Article:

Lehtonen, M., Soveri, A., Laine, A. et al. (3 more authors) (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. Psychological Bulletin. pp. 394-425. ISSN 0033-2909

https://doi.org/10.1037/bul0000142

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

Minna Lehtonen[1,2], Anna Soveri[1,3], Aini Laine[1], Janica Järvenpää[1], Angela de Bruin[4] & Jan Antfolk[1,3]

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

[2]Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

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

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

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

Abstract

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 We also included unpublished data, and considered the potential influence of a number of study-, task-, and participant-related variables.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

Keywords: Bilingualism, executive functions, cognitive control, bilingual advantage, meta-analysis

Significance of the meta-analysis

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

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

Theoretically, the bilingual advantage is assumed to stem from the demands that the use of two languages puts on the cognitive control system. Bilinguals' both languages have been shown to be active even when only one of them is used for communication (e.g., Marian & Spivey, 2003; Wu & Thierry, 2010). Producing a word in one language also activates the word in the other language, eliciting competition between the lexical alternatives. This means that cognitive control functions must work effectively to enable fluent use of the appropriate language and to prevent interference from the other language. Efficient use of two languages is assumed to require inhibition of items of the irrelevant language (Green, 1998) and flexible switching between languages. Further, it requires monitoring the activation levels of the two languages and the language context in order to choose the appropriate target language. This control of language use is assumed to involve domain-general EF processes, that is, control functions that are also used in other cognitive domains than language, such as monitoring behavior for conflict and inhibition of unwanted mental representations to minimize the conflict. Similarly, language switching and domain-general task switching share many common features, such as switch costs and their asymmetries (Prior & Gollan, 2011), as well as partly common neural substrates (De Baene, Duyck, Brass & Carreiras, 2015). Frequent language switching has therefore been suggested to train domain-general EF (e.g., Prior & MacWhinney, 2010; Soveri, Rodríguez-Fornells & Laine, 2011a; but see Jylkkä et al., 2017; Paap et al., 2017). Furthermore, distinct language use patterns and interactional contexts have been proposed to set differential demands on cognitive control and thus possibly lead to differential "training" gains (Green & Abutalebi, 2013). Due to often lifelong experience of cognitive control in the field of language, bilinguals are believed to have received more practice in domain-general EF processes than monolinguals (for narrative reviews, see, e.g., Bialystok, 2017; Bialystok et al., 2012).

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

Possible Moderators of the Bilingual Advantage

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

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

Assuming that the bilingual advantage is due to the training of continuous management of different languages, longer and more intensive bilingual experience should be associated with larger gains. In line with this view, bilingual participants with an early age of acquisition (AoA) of the second language (L2) have been proposed to show larger advantages (e.g., Luk, de Sa & Bialystok, 2011). Similarly, proficiency in the languages has been suggested to modulate training gains: A strong L2 should elicit higher interference to the first language (L1) than a weak L2, and thus lead to higher cognitive control demands. Furthermore, the structural and lexical similarity of bilinguals' two languages has been suggested to influence how much they interfere with each other, with more similar languages assumedly creating more interference and therefore larger training gains. Language similarity could thus affect the intensity of training of cognitive control (Bialystok, 2017; Costa, Santesteban & Ivanova, 2006).

The Controversy

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

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

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

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

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

Previous Systematic Reviews

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

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

Using a vote-count method, Paap et al. (2015) summarized the results of all studies published after the review by Hilchey and Klein (2011) investigating differences between bilingual and monolingual participants in nonverbal inhibition and set shifting. Their summary showed a bilingual advantage only in a small proportion of the included studies (proportions ranging from .125 to .217). Furthermore, their review showed that a bilingual advantage was typically reported in studies with small samples, while null results were only reported in studies with larger samples ($n > 50$). Based on this information, Paap et al. (2015) concluded that it is unlikely that a bilingual advantage in EF exists. Similarly, in an updated review, Hilchey, Saint-Aubin, and Klein (2015) concluded that, contrary to their 2011 review (Hilchey & Klein, 2011) showing a bilingual advantage on global RTs, there is little support for this claim in more recent publications.

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

Donnelly (2016) investigated the effects of bilingualism on interference control and set shifting in healthy children and adults. The meta-analysis on interference control included 168 effect sizes from 43 studies and showed a small overall effect of bilingualism ($d = .29$). The effect was significantly moderated by which research group had conducted the original studies. There was, however, no effect of task (e.g., Flanker task and Simon task) or type of measure (i.e., global RTs vs. interference cost) and the significant moderator effects of AoA and age were interpreted to be due to publication bias. The meta-analysis on set shifting was based on 30 effect sizes from 10 studies. The results showed no effect of bilingualism on set shifting (d = -.03) and there was no effect of research group.

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

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

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

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

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

The Present Study

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

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

Furthermore, we analyze whether the stimulus material used in the tasks is verbal or nonverbal in nature (see also de Bruin et al., 2015c; Grundy & Timmer, 2016). As bilingual participants have been reported to be at a disadvantage in verbal tasks (e.g., Bialystok et al., 2008b, Bialystok, 2009; Bialystok, Barac, Blaye & Poulin-Dubois, 2010; Bialystok & Luk, 2012), larger bilingual advantages may be observed for nonverbal than verbal tasks. For verbal fluency, which includes a strong language component, smaller bilingual advantages could be observed than for other EF tasks. This would be the case especially for category fluency, in which the demands on EF may be lower than for letter fluency. Letter fluency has been suggested to be more effortful because phonemic generation is not a task one commonly performs and it does not reflect the organization of words in the mental lexicon. With category fluency, participants can use existing links and practiced strategies to activate relevant representations (see, e.g., Luo et al., 2010). We also consider the language in which the EF tasks are performed, that is, whether the testing language is bilinguals participants' L1 or L2, in order to ensure that the group comparisons in verbal EF tasks are fair (Grundy & Timmer, 2016).

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

We also consider several participant-related variables: age group, AoA of L2, language proficiency, and immigrant status. First, we study whether older adults show a larger bilingual advantage than younger adults do (e.g., Hilchey & Klein, 2011; Bialystok, 2017). Second, we test the hypothesis that bilingual participants with an early AoA of L2 show a larger advantage than late bilinguals, due to the assumedly longer amount of training received (e.g., Luk, De Sa & Bialystok, 2011). Third, we analyze the effect of proficiency level in L2, with the assumption that high-proficiency bilingual participants have faced stronger demands and more training for cognitive control than lower-proficiency bilinguals. Fourth, we test whether possible immigrant status of bilinguals, a variable often discussed but not systematically analyzed in previous reviews, moderates the effects. Our focus is on adults: There was a vast amount of studies available even with the present focus. Also, while bilingual advantages have been reported in children as well (e.g., Grundy & Timmer, 2016; Hilchey et al., 2015; but see Antón et al., 2014; Duñabeitia et al. 2014; Gathercole et al., 2014), we would expect the advantage to be better observed in adults due to an assumedly longer "training period" of EF, at least in early bilinguals (who have decades of bilingual language control experience vs. a few years in children). Moreover, the significance of the phenomenon would naturally be limited if the positive effects were only observed in children and not in adults[2].

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

[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, longitudinal studies are largely lacking (but see Bak et al., 2014, and Woumans, Surmont, Struys & Duyck, 2016).

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

Primary Research Questions of the Present Study

1. In which EF domain (if any) do we observe a bilingual advantage? What are estimates for the advantage in each cognitive domain when correcting for possible publication bias?

2. Are possible advantages specific to particular task paradigms?

3. Are possible advantages of different magnitude in verbal than nonverbal tasks? In verbal tasks, have the tasks been performed in bilinguals' L1?

4. Are observed advantages affected by how participant groups have been matched for age, SES, vocabulary knowledge, or IQ?

5. Is there a larger advantage in older than younger bilingual adults?

6. Does AoA or proficiency in L2 moderate the advantages? Is the advantage related to possible immigration background of the bilingual participants?

