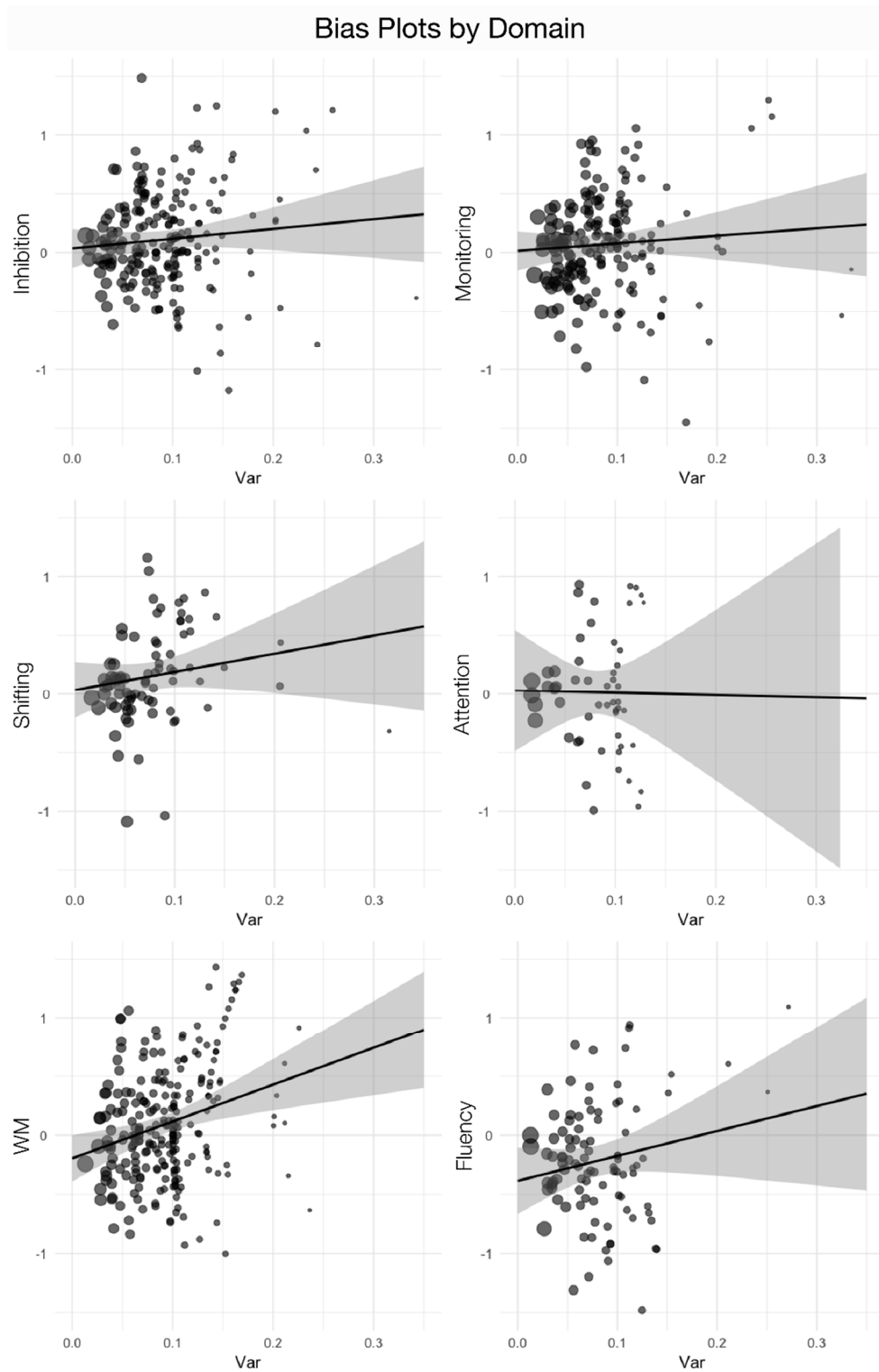


Figure 5. Scatter plots visualizing the PEESE-corrected effect sizes by each cognitive domain. The line depicts the regression slope for the association between the variance (Var; x-axis) and the effect size (y-axis). The shaded are gives the 95% confidence intervals. All panels include data after outlier exclusion.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Peer-Review Status

2 To investigate a possible source of the observed bias, we investigated whether
3 outcomes differed depending on whether the report had been peer-reviewed or not. To do
4 this, we conducted a set of analyses with peer-review status as a moderator. Peer-review
5 status did not moderate the outcome when including all domains, $Q_M [1] = 0.31, p = .580$. We
6 then analyzed whether outcomes differed depending on peer-review status for each of the
7 included domains. The estimated effect sizes were not statistically significant for inhibition,
8 $Q_M [1] = 1.26, p = .261$, monitoring, $Q_M [1] = 0.12, p = .732$, shifting, $Q_M [1] = 2.51, p =$
9 $.113$, attention, $Q_M [1] = 0.63, p = .426$, or verbal fluency, $Q_M [1] = 0.88, p = .349$. For WM,
10 the difference was statistically significant, $Q_M [1] = 5.29, p < .05$, such that the effect size
11 was smaller among peer-reviewed reports, $g = 0.02, [-0.08, 0.11], p = .745$, compared to
12 reports that had not been peer-reviewed, $g = 0.20, [0.07, 0.32], p < .01$.

13 To further investigate bias, we conducted separate PET-PEESE analyses for the two
14 groups of peer-review status. For peer-reviewed data, analyses revealed a statistically
15 significant association between effect sizes and their *SE* and variance ($p < .001$ and $p < .001$,
16 respectively). For other data, neither association was statistically significant ($p = .317$ and $p =$
17 $.436$, respectively).

18 Task Paradigms within Cognitive Domains

19 We then explored whether the outcomes within each domain were moderated by the
20 task used to measure EF. We found that task significantly moderated the estimates for
21 shifting and attention. In both cases, a medium-sized difference in favor of bilinguals was
22 found for TEA (for measures of attentional switching and selective attention), but not for
23 other tasks. TEA was, however, represented only by a low number of effect sizes. In shifting,
24 WCST also showed a significant positive effect, but it did not differ significantly from the
25 outcomes for the rest of the shifting tasks, which can be seen from the overlapping

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 confidence intervals. Due to the low numbers of effect sizes reported for TEA and WSCT, we
- 2 did not investigate the impact of bias with a PET-PEESE analysis. (See Table 3 for details).

BILINGUALISM AND EXECUTIVE FUNCTIONS

Table 3

Synthesized Effect Sizes and Confidence Intervals for Tasks within the Cognitive Domains

Domain	Task	Effect Size	95%CI		<i>p</i>	<i>k</i>	Moderator test		
		<i>g</i>	LB	UB			<i>Q_M</i>	<i>df</i>	<i>p</i>
Inhibition						212	0.25	4	.993
	Antisaccade	0.07	-0.21	0.35	.641	6			
	Flanker	0.11	0.00	0.21	.047	54			
	Go/Nogo	0.14	-0.10	0.38	.252	15			
	Simon	0.09	-0.01	0.20	.087	55			
	Stroop	0.12	0.03	0.20	.011	82			
Monitoring						184	9.12	4	.058
	Flanker	0.19	0.07	0.31	<.01	44			
	Go/Nogo	0.00	-0.43	0.43	.999	6			
	Simon	0.02	-0.11	0.14	.799	44			
	Stroop	0.10	-0.03	0.23	.143	44			
	TaskSwitching	-0.04	-0.16	0.09	.545	46			
Shifting						79	9.62	3	.022
	TEA(Switching)	0.55	0.22	0.88	<.01	7			
	TaskSwitching	0.10	-0.03	0.23	.127	45			
	TMT	-0.01	-0.22	0.21	.958	12			
	WCST	0.24	0.05	0.44	.016	15			
Attention						53	23.79	4	<.001
	Flanker(Alert)	-0.09	-0.28	0.10	.376	16			
	Flanker(Orient)	-0.07	-0.25	0.12	.470	20			
	SART	-0.16	-0.47	0.15	.309	7			
	TEA(Selective)	0.72	0.38	1.06	<.001	7			
	TEA(Sustained)	-0.05	-0.48	0.38	.818	3			
WM						243	3.50	3	.321
	Complex Span	0.08	-0.08	0.24	.325	35			

BILINGUALISM AND EXECUTIVE FUNCTIONS

	LNS	0.02	-0.13	0.17	.830	32			
	N-back	-0.29	-0.75	0.17	.211	5			
	Simple Span	0.09	0.01	0.18	.037	171			
Fluency						98	3.47	1	.062
	Category	-0.28	-0.41	-0.16	<.001	52			
	Letter	-0.17	-0.30	-0.04	.010	46			

Note: Positive effect sizes indicate a bilingual advantage; negative effect sizes indicate a bilingual disadvantage. g = Hedge's g , CI = Confidence intervals, LB = lower bound, UB = upper bound, k = number of effect sizes. TEA = Test of Everyday Attention; TMT = Trail Making Test; WCST = Wisconsin Card Sorting Test; Flanker(Alert) = ANT Alerting measure; Flanker(Orient) = ANT Orienting measure; SART = Sustained Attention to Response Task; WM = Working memory; LNS = Letter-Number Sequencing Task

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 **Verbal and Nonverbal Tasks**

2 Before running the analysis on the nature of the task, we removed the verbal fluency
 3 domain, as it only consists of verbal tasks, making comparisons to nonverbal tasks
 4 impossible. Overall, across all EF domains, there was a significant difference between the
 5 outcomes, $Q_M [1] = 17.35, p < .001$. The estimated positive effect size for nonverbal tasks, g
 6 $= 0.14 [0.09, 0.19], p < .001$, was larger compared to verbal tasks, $g = 0.01 [-0.04, 0.07], p =$
 7 $.603$.

8 We then repeated this analysis in each of the five remaining cognitive domains. For
 9 inhibition, $Q_M [1] = 0.54, p = .465$, and attention, $Q_M [1] = 1.13, p = .288$, there was no
 10 statistically significant difference between nonverbal and verbal tasks. The outcomes were
 11 moderated by whether a task was verbal or nonverbal in three domains: monitoring, $Q_M [1] =$
 12 $7.17, p < .01$, shifting, $Q_M [1] = 5.65, p < .05$, and WM, $Q_M [1] = 29.00, p < .001$. For
 13 monitoring, the estimated effect size was larger in nonverbal tasks, $g = 0.11 [0.03, 0.18], p <$
 14 $.001$, compared to verbal tasks, $g = -0.06 [-0.18, 0.05], p = .298$. For shifting, the estimated
 15 effect size was smaller in verbal tasks, $g = -0.01 [-0.17, 0.15], p = .912$, compared to
 16 nonverbal tasks, $g = 0.21 [0.10, 0.32], p < .001$. Also, for WM, the effect size was smaller in
 17 verbal tasks, $g = 0.01 [-0.08, 0.08], p = .962$, compared to nonverbal tasks, $g = 0.30 [0.18,$
 18 $0.41], p < .001$. Because previous analyses showed a strong bias in WM, we also investigated
 19 bias in nonverbal shifting and WM tasks. A PET-PEESE correcting for bias yielded a
 20 smaller, non-significant effect in both cases, $g = 0.06 [-0.14, 0.25], p = .585$ for shifting and g
 21 $= 0.10 [-0.11, 0.29], p = .376$ for WM.

22 **Testing Language**

23 The original studies included verbal tasks in languages that could be either the first or
 24 a second language of the bilingual sample. Because tasks performed in L2 could be expected
 25 to have an undue influence on the outcome, we also restricted our data to include only tasks

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 performed in the L1 of the bilinguals. In this case, the overall bilingual advantage was small
 2 and not statistically significant, $g = 0.07 [-0.05, 0.18]$, $p = .276$, $Q_E [108] = 336.91$. We then
 3 reran our analysis with cognitive domain as a moderator. Again, domain moderated the
 4 outcome, $Q_M [3] = 16.81$, $p < .001$. For inhibition, $g = 0.18 [-0.01, 0.37]$, $p = .060$, $k = 24$,
 5 monitoring, $g = 0.07 [-0.21, 0.34]$, $p = .639$, $k = 10$, and verbal fluency, $g = -0.17 [-0.34, -$
 6 $0.01]$, $p < .05$, $k = 34$, point estimates remained similar. For WM, $g = 0.30 [0.13, 0.47]$, $p <$
 7 $.001$, $k = 41$, the effect was slightly larger than before, but corrected towards null in a follow-
 8 up PET-PEESE, $g = 0.03 [-0.25, 0.32]$, $p = .824$. No data were available for shifting and
 9 attention.

10 **Matching of Groups**

11 Because not all studies matched their monolingual and bilingual samples for level of
 12 education, intelligence, size of vocabulary, and/or age, we also conducted follow-up analyses
 13 limiting our sample to only include data from studies with matched samples. We found few
 14 noteworthy differences between the outcomes of studies matching participants for our
 15 variables of interest and the outcomes when including data from all studies. For studies
 16 matching for vocabulary size, the previously estimated bilingual disadvantage for verbal
 17 fluency disappeared (i.e., was in the opposite direction but non-significant). For studies
 18 matching for intelligence and those matching for age, the estimated positive effect sizes in
 19 inhibition and shifting were slightly larger than previously but remained within the CI of
 20 prior estimates. A PET-PEESE analysis corrected outcomes towards null for shifting in
 21 studies matched for intelligence, $g = 0.13 [-0.04, 0.31]$, $p = .137$, and in studies matched for
 22 age, $g = -0.02 [-0.15, 0.19]$, $p = .830$. (See Table 4 for details).

BILINGUALISM AND EXECUTIVE FUNCTIONS

Table 4

Synthesized Effect Sizes from Studies Matching Samples for Education, Intelligence, Vocabulary, and Age

	<u>Education</u>			<u>Intelligence</u>			<u>Vocabulary</u>			<u>Age</u>		
	<i>g</i>	95% CI	<i>k</i>	<i>g</i>	95% CI	<i>k</i>	<i>g</i>	95% CI	<i>k</i>	<i>g</i>	95% CI	<i>k</i>
Overall	0.05	[-0.00, 0.10]	678	0.08*	[0.01, 0.14]	349	0.10**	[0.03, 0.16]	23	0.05	[-0.00, 0.11]	395
									0			
Inhibition	0.12**	[0.04, 0.19]	146	0.12**	[0.03, 0.22]	81	0.18**	[0.07, 0.30]	55	0.10*	[0.01, 0.18]	89
Monitoring	0.06	[-0.02, 0.13]	138	0.05	[-0.04, 0.14]	83	0.09	[-0.05, 0.22]	41	0.09	[-0.00, 0.18]	80
Shifting	0.11*	[0.01, 0.20]	71	0.23*	[0.10, 0.37]	31	0.21	[-0.00, 0.41]	16	0.27***	[0.14, 0.40]	34
Attention	-0.03	[-0.16, 0.09]	43	-0.05	[-0.26, 0.15]	13	-0.03	[-0.26, 0.20]	13	-0.01	[-0.20, 0.17]	15
WM	0.07	[-0.00, 0.14]	201	0.06	[-0.03, 0.15]	110	0.07	[-0.04, 0.18]	68	0.03	[-0.05, 0.11]	122
Fluency	-0.23***	[-0.32, -0.13]	79	-0.11	[-0.25, 0.03]	31	0.02	[-0.12, 0.16]	37	-0.15**	[-0.26, -0.05]	55

Note. Positive effects indicate a bilingual advantage; negative effects indicate a bilingual disadvantage. *g* = Hedge's *g*, CI = Confidence

intervals, *k* = number of effect sizes. * $p < .05$; ** $p < .01$; *** $p < .001$

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Age Group

2 We then investigated whether older participants showed more benefits of bilingualism
 3 than younger participants. Age group did not moderate the outcome across all EF domains, Q_M
 4 $[1] = 1.49, p = .222$. We also used the mean age of the bilinguals as a continuous predictor. We
 5 found no linear association between age and the difference between monolinguals and bilinguals,
 6 $g = -0.00 [-0.003, 0.002], p = .673$. Follow-up analyses revealed that there were no significant
 7 differences between age groups for any of the following domains considered separately:
 8 inhibition, $Q_M [1] = 0.41, p = .520$, monitoring, $Q_M [1] = 1.95, p = .163$, shifting, $Q_M [1] = 1.03,$
 9 $p = .311$, attention, $Q_M [1] = 0.92, p = .337$, and verbal fluency, $Q_M [1] = 0.36, p = .548$. For
 10 WM, age group moderated the outcomes, $Q_M [1] = 4.88, p < .05$. Samples with younger
 11 participants showed a very small difference in favor of bilinguals, $g = 0.11 [0.03, 0.19], p < .01,$
 12 and this estimate was larger than the estimated difference between older monolinguals and
 13 bilinguals, $g = -0.09 [-0.26, 0.07], p = .268$.

14 We also explored whether age group moderated the outcomes estimated for each of the
 15 25 included task paradigms (see Table S5). Age group moderated the outcomes only in the
 16 monitoring measure of Stroop (compared to younger participants, older participants showed less
 17 benefits of bilingualism) and the shifting measure of TaskSwitching (compared to younger
 18 participants, older participants showed more benefits of bilingualism). In both cases, only a
 19 limited amount of observations was available for older participants ($k = 7$ and $k = 5,$
 20 respectively).

21 Age of Acquisition

22 We then investigated whether AoA of L2 moderated the outcomes. To do this, each
 23 bilingual group was coded as either early acquisition or later acquisition (cut-off at 6 years).
 24 Outcomes across all EF domains was not moderated by AoA, $Q_M [1] = 0.95, p = .331$. Follow-up

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 analyses for each domain separately revealed that there was no significant moderation for
 2 inhibition, $Q_M [1] = 0.29$, $p = .592$, monitoring, $Q_M [1] = 0.42$, $p = .519$, shifting, $Q_M [1] = 0.02$,
 3 $p = .879$, or attention, $Q_M [1] = 0.25$, $p = .615$. For WM, AoA moderated the outcomes, $Q_M [1] =$
 4 5.12 , $p < .05$. Samples with later acquisition showed a smaller difference between monolinguals
 5 and bilinguals in WM, $g = 0.02 [-0.09, 0.12]$, $p = .735$, compared to samples with early
 6 acquisition, $g = 0.23 [0.07, 0.39]$, $p < .01$. A PET-PEESE analysis corrected the outcome for
 7 early acquisition towards null, $0.02 [-0.26, 0.29]$, $p = .912$. For verbal fluency, AoA also
 8 moderated the outcomes, $Q_M [1] = 4.92$, $p < .05$. In this case, samples with early acquisition
 9 showed a larger difference between monolinguals and bilinguals, $g = -0.52 [-0.82, -0.23]$, $p <$
 10 $.001$, compared to samples with later acquisition, $g = -0.15 [-0.30, -0.01]$, $p < .05$.