7. Does the country in which the study was conducted or language pair of the bilinguals moderate the effects?

Method

Literature Search

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

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

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

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

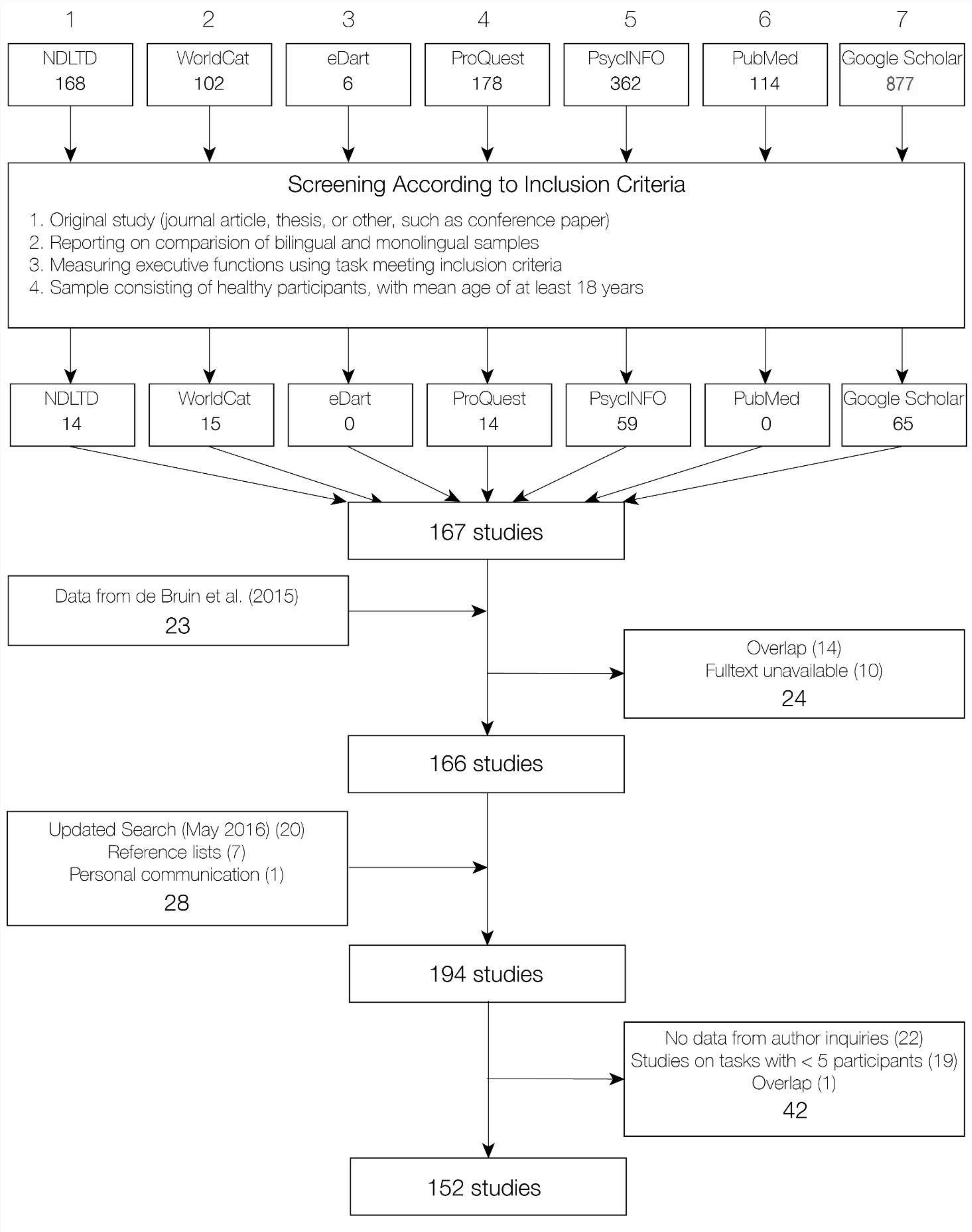

Figure 1. Flowchart of the screening process. Numbers 1-7 at the top of the figure refer to the order in which the results were screened. Other values refer to unique inclusions at each

stage. Inclusions reported to the left and exclusions reported to the right. NDLTD = Networked Digital Library of Theses and Dissertations.

Inclusion and Exclusion Criteria

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

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

Inclusion related to participants. We only included samples of healthy adult participants (mean age at least 18 years). Clinical groups, such as those including participants with deafness or neurological illnesses such as dementia, were excluded. We relied on the original studies' grouping of participants to bilinguals and monolinguals even though there was large variation in the operational definitions of bilingualism. However, we excluded studies that explicitly reported having blended bilingual and monolingual participants in a sample (e.g., Deslauries, 2008; Roth, 2003). Despite the variability in the definitions, the majority of studies included monolinguals with limited experience of a second language, and in all studies, monolingual participants had markedly less L2 experience than the bilinguals. Because we implemented rather liberal initial inclusion criteria for the participant groups, we also coded AoA and language proficiency in L2 and used this information to narrow down the groups in further analyses (see Participant characteristics under section Data Coding).

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

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

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

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

Data Coding

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

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

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

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

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

Matching of groups. We coded whether the bilingual and monolingual groups were matched for the following variables: age (mean and *SD*), income and education (as measures of SES), IQ (e.g., WAIS score), and measures of vocabulary size (expressive, receptive or both). Matching of groups for age was analyzed via calculating the effect size for the reported difference between the groups; if this effect size was between g = -0.3 and g = 0.3, the groups were considered matched[4]. For other variables, we relied on authors' own statements regarding the matching or checked the relevant statistics when reported. With regard to education, a common situation was one where both groups consisted of university students. In that case, the groups were considered matched.

[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.

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

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

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

The decision on which measures to include followed a priority order that was primarily based on the measures most typically used for these tasks in bilingual EF studies. In general, we prioritized measures that controlled for baseline performance such as cost effects in inhibitory control tasks or set shifting tasks. In these measures, we preferred a neutral baseline (if available) rather than a congruent one which can show facilitatory effects (e.g., Coderre, van Heuven & Conklin, 2013). The decision of inclusion was a tradeoff between reducing noise in data (using as "clean" and homogenous measures as possible) while including as much data as possible. In case the preferred domain-specific task measure (e.g., the Simon effect) was not available, we included the most difficult task condition (e.g., the incongruent condition). If the most difficult condition was also not available, we excluded the measure in question. In the following, we will outline the chosen measures, grouped by cognitive domain (see also Table 1).

Inhibitory control. Inhibitory control refers to the ability to deliberately inhibit dominant responses (Miyake et al., 2000) or competing representations (Stahl et al., 2014). Task paradigms with inhibitory control measures included Simon[5] (Simon & Rudell, 1967), (ANT) Flanker (Eriksen & Eriksen, 1974; Fan, McCandliss, Sommer, Raz, & Posner, 2002), Stroop (Stroop, 1935), Go-NoGo, and the Antisaccade task (Hallett, 1978). For interference control tasks (Flanker, Simon, Stroop), we used 1) the interference effect (incongruent vs. neutral trials; see, e.g., Bialystok, Craik & Luk, 2008a; Coderre et al., 2013); When that was not available, the following measures were used in the following order of preference 2) conflict effect (incongruent vs. congruent trials); or 3) incongruent condition. However, for Go-NoGo tasks, we only included accuracy measures (i.e., failure to inhibit responses), as

[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).

Go-trial RTs may also measure monitoring (Donkers & Van Boxtel, 2004). For antisaccade tasks, we included the antisaccade interference effect, antisaccade trial RTs, or antisaccade error rates in that preference order.

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

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

Monitoring. Monitoring refers to the ability to monitor conflict in information processing and to evaluate the need for cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001).

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

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

Attention. In the present meta-analysis, the Attention domain refers to the ability to selectively direct and maintain attention to stimuli. Task paradigms with attention measures included Sustained Attention to Response Task (SART; Robertson, Manly, Andrade,

[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 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.

Baddeley & Yiend, 1997), TEA, and Flanker. For SART, we included error rates. For TEA, we used “elevator counting” and “elevator counting with distraction” associated with sustained and selective attention, respectively. For ANT, we included orienting and alerting measures, and if not available, center cue condition for orienting and no cue condition for alerting.

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

Interrater reliability. The studies were coded by two raters with earlier experience in meta-analyses. Interrater reliability was addressed via the following process: First, both raters coded the same ten studies and checked that their coding was uniform. Disagreements were resolved through discussion. Then both raters independently coded approximately half of the remaining studies each. In addition, we randomly selected twenty[7] studies from the whole set, which both raters coded close to the end of the process. For these studies, the interrater reliability was calculated. The interrater reliability (Cohen’s Kappa) for the different variables ranged from strong $\kappa = .834$, $p < .001$ to perfect agreement $\kappa = 1.000$, $p < 0.001$.

[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.