11 We also categorized samples according to another criterion, separating between samples
 12 in which participants had learned the second language before 12 years of age and others. With
 13 this categorization, there was no evidence of outcomes being moderated by AoA, $Q_M [1] = 0.19$,
 14 $p = .659$. Follow-up analyses revealed that there was no significant moderation for any of the
 15 domains considered separately, inhibition, $Q_M [1] = 0.12$, $p = .734$, monitoring, $Q_M [1] = 3.12$, p
 16 $= .077$, shifting, $Q_M [1] = 0.84$, $p = .358$, attention, $Q_M [1] = 0.95$, $p = .330$, WM, $Q_M [1] = 2.48$,
 17 $p = .115$, and verbal fluency, $Q_M [1] = 2.19$, $p = .139$.

18 **Language Proficiency**

19 After this, we investigated the influence of language proficiency on the outcomes. We
 20 first tested whether the difference between monolinguals and bilinguals was larger in samples
 21 with high proficiency in L2 compared to other samples. There was no significant difference in
 22 outcomes between these two types of samples, $Q_M [1] = 0.35$, $p = .557$. Neither were there any
 23 significant differences in outcomes between samples with high proficiency in their second
 24 language and other samples in any of the six cognitive domains: inhibition, $Q_M [1] = 2.79$, $p =$

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 .095, monitoring, $Q_M [1] = 0.17, p = .677$, shifting, $Q_M [1] = 0.00, p = .961$, attention, $Q_M [1] =$
 2 1.22, $p = .270$, WM, $Q_M [1] = 0.20, p = .653$, or verbal fluency, $Q_M [1] = 1.20, p = .274$.

3 **Immigrant Status**

4 Next, we investigated the potential moderating effect of bilingual participants' immigrant
 5 background, specifically whether 1) more than half, 2) less than half, or 3) none of the bilinguals
 6 were first-generation immigrants. Across all EF domains, immigrant status did not moderate the
 7 outcome, $Q_M [2] = 2.89, p = .235$. We then repeated this analysis in each of the separate
 8 cognitive domains. Immigrant status did not moderate the outcome for inhibition, $Q_M [2] = 2.22,$
 9 $p = .329$, monitoring, $Q_M [2] = 1.90, p = .388$, shifting, $Q_M [2] = 3.12, p = .210$, attention (no
 10 immigrant 2 observations in the data), $Q_M [1] = 2.24, p = .134$, WM, $Q_M [2] = 0.57, p = .751$, or
 11 verbal fluency, $Q_M [2] = 5.53, p = .063$.

12 **Country**

13 We also conducted an analysis of country as a moderator. We found that country did not
 14 significantly moderate the overall outcome, $Q_M [11] = 19.35, p = .054$. Moderator analyses for
 15 each domain separately revealed that country moderated the outcome only for shifting, $Q_M [4] =$
 16 15.34, $p < .01$, and attention, $Q_M [4] = 29.58, p < .001$. (See Table 5 for details).

17 **Language Pair**

18 Our sample of studies included bilingual individuals with different language pairs. We
 19 tested whether the outcome was moderated by language pair. We found that the language pair
 20 did not significantly moderate the outcome across all EF domains, $Q_M [7] = 12.63, p = .082$.
 21 Moderator analyses for each domain separately revealed that country moderated the outcome
 22 only for monitoring, $Q_M [5] = 11.43, p < .05$ (See Table 6 for details).

23

BILINGUALISM AND EXECUTIVE FUNCTIONS

Table 5

Effect Size Estimates and Confidence Intervals by Cognitive Domain and Country

Country	<u>Inhibition</u>			<u>Monitoring</u>			<u>Shifting</u>			<u>Attention</u>			<u>WM</u>			<u>Fluency</u>		
	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>
AUS	0.15	-0.17, 0.48	11			2			0			4	0.10	-0.52, 0.73	9			0
BRA	0.07	-0.15, 0.29	17	0.23	-0.00, 0.47	16			0	-0.01	-0.28, 0.26	6	0.19	-0.21, 0.59	7			0
CAN	0.13*	0.00, 0.25	46	0.00	-0.17, 0.18	27	0.11	-0.06, 0.28	17	-0.18	-0.37, 0.01	10	0.05	-0.08, 0.19	99	-0.13	-0.29, 0.03	53
CHE			0	0.30	-0.20, 0.81	6	0.24	-0.23, 0.71	6			0			0			0
CHN			4			4			2			0			0			0
FRA	-0.04	-0.38, 0.30	7			1			0			0			3			0
GRE			4			0			0	-0.44**	-0.75, -0.13	5			0			0
ITA			3			2			0			0			0			0
NZL	0.26	-0.06, 0.58	8			2			0			2			0			0
SPA	0.20	-0.10, 0.50	5	0.27*	0.05, 0.50	9	0.18	-0.09, 0.45	6			2			1			0
UK	0.06	-0.22, 0.34	11	-0.03	-0.29, 0.23	16	0.53***	0.27, 0.80	9	0.45***	0.25, 0.66	10	0.08	-0.39, 0.56	6	-0.31	-0.69, 0.06	10
USA	0.09	-0.00, 0.18	72	-0.04	-0.14, 0.07	62	-0.04	-0.17, 0.09	27	0.07	-0.12, 0.26	7	0.02	-0.10, 0.14	88	-0.36***	-0.56, -0.17	30

Note: Positive effect sizes indicate a bilingual advantage; negative effect sizes indicate a bilingual disadvantage. *g* = Hedge's *g*; CI = Confidence intervals; *k* = number of effect sizes. AUS = Australia; BRA = Brazil; CAN = Canada; CHE = Switzerland; CHN = China; FRA = France; GRE = Greece; ITA = Italy; NZL = New Zealand; SPA = Spain. Countries with less than five reported effect sizes were removed before analysis.

BILINGUALISM AND EXECUTIVE FUNCTIONS

Table 6

Effect Size Estimates and Confidence Intervals by Cognitive Domain and Language Pair

Language Pair	<u>Inhibition</u>			<u>Monitoring</u>			<u>Shifting</u>			<u>Attention</u>			<u>WM</u>			<u>Fluency</u>		
	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>	<i>g</i>	95%CI	<i>k</i>
ENG-CHI	0.13	-0.11, 0.37	11	0.31*	0.07, 0.55	12	0.14	-0.13, 0.42	6			4			2			4
ENG-DUT	-0.11	-0.55, 0.33	9			0			0			0	0.11	-0.49, 0.71	9			0
ENG-FRE	0.04	-0.12, 0.20	30	-0.02	-0.26, 0.23	16	0.11	-0.14, 0.35	13	-0.16	-0.57, 0.26	8	0.13	-0.08, 0.34	30	-0.12	-0.36, 0.12	23
ENG-KOR			3			3			2			2	-0.08	-0.43, 0.28	18			0
ENG-SPA	0.03	-0.09, 0.16	43	-0.00	-0.15, 0.15	28	0.10	-0.10, 0.30	14			4	-0.02	-0.19, 0.15	52	-0.38***	-0.60, -0.16	26
POR-HUN	0.05	-0.22, 0.32	11	0.25	-0.05, 0.54	10			0			4	0.23	-0.25, 0.71	5			0
SPA-CAT	0.20	-0.10, 0.50	5	0.28*	0.04, 0.51	9	0.19	-0.13, 0.51	6			2			0			0
OTHER	0.15***	0.06, 0.23	100	0.02	-0.07, 0.10	106	0.16*	0.02, 0.30	38	-0.06	-0.26, 0.14	29	0.10	-0.01, 0.21	127	-0.19*	-0.37, -0.02	45

Note: Positive effect sizes indicate a bilingual advantage; negative effect sizes indicate a bilingual disadvantage. *g* = Hedge's *g*; CI = Confidence intervals; *k* = number of effect sizes.

CAT = Catalan; CHI = Chinese (both Cantonese and Mandarin were categorized as Chinese); DUT = Dutch; ENG = English; FRE = French; HUN = Hunsrückish; KOR = Korean; POR = Portuguese; SPA = Spanish. Language pairs with less than five effect sizes were removed before analysis.

Discussion

Despite the substantial amount of research conducted during the past 15 years, the question of whether bilinguals outperform monolinguals in EF is still debated. Our comprehensive meta-analysis, including 891 effect sizes from 152 studies, investigated whether there is evidence for a bilingual advantage in EF in healthy adults, and if so, in which cognitive domains and task paradigms the bilingual advantage is consistently observed. Previous systematic reviews that include also adults in their analyses have suggested that an advantage could be observed in the domains of WM (Adesope et al., 2010; Grundy & Timmer, 2016) and conflict monitoring (Hilchey & Klein, 2011), but also in inhibitory control (Donnelly, 2016) and attention (Adesope et al., 2010). Further, we investigated a possible advantage in the domain of shifting (e.g., Prior & MacWhinney, 2010). Moreover, we tested whether we would see smaller advantages in the verbal fluency domain than in other domains, especially in category fluency (e.g., Luo et al., 2010). However, we found no systematic evidence of a bilingual advantage in adults in any of these EF domains after correcting for an observed publication bias. We also examined a number of moderator variables in order to test critical assumptions behind the bilingual training hypothesis and to see whether the variation in the outcomes between studies were due to the kinds of tasks used or participant populations tested. These analyses did not reveal any consistent support for the theoretical presuppositions concerning the bilingual advantage hypothesis.

More specifically, our initial analysis across all EF domains estimated a very small¹⁰ positive difference in favor of bilinguals, corresponding to less than 1% of the explained variation in outcomes, and this difference was the likely result of bias that remained in the data

¹⁰ In the discussion section, we use the guidelines for interpretation of effect sizes suggested by Cohen (1988): > 0.0 = very small difference; .2 = small difference, .5 = medium difference, .8 = large difference.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 after removing outliers. After correcting for the remaining bias, our analysis across all EF
2 domains no longer estimated any difference between monolinguals and bilinguals. Before
3 accounting for bias in the data, the analysis focusing on each EF domain separately estimated
4 very small differences in favor of bilinguals for inhibitory control, shifting, and WM, and a very
5 small difference in favor of monolinguals was estimated for verbal fluency. After correcting for
6 bias, no bilingual advantages were seen in any of the investigated EF domains: inhibitory
7 control, monitoring, shifting, attention, WM, or verbal fluency. In fact, only a small bilingual
8 disadvantage for verbal fluency and a very small bilingual disadvantage for attention remained.