Table 1

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

Domain (*k*)[1]	Task Paradigm (*k*)	Task Version	Measure (type, *k*)[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;	
		Linguistic Flanker	
	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)

Domain (*k*)[1]	Task Paradigm (*k*)	Task Version	Measure (type, *k*)[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)

Domain (*k*)[1]	Task Paradigm (*k*)	Task Version	Measure (type, *k*)[2]
Monitoring (188)	Flanker (46)	ANT; Flanker Task; Go-No/Go Flanker; LANT; Linguistic Flanker	Global RT (RT, 28); Congruent Trials (RT, 18)
	Go-No/Go (6)	Go-No/Go Flanker	Global RT (RT, 2); Congruent trials (RT, 4)
	Simon (46)	Auditory Simon; Simon 2-colors; Simon Letters	Global RT (RT, 21); Congruent Trials (RT, 25)
	Stroop (44)	Auditory Stroop; Color-Word Stroop; Numerical Stroop; Spatial Stroop	Global RT (RT, 21; Acc, 1); Congruent Trials (RT, 18; Acc, 4)

Domain (*k*)[1]	Task Paradigm (*k*)	Task Version	Measure (type, *k*)[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-back effect (2-back minus 1-back; Acc)

Domain (k)[1]	Task Paradigm (k)	Task Version	Measure (type, k)[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; LANT	Alerting Effect (RT, 11); No Cue Trials (RT, 5)
	ANT Orienting (20)	ANT; LANT	Orienting Effect (RT, 14); 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

Note. SART = The Sustained Attention to Response Task; TEA = Test of Everyday Attention; TMT = Trail-Making Test; WCST = Wisconsin Card Sorting Test; LNS = Letter-Number Sequencing Task; ANT = Attention Network Task; LANT = Lateralized Attention Network Task; FW = forward; BW = backward.

[1] *k* refers to the number of effect sizes used in our final analyses (i.e., after pooling and before outlier exclusion).

[2] Measures are presented in order of preference. Acc = accuracy; RT = reaction time; O = other.

Statistical Analyses

For statistical analyses, we used metafor (Viechtbauer, 2010) for R (version 3.2.3; R Core Team, 2015). The R script including all reported analyses, an output of all the analyses, and the data file used in the analyses are available at (Lehtonen et al., 2018).

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

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

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

Multi-level modelling. In our data, effect sizes could not be considered entirely independent. Compared to independent effect sizes, dependent effect sizes are not as informative. When effect sizes are correlated, the information obtained from one estimate overlaps with information obtained from another estimate. Unless this overlap is taken into consideration, the amount of information is overestimated, and standard errors and confidence intervals are underestimated, leading to a high number of Type I errors (e.g., Becker, 2000). To consider the dependency between effect sizes, we used a multi-level meta-analysis (e.g., Van den Noortgate, López-López, Marín-Martínez, & Sánches-Meca, 2013) considering dependency of the following forms:

First, a unique *pair* of groups (a bilingual group vs. a monolingual group) could be repeatedly compared on more than one outcome measure (e.g., Simon and Flanker effects). To consider this form of dependency, we coded for repeated comparisons within pairs. Second, groups could also be repeatedly used in more than one pair. For example, two or more monolingual groups could be compared to one bilingual group, or vice versa.[8] To take this latter form of dependency into consideration, we also coded for *clusters* of pairs within which either a monolingual or a bilingual group was repeated. In this case, we made the reasonable assumption that there would be no systematic difference depending on the group type (monolingual or bilingual) that was repeated within the pairs.

These two forms of dependency were accounted for in a model with four levels of variance. The variance within effect sizes, which is accounted for in a fixed effects meta-analysis model, constitutes the first level. The second level is the variance between outcome measures, which is accounted for in a random effects model. The third level is the variance 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.

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

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

Table 2

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	96.44***	0.07	0.05	
4. Four	Clusters	950.98	-471.49	3 vs. 4	21.05***	0.05	0.02	0.05

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

The magnitude of the dependency of outcome measures within comparisons can be estimated with the intraclass correlation coefficient (*ICC*). The *ICC* is calculated by dividing the variance between comparisons by the sum of the variance between and within comparisons (i.e., σ^2_1 / [σ^2_1 + σ^2_2]). Hence, the *ICC* value also considers variance in the effect sizes that is attributed to differences between comparisons. When the variance within comparisons is small in relation to the variance between comparisons, the *ICC* value is high. If outcome measures within comparisons vary greatly so that each measure could equally well belong to any one of the included comparisons, the correlation will drop towards zero. In our final four-level model, the *ICC* for outcome measures within pairs was .129 and the *ICC* for pairs within clusters was .464.

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

In the precision-effect test (PET), the effect sizes are first regressed on their standard errors in a weighted least-squares regression. If there is a significant and positive association between effect sizes and their standard errors, this indicates a bias where studies with low precision tend to report larger effect sizes (or, equivalently, that studies with low precision and small effect sizes are underreported). The intercept (variance = 0) of the weighted least-squares regression is taken as an estimate of an unbiased effect size in a hypothetical study with perfect statistical power. In a simulation study (Stanley & Doucouliagos, 2014), the PET method performed well when the true, unbiased effect was zero. When the true, unbiased effect differed from zero, a better performance was observed when the standard error was replaced with the variance. This test is called precision-effect test with standard error (PEESE). The authors suggested that a PET test that reveals a significant association between the effect sizes and their SE is followed up by a PEESE test.

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

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

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

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

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

Results

Descriptive Results

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

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.

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

Assessment of Bias and Data Screening

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

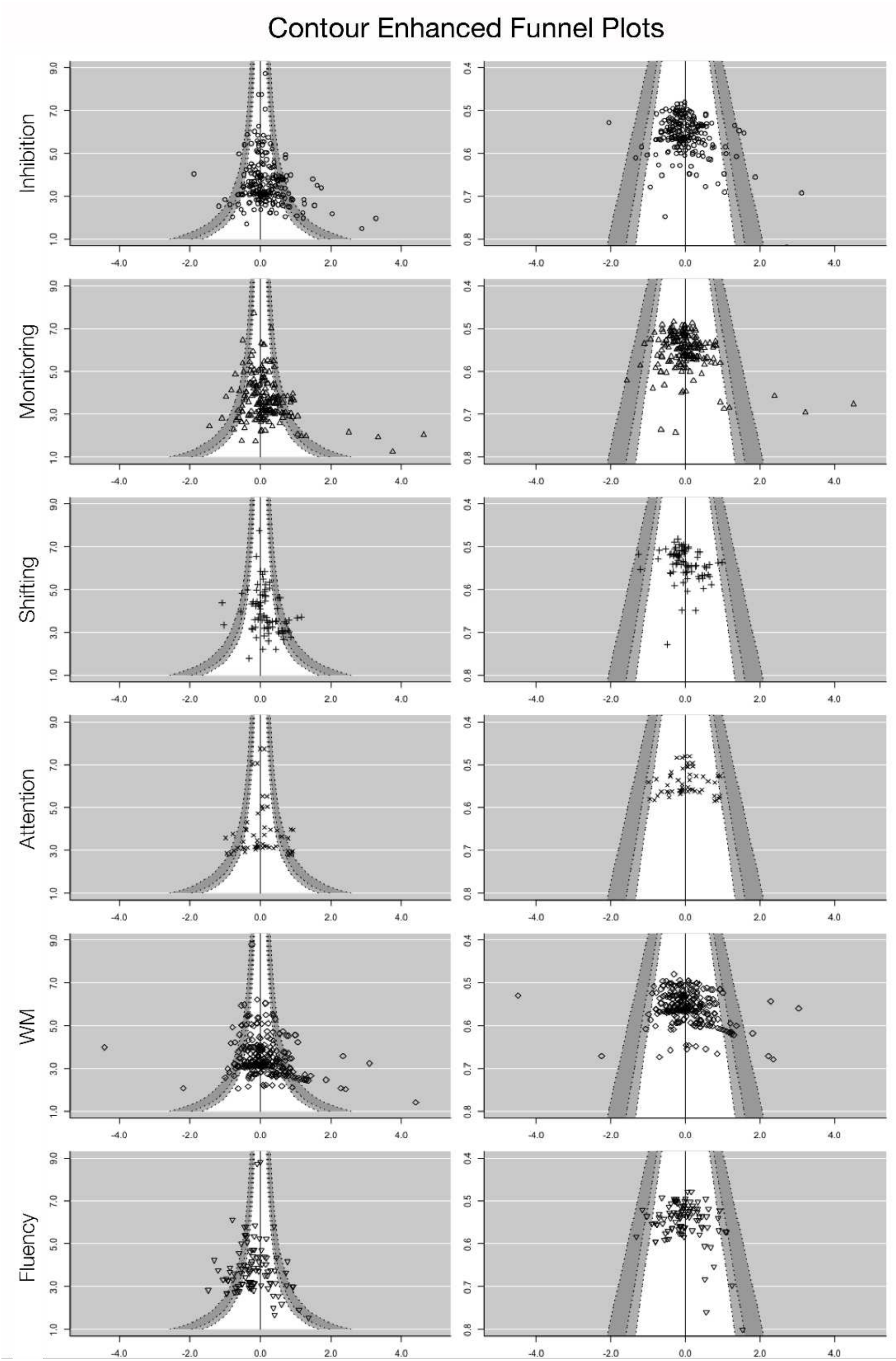

Figure 2. Contour enhanced funnel plots for each cognitive domain (by row). Contours change shades at *p*-levels .1 (white), .05 (light grey), and .01 (dark grey). In the left column, effect sizes are plotted against their precision (1/*SE*) and the reference line is set at Hedge's *g* = 0. In the right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