9 Our results are in line with findings presented in Hilchey et al. (2015) and Paap et al.
10 (2015) that question the hypothesized bilingual advantage. However, the results do not
11 corroborate some of the findings of previous systematic reviews that reported positive effects of
12 bilingualism on some types of EF (e.g., Adesope et al., 2010; de Bruin et al., 2015b; Donnelly,
13 2016; Grundy & Timmer, 2016; Hilchey and Klein, 2011) or the narrative review by Bialystok
14 (2017) which presented support of the same hypothesis. Because some of the contradictions
15 between these reviews and meta-analyses are likely due to variation in inclusion criteria and
16 methodology, we want to highlight that the statistical analyses used in the current study allowed
17 the inclusion of a larger amount of data than has been used in previous studies.

18 Publication Bias

19 Despite including unpublished studies, we observed bias in the distribution of the
20 reported results, as demonstrated in the funnel plots and the PET-PEESE analyses. Studies with
21 low precision (i.e., small sample sizes) tended to show stronger positive effects than studies with
22 high precision, whereas null or negative effects were underrepresented in studies with low
23 precision. A set of moderator analyses revealed no major differences between results reported in
24 peer-review publications and other studies. However, separate PET-PEESE analyses for peer-

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 reviewed data and other data revealed a significant association between effect sizes and their
2 precision only in the former case. In the latter case, there was no evidence of such an association.
3 This suggests that small studies with low precision and large, positive effect sizes might be
4 overrepresented in the peer-reviewed literature, or that comparably small studies with large,
5 negative effect sizes are underrepresented. There are several possible reasons for this: Journals'
6 publication processes may have favored strong positive outcomes in support of the purported
7 bilingual advantage. The bias may also stem from the researchers' own decisions regarding
8 whether to pursue a peer-review publication or not, or their decisions about whether or not to
9 report all findings when intending to publish their results.

10 In an attempt to correct for the observed bias, we used the PET-PEESE method. Recent
11 modelling studies show that the PET-PEESE method performs relatively well when the sample
12 size is large and the true effect is zero or close to zero (Carter, Schönbrodt, Gervais & Hilgard,
13 2017), which is likely to be the case in the current study. Importantly, in some of the cases
14 corrections were based on a relatively limited number of data points which increases the risk of
15 under- and over-estimates. It is, therefore, important to note that the corrected effect sizes should
16 not be taken as "true values". The PET-PEESE method, like any other method to correct for bias,
17 *estimates* the effect size in the absence of bias. This estimate is perhaps best understood as an
18 educated guess. Nevertheless, the systematic correction of very small or small effect sizes
19 towards null here suggests that not too much emphasis should be put on isolated outcomes.
20 Because different methods can be used to account for publication bias, we encourage other
21 researchers to use our openly available data to evaluate how employing different methods may
22 affect the outcomes of the current study.

23 Due to the problems inherent in meta-analyzing biased data, we encourage pre-
24 registration of studies investigating the bilingual advantage. This would ensure that reporting

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 bias does not affect the outcome of future meta-analyses. Recent evidence also shows that
2 publication trends in this field are changing, as suggested by the bibliometric analysis by
3 Sanchez-Azanza et al. (2017), possibly leading to more balanced reporting in the future (see also
4 de Bruin & Della Sala, 2015).

5 **Moderator Variables**

6 We analyzed a number of moderator variables in order to test several preset hypotheses
7 that have been proposed to affect the magnitude of the purported bilingual EF advantage.

8 **Task-Related Moderator Variables**

9 Due to questionable convergent validity of many commonly used EF tasks, we
10 considered it critical to study whether a bilingual advantage is only observed in particular task
11 paradigms. The type of task significantly moderated the outcome only in the domains of shifting
12 and attention, and not in inhibition, monitoring, WM, or verbal fluency. For shifting, small to
13 medium differences in favor of bilinguals were seen in TEA and WCST, but not in other shifting
14 tasks. In the attention domain, a medium-sized difference in favor of bilinguals was seen in the
15 selective attention measure of TEA. Importantly, these estimates were based on very limited data
16 (seven effect sizes from two studies for the TEA tasks, and 15 effect sizes from eight studies for
17 the WCST), and we must therefore be cautious of drawing any definite conclusions from these
18 findings. Both of the shifting measures, that is, Elevator Counting with Reversal in TEA and
19 Perseverative Errors in WCST have been most closely related to shifting (Chan, Lai &
20 Robertson, 2006; Miyake et al., 2000); however, they are based on rather complex executive
21 tasks and are assumedly reflecting also other cognitive functions (see, e.g., Chan, Hoosain &
22 Lee, 2002; Robertson et al., 1996; Miyake et al., 2000). It is therefore difficult to speculate which
23 specific functions might account for the larger differences between monolinguals and bilinguals
24 observed in TEA and WCST, if these differences were confirmed by further research.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 With the assumption of weaker bilingual performance in verbal than nonverbal tasks
2 (Bialystok, 2009), we tested whether clearer bilingual advantages are seen in tasks with
3 nonverbal than verbal stimulus material. For verbal fluency, the observed small bilingual
4 disadvantage is in line with previous reports suggesting that bilingual participants score lower
5 than monolinguals in language tasks, such as word production (e.g., Gollan et al., 2008) or
6 recognition (Lehtonen et al., 2012; Lehtonen & Laine, 2003). Across all EF domains, the
7 difference between monolinguals and bilinguals was smaller for verbal than nonverbal tasks, as a
8 very small difference in favor of bilinguals was estimated in nonverbal tasks but not in verbal
9 tasks. Differences between nonverbal and verbal tasks were found in the domains of shifting,
10 monitoring and WM, but the effect sizes estimated for nonverbal tasks were very small or small
11 and disappeared after corrected for bias. Differences between verbal and nonverbal tasks may in
12 some of the original studies reflect the fact that the testing language was not always the bilingual
13 participants' L1, leading to unfair comparisons with monolingual participants in verbal tasks.
14 This was seen here in the domain of WM: When only analyzing the cases in which the testing
15 language was reportedly L1, the outcome for this domain was larger than when the testing
16 language was reportedly L2 and likely a weaker language of the bilinguals, thus putting
17 bilinguals in an unfair comparison with monolinguals. A further complicating factor is that L1
18 might not in all cases refer to the dominant language of the bilinguals, as long use of and
19 exposure to L2 may have altered the dominance relations between the languages. In any case, as
20 also pointed out by Grundy and Timmer (2016), it would be important to more explicitly report
21 the languages of task administration in future studies.

Participant-Related Moderator Variables

23 It has been reported that bilingual advantages in EF are better observed in older than
24 younger adults, possibly because bilingualism may beneficially affect the typical EF decline in

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 the elderly. One could also hypothesize that the “EF training period” has been longer for older
2 than younger bilinguals. However, our results did not support this hypothesis. We did not find
3 evidence that larger advantages would be observed in older, relative to younger, bilingual
4 participants compared to monolinguals in any EF domain. On the contrary, in the domain of
5 WM, there was a very small difference in outcomes in favor of the bilinguals in the young
6 groups which was not present in the older groups. In the explorative task analysis for age groups,
7 in one single task paradigm (TaskSwitching), there was a larger difference in favor of bilinguals
8 in older than younger groups, but this outcome was based only on five samples in the older
9 adults’ age group. We thus conclude that no systematic bilingual advantages were observed in
10 older or younger adults.

11 Our initial inclusion criteria for definitions of bilingualism were rather liberal, because
12 EF advantages have been reported in both early balanced bilingual individuals and those learning
13 a L2 later in life and reaching varying proficiency levels. We, however, tested whether
14 advantages will be larger when L2 was acquired early, due to the assumedly longer training of
15 EF in early bilinguals. In addition, we assumed that a higher attained L2 proficiency level will be
16 associated with larger advantages, as a stronger language is likely to pose more interference on
17 the control systems than a weaker one¹¹. When analyzing early bilingual participants who had
18 acquired two languages before the age of six, we saw some differences in the studied domains:
19 There was a small advantage in WM in favor of the early bilingual groups compared to
20 monolinguals, but not for bilinguals who had acquired an L2 at a later age. This positive outcome
21 in early bilinguals, however, vanished when correcting for publication bias. In verbal fluency, a

¹¹ Note that Paap et al. (2014) also present an alternative, opposite hypothesis: a large proficiency difference between the two languages could lead to larger gains. This is because a frequently used but less fluent L2 could entail less automatized language control mechanisms and a stronger need to inhibit L1 than a strong L2. Paap et al. (2014), however, found no evidence for either of these hypotheses in their study.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 medium-sized disadvantage was observed when only including early bilingual participants; for
2 late bilinguals the disadvantage was very small. Early bilingual individuals tend to have used the
3 two languages more equally than late bilinguals, leading to less exposure to one particular
4 language than is the case for monolingual individuals. This could possibly lead to a disadvantage
5 in tasks that require access to linguistic units, such as words (see, e.g., Gollan et al., 2008;
6 Lehtonen et al., 2012).

7 Another AoA categorization with a cutoff at the 12 years did not moderate the effects.
8 Similarly, the effects were not significantly moderated by the reported proficiency level. In sum,
9 we found no evidence supporting the bilingual training hypothesis according to which longer
10 bilingual exposure and increased competition demands from the other language would lead to
11 enhanced EF performance.