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

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

Bilingual Advantage

After trimming the data, we first investigated the bilingual advantage across all included EF domains. We found a very small positive effect size in favor of bilingual groups, $g = 0.06$ [0.01, 0.10], $p < .05$, Q_E [868] = 2139.79. Because of the asymmetry of effect sizes observed in the contour-enhanced funnel plots, we used the PET-PEESE method to obtain a corrected, unbiased effect size. Both the PET analysis and the PEESE analysis showed significant negative associations between the effect sizes and their *SE* and variance ($p < .01$ and $p < .01$, respectively). The PET-PEESE corrected effect size was negative, $g = -0.08$ [-0.17, 0.01], $p = .099$, but not statistically significant. We then investigated whether the obtained results would be different if they were based on analyses conducted without trimming the data. With outliers included, the estimated effect size was $g = 0.08$ [0.03, 0.14], $p < .01$, Q_E [890] = 3173.75. A PET-PEESE correction yielded a statistically significant negative effect size, -0.29 [-0.38, -0.19], $p < .001$. This suggests that trimming data led to a less biased distribution of effect sizes and likely more reliable corrected and uncorrected estimates (See Figure 3). Because of the remaining bias, we decided to perform PET-PEESE analyses for estimates above $g = 0.2$ in the subsequent analyses.

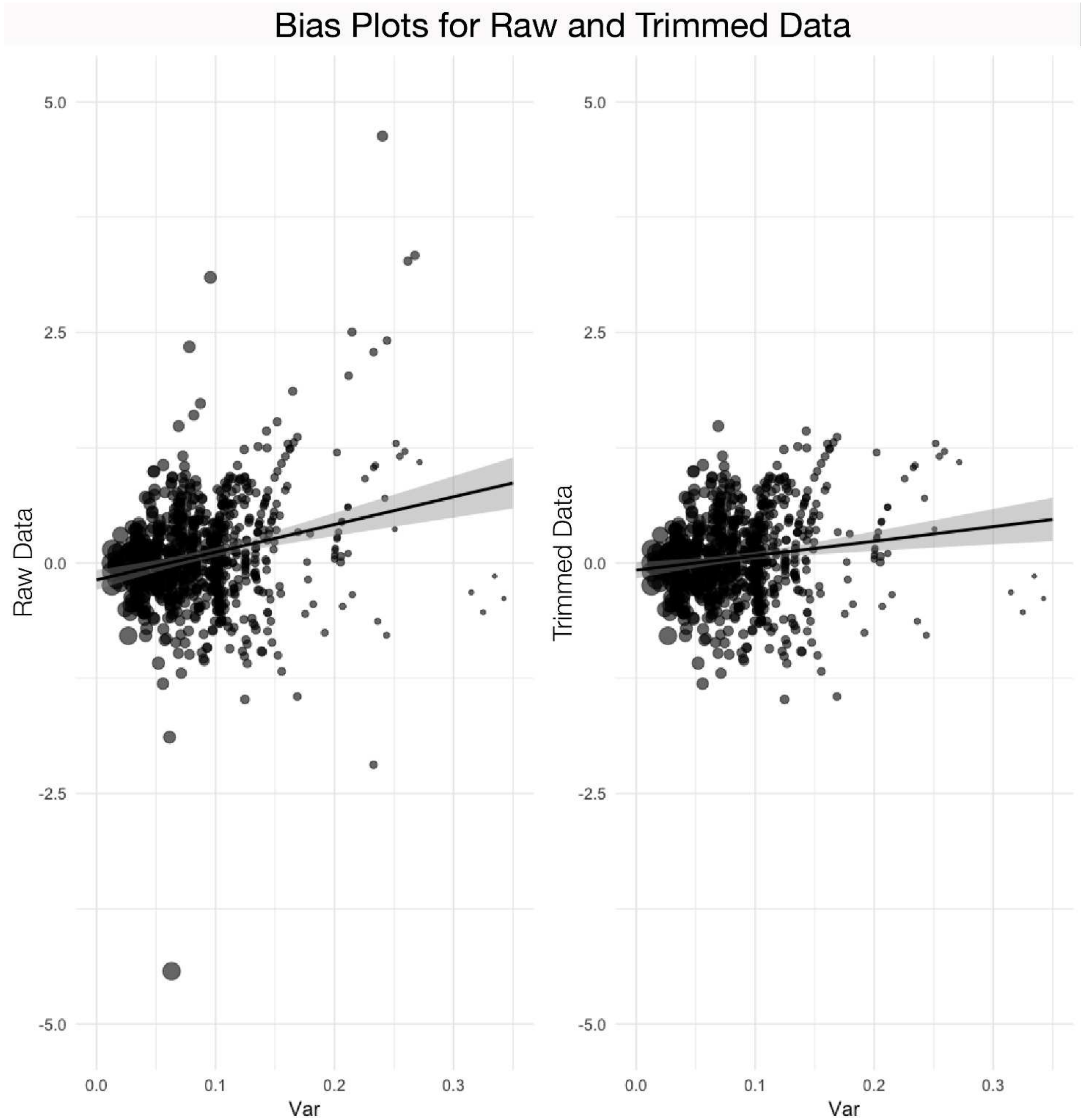

Figure 3. Plots visualizing the PEESE-corrected effect size. The line depicts the regression

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

Bilingual Advantage by Cognitive Domain

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

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

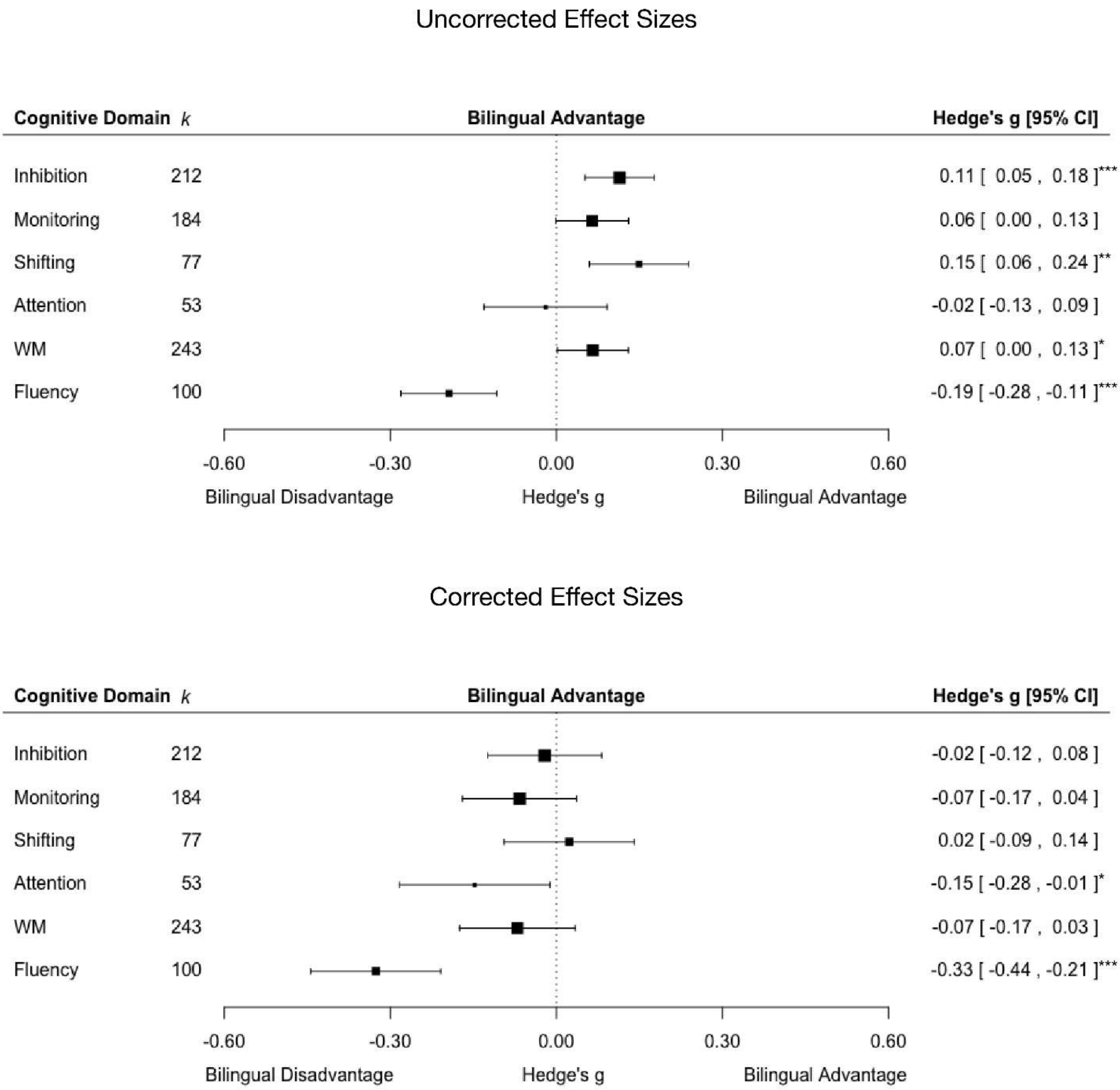

Figure 4. For each cognitive domain, the figure displays synthesized effect sizes and 95% confidence intervals (CI) for the comparison between monolinguals and bilinguals. Positive values indicate a bilingual advantage and negative values indicate a bilingual disadvantage. k = number of effect sizes. Uncorrected effect sizes are displayed in the upper panel and corrected effect sizes are displayed in the lower panel. * $p < .05$; ** $p < .01$; *** $p < .001$