12 Many studies have argued that differential matching of bilingual and monolingual
13 participants can underlie the disparities in results of different studies. We investigated this issue
14 by repeating the analyses for the EF domains without including such studies, in which the
15 monolingual and bilingual samples had not been matched according to age, education, IQ, or
16 vocabulary size, respectively. The results from these analyses roughly corresponded to the results
17 from the previous analyses including all samples. In other words, we did not find evidence for
18 the view that matching issues would explain disparity between results of different studies.
19 Similarly, differences in immigration status of the bilingual participants did not moderate the
20 outcomes in any EF domain.

21 It has been argued that meta-analyses and systematic reviews may miss particular
22 variables related to the environment from which the bilingual and monolingual participants have
23 been recruited (Bak, 2016). In an attempt to take into account some of this variation in the data,
24 we analyzed the country in which the original study had been conducted. We found no evidence

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 that country would moderate outcomes across all EF domains. A similar analysis for each
2 domain separately suggested that country significantly moderated the outcome in shifting and
3 attention. In both cases, small to medium-sized differences in favor of the bilinguals were
4 observed in studies conducted in the UK. These outcomes differed significantly from outcomes
5 in the US for shifting, and Greece for attention. The number of effect sizes from these countries
6 and domains was, however, small (equal to or less than 10), and hence they were not corrected
7 for bias. These findings may be associated with the use of particular tasks in a country, such as
8 the use of TEA in the UK (TEA was used in seven out of nine comparisons included from the
9 UK for shifting, and in all comparisons included for attention).

10 We also analyzed whether the language pair of the bilingual groups would moderate the
11 possible bilingual EF advantage. On the basis of previous proposals, this could be the case because
12 having two structurally or lexically similar languages might increase the competition demands they put
13 to one another and hence lead to more intensive training of inhibitory control. We found no evidence
14 that the language pair would moderate outcomes across all EF domains. A moderation effect was seen
15 only in the domain of monitoring. The small advantages observed in monitoring for English-Chinese
16 and Spanish-Catalan bilinguals compared to monolinguals were only larger than the difference
17 estimated for English-Spanish bilinguals. These differences were, however, based on 12 (English-
18 Chinese), nine (Spanish-Catalan), and 28 (English-Spanish) comparisons. Finding a larger difference in
19 favor of Spanish-Catalan bilinguals than other language groups would be in line with the
20 abovementioned hypothesis; however, this conclusion is opposed by the equally large difference in
21 favor of English-Chinese bilinguals, two languages with little structural or lexical overlap. There is also
22 no apparent theoretical reason for why these findings would only be observed specifically in
23 monitoring measures and not in other EF domains.

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 Language pair and country are variables likely to have interwoven different cultural or
2 environmental factors. For example, particular cultures have been associated with better EF
3 performance. Studies have, for instance, reported better performance in children from Eastern than
4 Western cultures (Yang & Yang, 2016; Tran, Arredondo & Yoshida, 2015). In studies comparing
5 monolinguals from one culture to bilinguals from another, it is thus possible that cultural factors may
6 account for some of the observed EF differences between monolinguals and bilinguals. In line with
7 this, the small advantage observed for Spanish-Catalan bilinguals may be explained by studies
8 comparing different kinds of bilingual vs. monolingual populations. In most studies investigating the
9 effects of bilingualism on the monitoring capacity with Spanish-Catalan bilinguals, the bilinguals came
10 from another geographical area in Spain than the monolinguals. Recruiting groups from an urban vs.
11 more rural region may introduce cultural or socio-economic confounds to the comparisons and thus in
12 fact account for the differences originally interpreted to be due to bilingualism of the participants.

13 One could also speculate along the lines of the Adaptive Control hypothesis (Green &
14 Abutalebi, 2013) that particular bilingual groups might use the languages more strictly with separate
15 speakers (dual-language context) which assumedly poses more demands on EF than using the
16 languages in contexts where both languages can be spoken interchangeably (so-called opportunistic
17 planning, see Green & Abutalebi, 2013). Further research needs to investigate whether such a dual-
18 language context, for instance, is a typical language use pattern in Chinese-English and Spanish-
19 Catalan bilinguals. In addition, future studies will have to empirically investigate whether such
20 language use patterns could be directly associated with differential EF gains, as proposed by Green and
21 Abutalebi (2013; for an example of such a study, see Hartanto & Yang, 2016).

Limitations and Future Directions

23 Taken together, our meta-analysis provides no systematic evidence for a general,
24 systematic bilingual advantage in EF in adult samples. If some enhancement of cognitive control

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 functions exists due to bilingualism, it is restricted to very specific circumstances, and its
2 magnitude and extent are modest.

3 Many authors have also commented that bilingualism is not a unitary phenomenon,
4 making it problematic to use it as a categorical variable (e.g., Bialystok, 2017). What is
5 admittedly complicating the research area is that a multitude of factors is likely to affect
6 individuals' cognitive abilities, and it is difficult to control for all of them in the studies of this
7 type. In fact, an important issue contributing to the mixed results in the field has been the
8 inherent weaknesses of the natural groups designs of the studies (Hakuta, 1986, as cited in Klein,
9 2016; Author & Author, submitted). When compared to a typical cognitive training study, the
10 setup represents a rather weak research design: In bilingualism studies that can be taken as
11 studies on "natural training" of EF, randomization to bilingual and monolingual groups and pre-
12 post comparisons are normally not possible, and the specific contents of the assumed EF training
13 are also not apparent (Author & Author, submitted).

14 Thus far only a few studies have introduced longitudinal intervention designs, including
15 language learning or training that assumedly resembles aspects of bilinguals' language behaviors, such
16 as language switching. Adult bilinguals participating in ten days of language switching training showed
17 improved performance at post-test in a cognitive control task when compared to a passive control
18 group (Zhang, Kang, Wu, Ma & Guo, 2015). Janus, Lee, Moreno, and Bialystok (2016), in turn,
19 investigated effects of short-term second-language training camp on 4–6-year-old children's nonverbal
20 abilities. They reported improvements in specific tasks involving EF, but the improvements were
21 similar to the children participating in a music camp. Sullivan, Janus, Moreno, Astheimer, and
22 Bialystok (2014) tested students taking either an introductory Spanish ("training group") or an
23 introductory Psychology course ("control group") before and after the 6-month courses. Modulations
24 were seen in the ERPs in a go-nogo task for language learners only, but no behavioral differences were

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 observed between the groups. Bak, Long, Vega-Mendoza, and Sorace (2016) compared EF
2 performance, as measured with TEA, in adult participants taking a one-week language course to the
3 performance of matched active and passive monolingual control groups. In the attentional switching
4 measure of TEA, the language learner group showed the largest improvement at posttest, significantly
5 different from that of the passive control group. The active control group showed intermediate
6 performance that did not significantly differ from either the language group or the passive control
7 group. Finally, Ramos, Fernández García, Antón, Casaponsa, and Duñabeitia (2017) studied healthy
8 monolingual seniors learning a new language for a year. Post-test performance in a nonverbal
9 switching task was not improved from pre-test performance for this group in comparison to a matched
10 passive control group. In sum, although clear behavioral EF improvements due to language learning or
11 language switching training in comparison to active control groups have not been observed in these
12 studies and although these studies have not used a fully random assignment to groups, they
13 nevertheless demonstrate how more solid experimental designs can be implemented in this field.

14 Another way to circumvent the problems of the cross-sectional designs and to make
15 progress in this area of research might be to utilize an individual differences approach and to
16 identify potential connections between features of the individuals' bilingual experience and
17 cognitive performance (Author & Author, submitted; Bialystok, 2017). Such studies, using
18 within-group correlative analyses, have already investigated how frequency of everyday
19 language switching and being involved in different kinds of interactional contexts (see, e.g.,
20 Green & Abutalebi, 2013) is associated with EF performance (see, e.g., Hartanto & Yang, 2016;
21 Jylkkä et al., 2017; Soveri et al., 2011a; Verreyt, Woumans, Vandelandotte & Szmalec, 2016).

22 The present meta-analysis only focused on healthy adults, and thus does not address the
23 proposed EF advantages in children or the question of possible later onset of dementia symptoms
24 in bilingual individuals. In their systematic review, Hilchey and colleagues (2015) reported that

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 larger advantages may in fact be observed in children than in adults. A challenge in comparing
2 and summarizing studies on children is the variety of task versions that children of different ages
3 need. Moreover, even if studies would consistently show that a bilingual advantage in children
4 exists, our results provide no reliable evidence for a bilingual advantage in adulthood, at least in
5 the cross-sectional data analyzed here. Observing an advantage only in children would naturally
6 limit the scale and significance of the putative phenomenon. With regard to risk of dementia in
7 older bilinguals, a recent meta-analysis by Mukadam, Sommerlad and Livingston (2017)
8 reported that prospective studies do not show compelling evidence for bilingualism protecting
9 from cognitive decline. According to their analysis, there is more evidence for such positive
10 effects in retrospective studies, but with these studies, the authors raise the issue of confounding
11 variables.

12 A meta-analysis by Zhou and Krott (2016) investigated the role of a seemingly trivial
13 aspect of data analysis of the bilingual advantage, namely the data trimming procedure. Their
14 hypothesis was that long RTs can be taken to reflect lapses of attentional control, and if
15 bilinguals have fewer long responses, there could be a difference to monolinguals in the tail of
16 the distribution. Their report on 68 effect sizes from 33 studies suggested that the time allowed to
17 respond affected the likelihood of seeing a bilingual advantage in healthy children and adults.
18 Studies including longer responses were more likely to report a bilingual advantage in nonverbal
19 inhibition tasks. This aspect of the original studies was not analyzed in the present study.

20 One well-known challenge in this field is that the tasks used to measure EF do not
21 correlate particularly strongly with one another. Thus, more work should be directed in studying
22 the general EF architecture and developing reliable and valid tests to measure its components.

23 Bilingualism, like several other sustained experiences such as practicing music (Münte,
24 Altenmüller & Jäncke, 2002), has been associated with particular neurocognitive signatures and

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 structural changes in the brain (for reviews, see, e.g., Abutalebi, 2008; Abutalebi & Green, 2016;
2 Bialystok, 2017; García-Pentón et al., 2015; Li et al., 2014). Some published fMRI studies have
3 shown different neural activation patterns in EF tasks or resting state connectivity for bilinguals
4 than monolinguals, with activation differences observed particularly in the anterior cingulate,
5 prefrontal regions, and subcortical structures. In addition, differences in ERPs have been
6 observed in brain responses associated to EF and attention and often assumed to reflect better
7 processing capacity in bilinguals. Notably, such effects in neural activation have often been
8 reported in the absence of behavioral differences between groups. In such cases, it may be
9 difficult to know whether bilingualism-related activation increases or decreases or ERP
10 modulations truly reflect increased processing efficiency, as the results have often been
11 interpreted (see, e.g., Bialystok, 2017; Sullivan et al., 2014; see also Paap et al., 2015).

12 Furthermore, structural differences related to bilingualism or language learning have been
13 shown in regions and pathways associated with language processing and cognitive control. Such
14 results have been reported both in grey-matter measures and in the integrity of white-matter
15 tracts. The reported differences, particularly in the grey-matter measures, have been quite
16 variable, likely at least partly because of heterogeneity in the analysis methods used and
17 populations studied (García-Pentón et al., 2015).

18 These kinds of examples of experience-dependent brain plasticity are interesting in their
19 own right, but what remains to be investigated in future research are the underlying reasons as
20 well as the possible functional significance and behavioral correlates of these modulations.
21 Neural measures were outside the scope of the present study. However, based on the current
22 results, the reported neural differences between bilingual and monolingual adults are unlikely to
23 reflect any general bilingualism-related EF advantages with direct behavioral consequences. Our
24 meta-analysis also leaves out particular other cognitive skills that have previously been

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 associated with superior performance in bilingual individuals. Such domains include, for
2 example, metalinguistic abilities and divergent thinking in which Adesope and colleagues (2010)
3 demonstrated a bilingual advantage. It is possible that future studies will accumulate evidence on
4 such other types of cognitive advantages of bilingualism. However, even if no extra-linguistic
5 cognitive consequences are found, the main advantage of bilingualism—the ability to
6 communicate in different languages with its personal and social consequences—will always
7 remain.

8 **Conclusions**

9 The present meta-analysis of 152 studies and 891 comparisons of bilinguals' and
10 monolinguals' performance in six EF domains does not support the view of bilingualism being
11 associated with an advantage in cognitive control functions in adults. The observed very small
12 effect sizes in the domains of inhibitory control, shifting, and WM disappeared when correcting
13 for publication bias. We also did not find systematic evidence supporting the bilingual advantage
14 hypothesis, and studies that included better matched participant groups did not show consistently
15 stronger advantages, either. In verbal fluency tasks, evidence for a small bilingual disadvantage
16 was observed, assumedly because balanced use of two languages may lead to less exposure to
17 and experience of using each individual language. We also observed that null and negative
18 findings were underreported in studies with small samples, which highlights the need of pre-
19 registration practices to be more widely adopted in the field.

20 **Acknowledgements**

21 We are indebted to all the authors who kindly responded to our queries and who provided
22 additional data to the meta-analysis (for a list of authors who were able to provide additional
23 data, see Table S2).

24

BILINGUALISM AND EXECUTIVE FUNCTIONS

1 **References**

2

3 For the studies included in the meta-analysis, see Table S6.

4

5 Abutalebi, J. (2008). Neural aspects of second language representation and language control.

6 *Acta Psychologica*, 128, 466-478. <https://doi.org/10.1016/j.actpsy.2008.03.014>

7 Abutalebi, J. & Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural

8 adaptation and reserve. *Bilingualism: Language and Cognition*, 19(4), 689-698.9 <https://doi.org/10.1017/S1366728916000225>

10 Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and

11 Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*,12 80(2), 207–245. <http://doi.org/10.3102/0034654310368803>

13 Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., ...

14 Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from

15 children, *Frontiers in Psychology*, 5, 1–12. <http://doi.org/10.3389/fpsyg.2014.00398>

16 Baddeley, A. (2000). The episodic buffer: a new component of working memory? Trends in

17 Cognitive Sciences, 4(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

18 Bak, T. H. (2013). The importance of looking in dark places. *Amyotrophic Lateral Sclerosis &*19 *Frontotemporal Degeneration*, 14(1), 1–2. <http://doi.org/10.3109/21678421.2013.760150>20 Bak, T. H. (2016). Cooking pasta in La Paz. *Linguistic Approaches to Bilingualism*, 6(5), 699–21 717. <http://doi.org/10.1075/lab.16002.bak>

22 Bak, T. H., & Alladi, S. (2016). Bilingualism, dementia and the tale of many variables: Why we need

23 to move beyond the Western World. Commentary on Lawton et al. (2015) and Fuller-Thomson

24 (2015). *Cortex*, 74, 315-317. doi: 10.1016/j.cortex.2015.04.025.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and
2 practice: The impact of intensive language learning on attentional functions. *PLoS ONE*,
3 *11*(4), 1–11. <http://doi.org/10.1371/journal.pone.0153485>
- 4 Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence
5 cognitive aging? *Annals of Neurology*, *75*(6), 959–963. <http://doi.org/10.1002/ana.24158>
- 6 Barkley, R. A. (2012). *Executive Functions: What they are, how they work, and why they evolved*. New
7 York: The Guildford Press.
- 8 Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & E. D. Brown (Eds.), *Handbook*
9 *of applied multivariate statistics and mathematical modeling* (pp. 499–525). Orlando: Academic
10 Press.
- 11 Bennett, J. (2012). *Linguistic and cultural factors associated with phonemic fluency*
12 *performance in bilingual hispanics*. (Doctoral dissertation). Retrieved from ProQuest
13 Dissertations and Theses Database. (UMI No. 3553808)
- 14 Bialystok, E. (2001). *Bilingualism in Development. Language, Literacy, and Cognition*. Cambridge,
15 UK: Cambridge University Press.
- 16 Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism:*
17 *Language and Cognition*, *12*(1), 3. <http://doi.org/10.1017/S1366728908003477>
- 18 Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience.
19 *Psychological Bulletin*, *143*(3), 233–262. <http://doi.org/10.1037/bul0000099>
- 20 Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D., (2010). Word mapping and executive
21 functioning in young monolingual and bilingual children. *Journal of Cognition and Development:*
22 *Official Journal of the Cognitive Development Society*, *11*, 485-508. doi:
23 10.1080/15248372.2010.516420

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and
2 cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290–303.
3 <http://doi.org/10.1037/0882-7974.19.2.290>
- 4 Bialystok, E., Craik, F., & Luk, G. (2008a). Cognitive control and lexical access in younger and
5 older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*
6 *34*(4), 859.
- 7 Bialystok, E., Craik, F. I., & Luk, G. (2008b). Lexical access in bilinguals: Effects of vocabulary
8 size and executive control. *Journal of Neurolinguistics, 21*(6), 522-538.
- 9 Bialystok, E., Craik, F., & Luk, G. (2009). “Cognitive control and lexical access in younger and
10 older bilinguals”: Correction. *Journal of Experimental Psychology: Learning, Memory, and*
11 *Cognition, 35*(3), 828–828. <http://doi.org/10.1037/a0015638>
- 12 Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain.
13 *Trends in Cognitive Sciences, 16*(4), 240-250. <http://doi.org/10.1016/j.tics.2012.03.001>
- 14 Bialystok, E., Kroll, J. F., Green, D. W., Macwhinney, B., & Craik, F. I. M. (2015). Publication
15 Bias and the Validity of Evidence : What’s the Connection? *Psychological Science, 26*(6),
16 944–946. <http://doi.org/10.1017/S01427>
- 17 Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults
18 *Bilingualism. 15*, 397-401. doi: 10.1017/S136672891100040X
- 19 Bialystok, E. & Viswanathan, M. (2009). Components of executive control with advantages for
20 bilingual children in two cultures. *Cognition, 112*(3), 494-500.
21 <http://doi.org/10.1016/j.cognition.2009.06.014>.
- 22 Bishop, D. V., & Thompson, P. A. (2016). Problems in using p-curve analysis and text-mining to
23 detect rate of p-hacking and evidential value. *Peer J, 4*.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in
2 Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*(3), 610–629.
3 <http://doi.org/10.1017/S1366728913000564>
- 4 Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict
5 monitoring and cognitive control. *Psychological Review*, *108*, 624–652.
- 6 Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017, October 2). Correcting for
7 bias in psychology: A comparison of meta-analytic methods. Retrieved from
8 psyarxiv.com/9h3nu
- 9 Chan, R. C. K., Hoosain, R., & Lee, T. M. C. (2002). Reliability and validity of the
10 Cantonese version of the Test of Everyday Attention among normal Hong Kong
11 Chinese: a preliminary report. *Clinical Rehabilitation*, *16*, 900–909.
12 <http://doi.org/10.1191/0269215502cr574oa>
- 13 Chan, R. C. K., Lai, M. K., & Robertson, I. H. (2006). Latent structure of the Test of
14 Everyday Attention in a non-clinical Chinese sample. *Archives of Clinical
15 Neuropsychology*, *21*(5), 477–485. <http://doi.org/10.1016/j.acn.2006.06.007>
- 16 Chong De La Cruz, I. A. (2016). *The role of language profiles in complex driving environments*.
17 (Doctoral dissertation). Retrieved from ProQuest. (10007414)
- 18 Coderre, E. L., van Heuven, W. J. B., & Conklin, K. (2013). The timing and magnitude of Stroop
19 interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and
20 Cognition*, *16*(2), 420–441. <http://doi.org/10.1017/S1366728912000405>
- 21 Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual
22 advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–
23 149. <http://doi.org/10.1016/j.cognition.2009.08.001>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control
2 their lexicalization process? Inhibitory and language-specific selection mechanisms are both
3 functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5),
4 1057–1074. <http://doi.org/10.1037/0278-7393.32.5.1057>
- 5 De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control
6 is shared by task and language switching. *Journal of Cognitive Neuroscience*, 27(9), 1752-
7 65. doi: 10.1162/jocn_a_00817
- 8 de Bruin, A., Bak, T. H., & Della Sala, S. (2015a). Examining the effects of active versus
9 inactive bilingualism on executive control in a carefully matched non-immigrant sample.
10 *Journal of Memory and Language*, 85, 15–26. <http://doi.org/10.1016/j.jml.2015.07.001>
- 11 de Bruin, A., & Della Sala, S. (2015). The decline effect: How initially strong results tend to
12 decrease over time. *Cortex*, 73, 375–377. <http://doi.org/10.1016/j.cortex.2015.05.025>
- 13 de Bruin, A., Treccani, B., & Della Sala, S. (2015b). Cognitive advantage in bilingualism an
14 example of publication bias? *Psychological Science*, 26(1), 99-107.
- 15 de Bruin, A., Treccani, B., & Della Sala, S. (2015c). The connection is in the data: We should
16 consider them all. *Psychological Science*, 26(6), 946-949.
- 17 Deslauries, G. (2008). *The role of selective attention in foreign accented speech perception*. (Master's
18 thesis). Retrieved from ProQuest. (1466563)
- 19 Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict
20 monitoring not response inhibition. *Brain and Cognition*, 56(2), 165-176.
- 21 Donnelly, S. (2016). *Re-examining the bilingual advantage on interference-control and task-*
22 *switching tasks: A meta-analysis*. (Doctoral Dissertation). *CUNY Academic Works*.
23 http://academicworks.cuny.edu/gc_etds/762