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

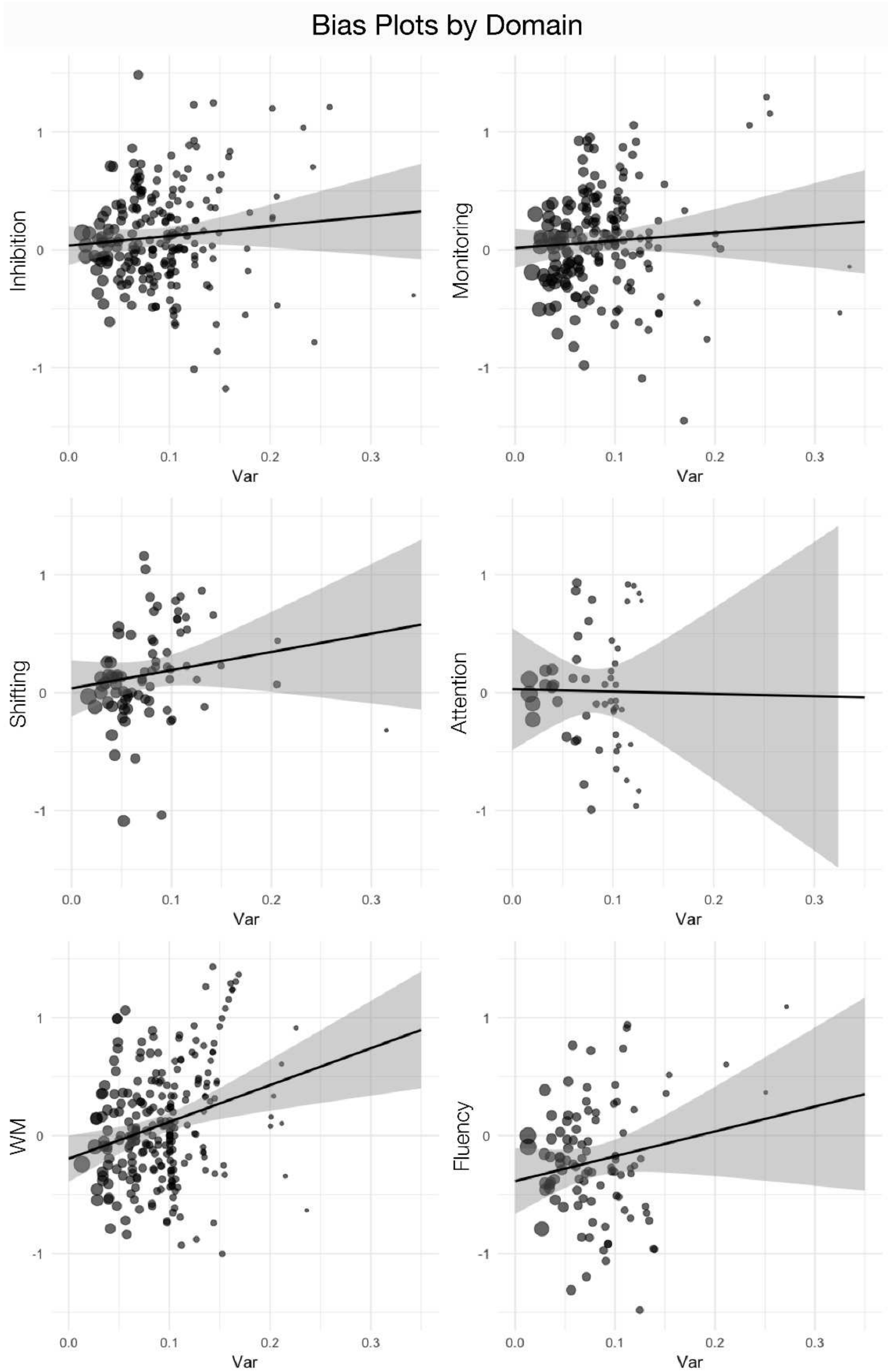

Figure 5. Scatter plots visualizing the PEESE-corrected effect sizes by each cognitive domain. The line depicts the regression slope for the association between the variance (Var; x-axis) and the effect size (y-axis). The shaded are gives the 95% confidence intervals. All panels include data after outlier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

Peer-Review Status

To investigate a possible source of the observed bias, we investigated whether outcomes differed depending on whether the report had been peer-reviewed or not. To do this, we conducted a set of analyses with peer-review status as a moderator. Peer-review status did not moderate the outcome when including all domains, Q_M [1] = 0.31, $p = .580$. We then analyzed whether outcomes differed depending on peer-review status for each of the included domains. The estimated effect sizes were not statistically significant for inhibition, Q_M [1] = 1.26, $p = .261$, monitoring, Q_M [1] = 0.12, $p = .732$, shifting, Q_M [1] = 2.51, $p = .113$, attention, Q_M [1] = 0.63, $p = .426$, or verbal fluency, Q_M [1] = 0.88, $p = .349$. For WM, the difference was statistically significant, Q_M [1] = 5.29, $p < .05$, such that the effect size was smaller among peer-reviewed reports, $g = 0.02$, [-0.08, 0.11], $p = .745$, compared with reports that had not been peer-reviewed, $g = 0.20$, [0.06, 0.33], $p < .01$.

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

Task Paradigms within Cognitive Domains

We then explored whether the outcomes within each domain were moderated by the task used to measure EF. We found that task significantly moderated the estimates for shifting and attention. In both cases, a medium-sized difference in favor of bilinguals was found for TEA (for measures of attentional switching and selective attention), but not for other tasks. TEA was, however, represented only by a low number of effect sizes. In shifting, WCST also showed a statistically significant positive effect, but it did not differ significantly from the outcomes for the rest of the shifting tasks, which can be seen from the overlapping confidence intervals. Due to the low numbers of effect sizes reported for TEA and WSCT, we did not investigate the impact of bias with a PET-PEESE analysis. (See Table 3 for details).

Table 3

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

Domain	Task	Effect Size	95%CI				Moderator test		
		g	LB	UB	*p*	*k*	Q_M	*df*	*p*
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			

	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

Verbal and Nonverbal Tasks

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

We then repeated this analysis in each of the five remaining cognitive domains. For inhibition, Q_M [1] = 0.53, $p = .465$, and attention, Q_M [1] = 1.13, $p = .288$, there was no statistically significant difference between nonverbal and verbal tasks. The outcomes were moderated by whether a task was verbal or nonverbal in three domains: monitoring, Q_M [1] = 7.17, $p < .01$, shifting, Q_M [1] = 5.65, $p < .05$, and WM, Q_M [1] = 29.00, $p < .001$. For monitoring, the estimated effect size was larger in nonverbal tasks, $g = 0.11$ [0.03, 0.18], $p < .01$, compared with verbal tasks, $g = -0.06$ [-0.18, 0.05], $p = .298$. For shifting, the estimated effect size was smaller in verbal tasks, $g = -0.01$ [-0.17, 0.15], $p = .912$, compared with nonverbal tasks, $g = 0.21$ [0.10, 0.32], $p < .001$. Also, for WM, the effect size was smaller in verbal tasks, $g = -0.00$ [-0.08, 0.08], $p = .962$, compared with nonverbal tasks, $g = 0.30$ [0.18, 0.41], $p < .001$. Because previous analyses showed a strong bias in WM, we also investigated bias in nonverbal shifting and WM tasks. A PET-PEESE correcting for bias yielded a smaller, non-significant effect in both cases, $g = 0.06$ [-0.14, 0.25], $p = .585$ for shifting and $g = 0.10$ [-0.11, 0.29], $p = .376$ for WM.

Testing Language

The original studies included verbal tasks in languages that could be either the first or a second language of the bilingual sample. Because tasks performed in L2 could be expected to have an undue influence on the outcome, we also restricted our data to include only tasks performed in the L1 of the bilinguals. In this case, the overall bilingual advantage was small and not statistically significant, g = 0.07 [-0.05, 0.18], p = .276, Q_E [108] = 336.90. We then reran our analysis with cognitive domain as a moderator. Again, domain moderated the outcome, Q_M [3] = 16.81, p < .001. For inhibition, g = 0.18 [-0.01, 0.37], p = .060, k = 24, monitoring, g = 0.07 [-0.21, 0.34], p = .639, k = 10, and verbal fluency, g = -0.17 [-0.34, -0.01], p < .05, k = 34, point estimates remained similar. For WM, g = 0.30 [0.13, 0.47], p < .001, k = 41, the effect was slightly larger than before, but corrected towards null in a follow-up PET-PEESE, g = 0.03 [-0.25, 0.32], p = .824. No data were available for shifting and attention.

Matching of Groups

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

Table 4

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

	Education			Intelligence			Vocabulary			Age		
	g	95% CI	*k*	*g*	95% CI	*k*	*g*	95% CI	*k*	*g*	95% CI	*k*
Overall	0.05	[-0.00, 0.10]	678	0.08*	[0.01, 0.14]	349	0.10**	[0.03, 0.17]	218	0.05	[-0.00, 0.11]	395
Inhibition	0.12**	[0.04, 0.19]	146	0.12**	[0.03, 0.22]	81	0.14**	[0.02, 0.26]	51	0.10*	[0.01, 0.18]	89
Monitoring	0.06	[-0.02, 0.13]	138	0.05	[-0.04, 0.14]	83	0.08	[-0.06, 0.22]	36	0.09	[-0.00, 0.18]	80
Shifting	0.11*	[0.01, 0.20]	71	0.23***	[0.10, 0.37]	31	0.21*	[0.01, 0.41]	16	0.27***	[0.14, 0.39]	34
Attention	-0.03	[-0.16, 0.09]	43	-0.05	[-0.26, 0.15]	13	0.21	[-0.04, 0.47]	10	-0.01	[-0.20, 0.17]	15
WM	0.07	[-0.00, 0.14]	201	0.06	[-0.03, 0.15]	110	0.06	[-0.04, 0.17]	69	0.03	[-0.05, 0.11]	122
Fluency	-0.23***	[-0.32, -0.13]	79	-0.11	[-0.25, 0.03]	31	0.04	[-0.09, 0.18]	36	-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

Age Group

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

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

Age of Acquisition

We then investigated whether AoA of L2 moderated the outcomes. To do this, each bilingual group was coded as either early acquisition or later acquisition (cut-off at 6 years). Outcomes across all EF domains was not moderated by AoA, Q_M [1] = 0.95, $p = .331$. Follow-up analyses for each domain separately revealed that there was no significant moderation for inhibition, Q_M [1] = 0.29, p = .592, monitoring, Q_M [1] = 0.42, p = .519, shifting, Q_M [1] = 0.02, p = .879, or attention, Q_M [1] = 0.25, p = .615. For WM, AoA moderated the outcomes, Q_M [1] = 5.12, p < .05. Samples with later acquisition showed a smaller difference between monolinguals and bilinguals in WM, g = 0.02 [-0.09, 0.12], p = .735, compared to samples with early acquisition, g = 0.23 [0.07, 0.39], p < .01. A PET-PEESE analysis corrected the outcome for early acquisition towards null, 0.02 [-0.26, 0.29], p = .912. For verbal fluency, AoA also moderated the outcomes, Q_M [1] = 4.92, p < .05. In this case, samples with early acquisition showed a larger difference between monolinguals and bilinguals, g = -0.52 [-0.82, -0.23], p < .001, compared to samples with later acquisition, g = -0.15 [-0.30, -0.01], p < .05.

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

Language Proficiency

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

.095, monitoring, Q_M [1] = 0.17, p = .677, shifting, Q_M [1] = 0.00, p = .961, attention, Q_M [1] = 1.22, p = .270, WM, Q_M [1] = 0.20, p = .653, or verbal fluency, Q_M [1] = 1.19, p = .274.

Immigrant Status

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

Country

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

Language Pair

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

Table 5

Effect Size Estimates and Confidence Intervals by Cognitive Domain and Country

Country	Inhibition			Monitoring			Shifting			Attention			WM			Fluency		
	g	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*
AUS	0.15	-0.17, 0.48	11			2			0			4	0.10	-0.53, 0.73	9			0
BRA	0.07	-0.15, 0.29	17	0.23	-0.00, 0.47	16			0	-0.01	-0.28, 0.27	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	-0.40, 0.56	6	-0.31	-0.69, 0.06	10
USA	0.09	-0.00, 0.18	72	-0.04	-0.14, 0.06	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. * = p < .05. ** = p < .01. *** = p < .001.

Table 6

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

Language Pair	Inhibition			Monitoring			Shifting			Attention			WM			Fluency		
	g	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*	*g*	95%CI	*k*
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. * = $p < .05$. *** = $p < .001$.

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 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 small[10] 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.

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

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

Publication Bias

Despite including unpublished studies, we observed bias in the distribution of the reported results, as demonstrated in the funnel plots and the PET-PEESE analyses. Studies with low precision (i.e., small sample sizes) tended to show stronger positive effects than studies with high precision, whereas null or negative effects were underrepresented in studies with low precision. A set of moderator analyses revealed no major differences between results reported in peer-review publications and other studies. However, separate PET-PEESE analyses for peer-reviewed data and other data revealed a significant association between effect sizes and their precision only in the former case. In the latter case, there was no evidence of such an association. This suggests that small studies with low precision and large, positive effect sizes might be overrepresented in the peer-reviewed literature, or that comparably small studies with large, negative effect sizes are underrepresented. There are several possible reasons for this: Journals' publication processes may have favored strong positive outcomes in support of the purported bilingual advantage. The bias may also stem from the researchers' own decisions regarding whether to pursue a peer-review publication or not, or their decisions about whether or not to report all findings when intending to publish their results.

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

Due to the problems inherent in meta-analyzing biased data, we encourage pre-registration of studies investigating the bilingual advantage. This would ensure that reporting bias does not affect the outcome of future meta-analyses. Recent evidence also shows that publication trends in this field are changing, as suggested by the bibliometric analysis by Sanchez-Azanza et al. (2017), possibly leading to more balanced reporting in the future (see also de Bruin & Della Sala, 2015).

Moderator Variables

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

Task-Related Moderator Variables

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

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

Participant-Related Moderator Variables

It has been reported that bilingual advantages in EF are better observed in older than younger adults, possibly because bilingualism may beneficially affect the typical EF decline in the elderly. One could also hypothesize that the "EF training period" has been longer for older than younger bilinguals. However, our results did not support this hypothesis. We did not find evidence that larger advantages would be observed in older, relative to younger, bilingual participants compared to monolinguals in any EF domain. On the contrary, in the domain of WM, there was a very small difference in outcomes in favor of the bilinguals in the young groups which was not present in the older groups. In the explorative task analysis for age groups, in one single task paradigm (TaskSwitching), there was a larger difference in favor of bilinguals in older than younger groups, but this outcome was based only on five samples in the older adults' age group. We thus conclude that no systematic bilingual advantages were observed in older or younger adults.

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

[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.

late bilinguals the disadvantage was very small. Early bilingual individuals tend to have used the two languages more equally than late bilinguals, leading to less exposure to one particular language than is the case for monolingual individuals. This could possibly lead to a disadvantage in tasks that require access to linguistic units, such as words (see, e.g., Gollan et al., 2008; Lehtonen et al., 2012).

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

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

It has been argued that meta-analyses and systematic reviews may miss particular variables related to the environment from which the bilingual and monolingual participants have been recruited (Bak, 2016). In an attempt to take into account some of this variation in the data, we analyzed the country in which the original study had been conducted. We found no evidence that country would moderate outcomes across all EF domains. A similar analysis for each domain separately suggested that country significantly moderated the outcome in shifting and attention. In both cases, small to medium-sized differences in favor of the bilinguals were observed in studies conducted in the UK. These outcomes differed significantly from outcomes in the US for shifting, and Greece for attention. The number of effect sizes from these countries and domains was, however, small (equal to or less than 10), and hence they were not corrected for bias. These findings may be associated with the use of particular tasks in a country, such as the use of TEA in the UK (TEA was used in seven out of nine comparisons included from the UK for shifting, and in all comparisons included for attention).

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

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

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

Limitations and Future Directions

Taken together, our meta-analysis provides no systematic evidence for a general, systematic bilingual advantage in EF in adult samples. If some enhancement of cognitive control functions exists due to bilingualism, it is restricted to very specific circumstances, and its magnitude and extent are modest.