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., &
2 Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or
3 reality? *Experimental Psychology*, *61*(3), 234-251. [http://doi.org/10.1027/1618-](http://doi.org/10.1027/1618-3169/a000243)
4 [3169/a000243](http://doi.org/10.1027/1618-3169/a000243)
- 5 Duncan, H. D., Segalowitz, N., & Phillips, N. A. (2016). Differences in L1 linguistic attention
6 control between monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *19*(1),
7 106-121. <http://doi.org/0.1017/S136672891400025X>
- 8 Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot based method of testing and
9 adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.
- 10 Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by
11 a simple, graphical test. *BMJ*, *315*(7109), 629–634.
- 12 Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical
13 representation of the fingers of the left hand in string players. *Science*, *270*, 305-307.
- 14 Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a
15 target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*(1), 143-149.
- 16 Fan, J., McCandliss, B., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and
17 independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340-347.
- 18 Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control
19 functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*,
20 *133*(1), 101–135. <http://doi.org/10.1037/0096-3445.133.1.101>
- 21 Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006).
22 Not all executive functions are related to intelligence. *Psychological Science (Wiley-*
23 *Blackwell)*, *17*(2), 172–179. <http://doi.org/10.1111/j.1467-9280.2006.01681.x>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodríguez-Pujadas, A., Belloch, V., ... &
2 Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism
3 on cognitive control. *Neuroimage*, *53*(4), 1272-1278.
- 4 García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016).
5 The neuroanatomy of bilingualism: how to turn a hazy view into the full picture. *Language*,
6 *Cognition and Neuroscience*, *31*(3), 303–327.
7 <http://doi.org/10.1080/23273798.2015.1068944>
- 8 Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means
9 a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of*
10 *Memory and Language*, *58*(3), 787-814.
- 11 Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of
12 shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental*
13 *Psychology*, *38*(4), 404.
- 14 Green, D., W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism:*
15 *Language and Cogntion*, *1*, 67-81.
- 16 Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual
17 speakers. *Frontiers in Psychology*, *2*, 103.
- 18 Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control
19 hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530.
20 <http://doi.org/10.1080/20445911.2013.796377>
- 21 Grundy, J. G., & Timmer, K. (2016). Bilingualism and working memory capacity: A
22 comprehensive meta-analysis. *Second Language Research*.
23 <http://doi.org/10.1177/0267658316678286>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Gutierrez, M. (2009) *A study of possible pre-cognitive advantages of bilingualism*. (Doctoral
2 dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No.
3 1473867)
- 4 Hakuta, J. (1986). *Mirror of Language: The debate of bilingualism*. New York, NY: Basic
5 Books.
- 6 Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision
7 Research, 18(10)*, 1279-1296.
- 8 Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching
9 advantages: A diffusion-model analysis of the effects of interactional context on switch costs.
10 *Cognition, 150*, 10–19. <http://doi.org/10.1016/j.cognition.2016.01.016>
- 11 Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic
12 interference tasks? Implications for the plasticity of executive control processes.
13 *Psychonomic Bulletin & Review, 18(4)*, 625–658. <http://doi.org/10.3758/s13423-011-0116-7>
- 14 Hilchey, M. O., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance
15 cognitive fitness in traditional non-linguistic executive processing tasks? In: J. H. Schwieter
16 (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 586–613). Cambridge
17 University Press.
- 18 Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016) Effects of short-term music and second-
19 language training on executive control. *Journal of Experimental Child Psychology, 144*, 84-
20 97. doi: 10.1016/j.jecp.2015.11.009
- 21 Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our
22 current understanding. *Neuropsychology Review, 17*, 213–233.
- 23 Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017).
24 Relationship between language switching experience and executive functions in bilinguals:

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 An Internet-based study. *Journal of Cognitive Psychology*, 29(4), 1–16.
2 <http://doi.org/10.1080/20445911.2017.1282489>
- 3 Kane, M.J., Hambrick, D.Z., & Conway, A.R. (2005). Working memory capacity and fluid intelligence
4 are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological*
5 *Bulletin*, 131, 66-71.
- 6 Klein, R. M. (2016) What cognitive processes are likely to be exercised by bilingualism and does
7 this exercise lead to extra-linguistic cognitive benefits? *Linguistic Approaches to*
8 *Bilingualism*, 6(5), 549-564.
- 9 Kouzaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: absence of a "bilingual advantage" in
10 stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology*,
11 65, doi: 10.1080/17470218.2011.604788
- 12 Kovelman, I. (2006). Bilingual and monolingual brains compared: A fmri study of semantic
13 processing (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database.
14 (UMI No. 3219722)
- 15 Kroll, J., & Gollan, T. H. (2014). Speech planning in two languages: What bilinguals tell us about
16 language production. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford Handbook of*
17 *Language Production*. doi: 10.1093/oxfordhb/9780199735471.013.001
- 18 Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M.
19 (2012). Differences in word recognition between early bilinguals and monolinguals:
20 Behavioral and ERP evidence. *Neuropsychologia*, 50(7), 1362–1371.
21 <http://doi.org/10.1016/j.neuropsychologia.2012.02.021>
- 22 Lehtonen, M. & Laine, M. (2003). How word frequency affects morphological processing in mono- and
23 bilinguals. *Bilingualism: Language and Cognition*, 6, 213 – 225.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language
2 learning: Anatomical changes in the human brain. *Cortex*, 58, 301–324.
3 <http://doi.org/10.1016/j.cortex.2014.05.001>
- 4 Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism
5 and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14(4), 588-
6 595. <https://doi.org/10.1017/S1366728911000010>
- 7 Luo, L., Craik, F. I., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a
8 working memory task: Evidence from aging. *Psychology and Aging*, 28(1), 28.
- 9 Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on
10 verbal fluency performance in bilinguals. *Cognition*, 114, 29-41.
11 [doi:10.1016/j.cognition.2009.08.014](https://doi.org/10.1016/j.cognition.2009.08.014)
- 12 Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., &
13 Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers.
14 *Proceedings of the National Academy of Sciences*, 97(8), 4398-4403.
- 15 Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing:
16 Within- and between-language competition. *Bilingualism: Language and Cognition*, 6(2),
17 S1366728903001068. <http://doi.org/10.1017/S1366728903001068>
- 18 Meuter, R. F. I., & Ehrich, J. F. (2012). The acquisition of an artificial logographic script and
19 bilingual working memory: Evidence for L1-specific orthographic processing skills transfer
20 in Chinese–English bilinguals. *Writing Systems Research*, 4, 8–29.
21 <http://doi.org/10.1080/17586801.2012.665011>
- 22 Miyake, A., & Friedman, N. P. (2012). The Nature and Organisation of Individual Differences in
23 Executive Functions : Four General Conclusions. *Current Directions in Psychological
24 Science*, 21(1), 8–14. <http://doi.org/10.1177/0963721411429458>.The