Many authors have also commented that bilingualism is not a unitary phenomenon, making it problematic to use it as a categorical variable (e.g., Bialystok, 2017). What is admittedly complicating the research area is that a multitude of factors is likely to affect individuals' cognitive abilities, and it is difficult to control for all of them in the studies of this type. In fact, an important issue contributing to the mixed results in the field has been the inherent weaknesses of the natural groups designs of the studies (Hakuta, 1986, as cited in Klein, 2016; Laine & Lehtonen, 2018). When compared to a typical cognitive training study, the setup represents a rather weak research design: In bilingualism studies that can be taken as studies on "natural training" of EF, randomization to bilingual and monolingual groups and pre-post comparisons are normally not possible, and the specific contents of the assumed EF training are also not apparent (Laine & Lehtonen, 2018).

Thus far only a few studies have introduced longitudinal intervention designs, including language learning or training that assumedly resembles aspects of bilinguals' language behaviors, such as language switching. Adult bilinguals participating in ten days of language switching training showed improved performance at post-test in a cognitive control task when compared to a passive control group (Zhang, Kang, Wu, Ma & Guo, 2015). Janus, Lee, Moreno, and Bialystok (2016), in turn, investigated effects of short-term second-language training camp on 4–6-year-old children's nonverbal abilities. They reported improvements in specific tasks involving EF, but the improvements were similar to the children participating in a music camp. Sullivan, Janus, Moreno, Astheimer, and Bialystok (2014) tested students taking either an introductory Spanish ("training group") or an introductory Psychology course ("control group") before and after the 6-month courses. Modulations were seen in the ERPs in a go-nogo task for language learners only, but no behavioral differences were observed between the groups. Bak, Long, Vega-Mendoza, and Sorace (2016) compared EF performance, as measured with TEA, in adult participants taking a one-week language course to the performance of matched active and passive monolingual control groups. In the attentional switching measure of TEA, the language learner group showed the largest improvement at posttest, significantly different from that of the passive control group. The active control group showed intermediate performance that did not significantly differ from either the language group or the passive control group. Finally, Ramos, Fernández García, Antón, Casaponsa, and Duñabeitia (2017) studied healthy monolingual seniors learning a new language for a year. Post-test performance in a nonverbal switching task was not improved from pre-test performance for this group in comparison to a matched passive control group. In sum, although clear behavioral EF improvements due to language learning or language switching training in comparison to active control groups have not been observed in these studies and although these studies have not used a fully random assignment to groups, they nevertheless demonstrate how more solid experimental designs can be implemented in this field.

Another way to circumvent the problems of the cross-sectional designs and to make progress in this area of research might be to utilize an individual differences approach and to identify potential connections between features of the individuals' bilingual experience and cognitive performance (Bialystok, 2017; Laine & Lehtonen, 2018). Such studies, using within-group correlative analyses, have already investigated how frequency of everyday language switching and being involved in different kinds of interactional contexts (see, e.g., Green & Abutalebi, 2013) is associated with EF performance (see, e.g., Hartanto & Yang, 2016; Jylkkä et al., 2017; Soveri et al., 2011a; Verreyt, Woumans, Vandelanotte & Szmalec, 2016).

The present meta-analysis only focused on healthy adults, and thus does not address the proposed EF advantages in children or the question of possible later onset of dementia symptoms in bilingual individuals. In their systematic review, Hilchey and colleagues (2015) reported that larger advantages may in fact be observed in children than in adults. A challenge in comparing and summarizing studies on children is the variety of task versions that children of different ages need. Moreover, even if studies would consistently show that a bilingual advantage in children exists, our results provide no reliable evidence for a bilingual advantage in adulthood, at least in the cross-sectional data analyzed here. Observing an advantage only in children would naturally limit the scale and significance of the putative phenomenon. With regard to risk of dementia in older bilinguals, a recent meta-analysis by Mukadam, Sommerlad and Livingston (2017) reported that prospective studies do not show compelling evidence for bilingualism protecting from cognitive decline. According to their analysis, there is more evidence for such positive effects in retrospective studies, but with these studies, the authors raise the issue of confounding variables.

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

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

Bilingualism, like several other sustained experiences such as practicing music (Münte, Altenmüller & Jäncke, 2002), has been associated with particular neurocognitive signatures and structural changes in the brain (for reviews, see, e.g., Abutalebi, 2008; Abutalebi & Green, 2016; Bialystok, 2017; García-Pentón et al., 2015; Li et al., 2014). Some published fMRI studies have shown different neural activation patterns in EF tasks or resting state connectivity for bilinguals than monolinguals, with activation differences observed particularly in the anterior cingulate, prefrontal regions, and subcortical structures. In addition, differences in ERPs have been observed in brain responses associated to EF and attention and often assumed to reflect better processing capacity in bilinguals. Notably, such effects in neural activation have often been reported in the absence of behavioral differences between groups. In such cases, it may be difficult to know whether bilingualism-related activation increases or decreases or ERP modulations truly reflect increased processing efficiency, as the results have often been interpreted (see, e.g., Bialystok, 2017; Sullivan et al., 2014; see also Paap et al., 2015).

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

These kinds of examples of experience-dependent brain plasticity are interesting in their own right, but what remains to be investigated in future research are the underlying reasons as well as the possible functional significance and behavioral correlates of these modulations. Neural measures were outside the scope of the present study. However, based on the current results, the reported neural differences between bilingual and monolingual adults are unlikely to reflect any general bilingualism-related EF advantages with direct behavioral consequences. Our meta-analysis also leaves out particular other cognitive skills that have previously been associated with superior performance in bilingual individuals. Such domains include, for example, metalinguistic abilities and divergent thinking in which Adesope and colleagues (2010) demonstrated a bilingual advantage. It is possible that future studies will accumulate evidence on such other types of cognitive advantages of bilingualism. However, even if no extra-linguistic cognitive consequences are found, the main advantage of bilingualism—the ability to communicate in different languages with its personal and social consequences—will always remain.

Conclusions

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

Acknowledgements

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, and 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 list of authors who were able to provide additional data, see Table S2).

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*, 466-478. https://doi.org/10.1016/j.actpsy.2008.03.014

*Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., … Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex, 22*(9), 2076–2086. http://doi.org/10.1093/cercor/bhr287

Abutalebi, J. & Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689-698.

*Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Rosa, P. A. Della, Parris, B. A., & Weekes, B. S. (2015). Bilingualism provides a neural reserve for aging populations. *Neuropsychologia*, *69*, 201-210.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and Meta-Analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245. http://doi.org/10.3102/0034654310368803

*Ansaldo, A. I., Ghazi-Saidi, L., & Adrover-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of Clinical and Experimental Neuropsychology, 2*(1), 1–16. http://doi.org/10.1080/13803395.2014.990359

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children, *Frontiers in Psychology*, *5*, 1–12. http://doi.org/10.3389/fpsyg.2014.00398

Baddeley, A. (2000). The episodic buffer: A new component of working memory? Trends in Cognitive Sciences, *4(*11*)*, 417–423. doi:10.1016/S1364-6613(00)01538-2

*Badzakova-Trajkov, G. (2008). *A behavioural and functional imaging investigation of Stroop task performance in late proficient bilinguals*. (Doctoral dissertation). The University of Auckland, Auckland, New Zealand. Retrieved from NDLTD Theses Database.

Bak, T. H. (2016). Cooking pasta in La Paz. *Linguistic Approaches to Bilingualism, 6(*5*)*, 699–717. http://doi.org/10.1075/lab.16002.bak

Bak, T. H., & Alladi, S. (2016). Bilingualism, dementia and the tale of many variables: Why we need to move beyond the Western World. Commentary on Lawton et al. (2015) and Fuller-Thomson (2015). *Cortex*, *74*, 315-317. doi: 10.1016/j.cortex.2015.04.025.

Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. *PLoS ONE, 11(*4*)*, 1–11. http://doi.org/10.1371/journal.pone.0153485

Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging? *Annals of Neurology*, *75*(6), 959–963. http://doi.org/10.1002/ana.24158

*Bak, T. H., Vega-Mendoza, M., & Sorace, A. (2014). Never too late? An advantage on tests of auditory attention extends to late bilinguals. *Frontiers in Psychology*, *5*, 1–6. http://doi.org/10.3389/fpsyg.2014.00485

Barkley, R. A. (2012). *Executive Functions: What they are, how they work, and why they evolved.* New York: The Guildford Press.

*Batres, K. M. (2013). *Bilingual cognitive control and perspective-monitoring in dialogue* (Doctoral dissertation). State University of New York at Stony Brook, NY. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3594648)

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & E. D. Brown (Eds.), Handbook of applied multivariate statistics and mathematical modeling (pp. 499–525). Orlando: Academic Press.