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D.
2 (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex
3 “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100.
4 <http://doi.org/10.1006/cogp.1999.0734>
- 5 Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper,
6 N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a
7 comprehensive simulation study. *BMC Medical Research Methodology*, *9*(2).
- 8 Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage.
9 *Developmental Science*, *10*(6), 719-726. <http://doi.org/10.1111/j.1467-7687.2007.00623.x>
- 10 Mukadam, N., Sommerlad, A., & Livingston, G. (2017). The relationship of bilingualism compared to
11 monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-
12 analysis. *Journal of Alzheimer's Disease*, *58*, 45-54. doi: 10.3233/JAD-170131.
- 13 Münte, T., Altenmüller, E. & Jäncke, L. (2002). The musician's brain as a model of
14 neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473-478.
- 15 Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2013).
16 Meta-analytic evidence for a superordinate cognitive control network subserving diverse
17 executive functions. *Cogn Affect Behav Neurosci*. 2012, *12*(2): 241–268.
18 doi:10.3758/s13415-011-0083-5.
- 19 Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.M. (2005). Working memory and intelligence--their
20 correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychological*
21 *Bulletin*, *131*, 61-65. doi: 10.1037/0033-2909.131.1.61
- 22 Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability
23 and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of*
24 *Cognitive Psychology*, *26*(3), 242–255. <http://doi.org/10.1080/20445911.2014.891597>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage
2 in executive processing. *Cognitive Psychology*, *66*(2), 232–258.
3 <http://doi.org/10.1016/j.cogpsych.2012.12.002>
- 4 Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon
5 specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6),
6 615–639. <http://doi.org/10.1080/20445911.2014.944914>
- 7 Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning
8 either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*,
9 *69*, 265–278. <http://doi.org/10.1016/j.cortex.2015.04.014>
- 10 Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group
11 differences in executive functioning. *Journal of Neuroscience Methods*, *1*, 81–93. doi:
12 [10.1016/j.jneumeth.2016.10.002](http://doi.org/10.1016/j.jneumeth.2016.10.002)
- 13 Pelham, S. D. (2014). *Monolinguals' and bilinguals' attentional control in the presence of cognitive*
14 *and emotional distraction* (Doctoral dissertation). University of Florida, FL. Retrieved from
15 ProQuest Dissertations and Theses Database.
- 16 Peters, J. L., Sutton, A. J., Jobes, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced
17 meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry.
18 *Journal of Clinical Epidemiology*, *31*(10), 991–996.
- 19 Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of
20 the trim and fill method in the presence of publication bias and between-study heterogeneity.
21 *Statistics in Medicine*, *26*(25), 4544–4562.
- 22 Prior, A., & Gollan, T. (2011). Good language-switchers are good task-switchers: Evidence from
23 Spanish-English and Mandarin-English bilinguals. *Journal of the International*
24 *Neuropsychological Society*, *17*, 682–691. doi:[10.1017/S1355617711000580](http://doi.org/10.1017/S1355617711000580)

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism:*
2 *Language and Cognition*, 13(2), 253-262.
- 3 R Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R
4 Foundation for Statistical Computing.
- 5 Ramos, S., Fernández García, Y., Antón, E., Casaponsa, A., & Duñabeitia, J. A. (2017). Does
6 learning a language in the elderly enhance switching ability? *Journal of Neurolinguistics*,
7 43, 39–48. <http://doi.org/10.1016/j.jneuroling.2016.09.001>
- 8 Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage.
9 *Perceptual and Motor Skills*, 8(3), 271-276.
- 10 Rey-Mermet, A., Gade, M., & Oberauer, K. (2017, September 28). Should We Stop Thinking About
11 Inhibition? Searching for Individual and Age Differences in Inhibition Ability. *Journal of*
12 *Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.
13 <http://dx.doi.org/10.1037/xlm0000450>
- 14 Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997). 'Oops!': performance
15 correlates of everyday attentional failures in traumatic brain injured and normal subjects.
16 *Neuropsychologia*, 35, 747-758.
- 17 Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal
18 human attention: The test of everyday attention. *Journal of the International*
19 *Neuropsychological Society : JINS*, 2(6), 525–534.
20 <http://doi.org/10.1017/S1355617700001697>
- 21 Robertson, I. H., Ward, T., Ridgeway, V., Nimmo-Smith, I., & McAnespie, A. W. (1994). The
22 test of everyday attention (TEA). Bury St. Edmonds, United Kingdom: Thames Valley Test
23 Company.

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive
2 tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.
3 <http://doi.org/10.1037/0096-3445.124.2.207>
- 4 Roth, D. (2003). *Effect of language status on neuropsychological test performance in elderly nursing*
5 *home residents*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses
6 Database. (UMI No. 3098138)
- 7 Samuel, S. (2015). *An investigation of a proposed bilingual advantage in aspects of executive function:*
8 *Evidence from visual perspective taking and Simon tasks* (Doctoral dissertation). University of
9 Essex, UK. Retrieved from NDLTD Theses Database.
- 10 Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig,
11 D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *PLoS One*, *12*(4),
12 e0176151. <http://doi.org/10.1371/journal.pone.0176151>
- 13 Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure?
14 Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*(JUL), 1–
15 10. <http://doi.org/10.3389/fpsyg.2014.00772>
- 16 Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue
17 on information processing. *Journal of Applied Psychology*, *51*(3), 300
- 18 Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size. Correcting for
19 publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6),
20 666–681.
- 21 Soveri, A., Antfolk, J., Karlsson, L., Salo, B. & Laine, M. (2017). Working memory training
22 revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin &*
23 *Review*, doi: 10.3758/s13423-016-1217-0
- 24 Soveri, A., Laine, M., Hämäläinen, H. & Hugdahl, K. (2011b). Bilingual advantage in attentional

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 control: Evidence from the forced attention dichotic listening paradigm. *Bilingualism:*
2 *Language and Cognition*, 14(3), 371-378.
- 3 Soveri, A., Lehtonen, M., Karlsson, L.C., Lukasik, K., Antfolk, J., & Laine, M. (2016). Test–retest
4 reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology:*
5 *Adult*, doi: 10.1080/23279095.2016.1263795
- 6 Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011a). Is there a relationship between
7 language switching and executive functions in bilingualism? Introducing a within-group
8 analysis approach. *Frontiers in Psychology*, 2, 1–8. <http://doi.org/10.3389/fpsyg.2011.00183>
- 9 Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014).
10 Behavioral components of impulsivity. *Journal of Experimental Psychology: General*,
11 143(2), 850-886. doi: 10.1037/a0033981
- 12 Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce
13 publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.
- 14 Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*
15 *Psychology*, 18(6), 643.
- 16 Sullivan, M. D., Janus, M., Moreno, S., Astheimer, L., & Bialystok, E. (2014). Early stage
17 second-language learning improves executive control: Evidence from ERP. *Brain and*
18 *Language*, 139, 84–98. <http://doi.org/10.1016/j.bandl.2014.10.004>
- 19 Tabares, J.G. (2012). *Phonological influences in verbal working memory in monolinguals and*
20 *bilinguals* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses
21 Database. (UMI No. 3494340)
- 22 Tran, C. D., Arredondo, M. M., & Yoshida, H. (2015). Differential effects of bilingualism and
23 culture on early attention: a longitudinal study in the U.S., Argentina, and Vietnam.
24 *Frontiers in psychology*, 6. <http://doi.org/10.3389/fpsyg.2015.00795>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Wagenmakers, E.-J. (2015). A quartet of interactions. *Cortex*, 73, 334-335.
2 <https://doi.org/10.1016/j.cortex.2015.07.031>
- 3 Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, 18(1),
4 (3-24). [http:// doi:10.1017/S1366728914000522](http://doi:10.1017/S1366728914000522)
- 5 Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013).
6 Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2),
7 576–594. <http://doi.org/10.3758/s13428-012-0261-6>
- 8 Waris, O., Soveri, A., Ahti, M., Hoffing, R. C., Ventus, D., Jaeggi, S. M., ... Laine, M. (2017).
9 A latent factor analysis of working memory measures using large-scale data. *Frontiers in*
10 *Psychology*, 8(JUN), 1–14. <http://doi.org/10.3389/fpsyg.2017.01062>
- 11 Verreyt, N., Woumans, E., Vandelanotte, D. & Szmalec, A. (2016). The influence of language-
12 switching experience on the bilingual executive control advantage. *Bilingualism: Language*
13 *and Cognition*, 19(1), 181-190. <https://doi.org/10.1017/S1366728914000352>
- 14 Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of*
15 *Statistical Software*, 36(3), 1–48.
- 16 Wu, Y. J., & Thierry, G. (2010). Chinese-English Bilinguals Reading English Hear Chinese.
17 *Journal of Neuroscience*, 30(22), 7646–7651. [http://doi.org/10.1523/JNEUROSCI.1602-](http://doi.org/10.1523/JNEUROSCI.1602-10.2010)
18 [10.2010](http://doi.org/10.1523/JNEUROSCI.1602-10.2010)
- 19 Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in
20 linguistically and culturally homogeneous children and adults. *Journal of Experimental*
21 *Child Psychology*, 146, 121–136. <http://doi.org/10.1016/j.jecp.2016.01.011>
- 22 Yang, S., Yang, H. & Lust, B. (2011). Early childhood bilingualism leads to advances in
23 executive attention: Dissociating culture and language. *Bilingualism: Language and*
24 *Cognition*, 14(3), 412-422. <https://doi.org/10.1017/S1366728910000611>

BILINGUALISM AND EXECUTIVE FUNCTIONS

- 1 Zhang, H., Kang, C., Wu, Y., Ma, F., & Guo, T. (2015). Improving proactive control with
2 training on language switching in bilinguals. *Neuroreport*, 26(6), 354–9.
3 <http://doi.org/10.1097/WNR.0000000000000353>
- 4 Zhou, B., & Krott, A. (2016). Data trimming procedure can eliminate bilingual cognitive
5 advantage. *Psychonomic Bulletin & Review*, 23(4), 1221–1230.
6 <http://doi.org/10.3758/s13423-015-0981-6>