*Bennett, J. (2012). *Linguistic and cultural factors associated with phonemic fluency performance in bilingual hispanics* (Doctoral dissertation). The University of New Mexico, NM. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3553808)

Bialystok, E. (2001). *Bilingualism in Development. Language, Literacy, and Cognition*. Cambridge, UK: Cambridge University Press.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(1), 3. http://doi.org/10.1017/S1366728908003477

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233–262. http://doi.org/10.1037/bul0000099

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D., (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *11*, 485-508. doi: 10.1080/15248372.2010.516420

*Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage*, *24*(1), 40–49. http://doi.org/10.1016/j.neuroimage.2004.09.044

*Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303. http://doi.org/10.1037/0882-7974.19.2.290

*Bialystok, E., Craik, F. I. M., & Luk, G. (2008a). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538. http://doi.org/10.1016/j.jneuroling.2007.07.001

*Bialystok, E., Craik, F. I. M., & Luk, G. (2008b). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(4), 859–873. http://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240-250. http://doi.org/ 10.1016/j.tics.2012.03.001

*Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*(6), 1341–1354. http://doi.org/10.1037/0278-7393.32.6.1341

*Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance. 35*(2), 565–574. http://doi.org/10.1037/a0012735

Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. M. (2015). Publication bias and the validity of evidence : What's the connection? *Psychological Science*, *26*(6), 944–946. http://doi.org/10.1017/S01427

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism*. *15*, 397-401. doi: 10.1017/S136672891100040X

*Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*(1), 17. http://doi.org/10.1177/13670069050090010701

*Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging, 29*(3), 696–705. http://doi.org/10.1037/a0037254

Bialystok, E. & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494-500. http://doi.org/10.1016/j.cognition.2009.06.014

*Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport 26*(16), 966–971. http://doi.org/10.1097/WNR.0000000000000453

*Billig, J. D., & Scholl, A. P. (2011). The impact of bilingualism and aging on inhibitory control and working memory. *Organon, 26*(51), 39–52.

Bishop, D. V., & Thompson, P. A. (2016). Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value. *Peer J*, 18(*4*). doi: 10.7717/peerj.1715.

*Blumenfeld, H. K. (2008). *Suppression mechanisms in monolingual and bilingual language comprehension: Evidence from eye-tracking, priming, and inhibitory control tasks.* (Doctoral dissertation). Northwestern University, IL. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3303706)

*Blumenfeld, H. K., & Adams, A. M. (2014). Learning and processing of nonverbal symbolic information in bilinguals and monolinguals. *Frontiers in Psychology, 5*, 1–16. http://doi.org/10.3389/fpsyg.2014.01147

*Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition, 118*(2), 245–257. http://doi.org/10.1016/j.cognition.2010.10.012

*Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology, 25*(5), 547–567. http://doi.org/10.1080/20445911.2013.812093

*Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism (Cambridge, England), 17*(3), 610–629. http://doi.org/10.1017/S1366728913000564

*Blumenfeld, H. K., Schroeder, S. R., Bobb, S. C., Freeman, M. R., & Marian, V. (2016). Auditory word recognition across the lifespan: Links between linguistic and nonlinguistic inhibitory control in bilinguals and monolinguals. *Linguistic Approaches to Bilingualism*, *6*(1), 119–146. http://doi.org/10.1075/lab.14030.blu

*Bogulski, C. A. (2013). *Are bilinguals better learners? A neurocognitive investigation of the bilingual advantage* (Doctoral dissertation). The Pennsylvania State University, PA. Retrieved from ProQuest Dissertations and Theses Database. (Accession Order No. AAT 3585561)

*Bogulski, C. A., Rakoczy, M., Goodman, M., & Bialystok, E. (2015). Executive control in fluent and lapsed bilinguals. *Bilingualism: Language and Cognition*, *18*(3), 561–567. http://doi.org/10.1017/S1366728914000856

*Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science, 14*(2), 256–269. http://doi.org/10.1111/j.1467-7687.2010.00974.x

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624-652.

*Boutla, M. (2003). Cognitive mechanisms underlying capacity limits in working memory: Insights from American Sign Language (Doctoral dissertation). University of Rochester, NY. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*(8-).

*Brito, N. H., Murphy, E. R., Vaidya, C., & Barr, R. (2016). Do bilingual advantages in attentional control influence memory encoding during a divided attention task? *Bilingualism: Language and Cognition, 19*(3), 621–629. http://doi.org/10.1017/S1366728915000851

*Brown, J. M. (2015). *Task-switching in bilinguals: Further investigation of the bilingual advantage* (Doctoral dissertation). Illinois State University, IL. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 1599042)

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved from psyarxiv.com/9h3nu

*Cates, D.M. (2015). *Predictors of reading comprehension skill in deaf and hearing bilingual readers* (Doctoral dissertation). University of California, Davis, CA. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3706567)

*Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics, 77*(8), 2684–2693. http://doi.org/10.3758/s13414-015-0973-7

Chan, R. C. K., Hoosain, R., & Lee, T. M. C. (2002). Reliability and validity of the Cantonese version of the Test of Everyday Attention among normal Hong Kong Chinese: a preliminary report. *Clinical Rehabilitation*, *16*, 900–909. http://doi.org/10.1191/0269215502cr574oa

Chan, R. C. K., Lai, M. K., & Robertson, I. H. (2006). Latent structure of the Test of Everyday Attention in a non-clinical Chinese sample. *Archives of Clinical Neuropsychology*, *21*(5), 477–485. http://doi.org/10.1016/j.acn.2006.06.007

Chichibu, S., Bergstrom, J. C. R., Howard, D. V, Howard, J. H., Romano, J. C., Leblanc, N., & Howard, J. H. (2008). Japanese-English bilingualism influences control of attention but not implicit sequence learning*. Retrieved from http://www.researchgate.net/publication/242656805

Chong De La Cruz, I. A. (2016). *The role of language profiles in complex driving environments*. (Doctoral dissertation). Retrieved from ProQuest. (10007414)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J: L. Erlbaum Associates.

*Coderre, E, L., Smith, J. F., van Heuven, W. J. B., & Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Bilingualism: Language and Cognition, 19*(3), 1–18. http://doi.org/10.1017/S1366728915000188

*Coderre, E. L., van Heuven, W. J. B., Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition, 16*(2), 420–441. http://doi.org/10.1017/S1366728912000405

*Coderre, E. L., & van Heuven, W. (2014a). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology*, 5, 1–16. http://doi.org/10.3389/fpsyg.2014.01070

*Coderre, E. L., & van Heuven, W. (2014b). Electrophysiological explorations of the bilingual advantage: Evidence from a Stroop task. *PLoS ONE, 9*(7). http://doi.org/10.1371/journal.pone.0103424

*Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59–86. http://doi.org/10.1016/j.cognition.2006.12.013

*Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149. http://doi.org/10.1016/j.cognition.2009.08.001

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057–1074. http://doi.org/10.1037/0278-7393.32.5.1057

*Costumero, V., Rodríguez-Pujadas, A., Fuentes-Claramonte, P., & Ávila, C. (2015). How bilingualism shapes the functional architecture of the brain: A study on executive control in early bilinguals and monolinguals. *Human Brain Mapping, 36*(12), 5101–5112. http://doi.org/10.1002/hbm.22996

*Craik, F. I. M., & Bialystok, E. (2006). Planning and task management in older adults: Cooking breakfast. *Memory & Cognition, 34*(6), 1236–1249. http://doi.org/10.3758/BF03193268

*Dana-Gordon, C. (2013). *Bilinguisme et Fonctions Executives: Une approche développementale* [Bilingualism and executive functions: A developmental approach] (Doctoral dissertation). Universite Bordeaux Segalen 2, France. Retrieved from: http://www.theses.fr/2013BOR22079/abes

De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. *Journal of Cognitive Neuroscience*, *27(9)*, 1752-65. doi: 10.1162/jocn_a_00817

*de Bruin, A., Bak, T. H., & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language, 85*, 15–26. http://doi.org/10.1016/j.jml.2015.07.001

de Bruin, A., & Della Sala, S. (2015a). The decline effect: How initially strong results tend to decrease over time. *Cortex, 73*, 375–377. http://doi.org/10.1016/j.cortex.2015.05.025

de Bruin, A., Treccani, B., & Della Sala, S. (2015b). Cognitive advantage in bilingualism an example of publication bias? *Psychological Science*, *26(*1*)*, 99-107.

de Bruin, A., Treccani, B., & Della Sala, S. (2015c). The connection is in the data: We should consider them all. *Psychological Science*, *26(*6*)*, 946-949.

*de Leeuw, E. & Bogulski, C. (2015). L2 pronunciation proficiency, language use and age of acquisition as predictors of executive control in bilinguals. Proceedings of the 18th International Congress of Phonetic Sciences. University of Glasgow, UK. Paper number 1041.1-9, retrieved from http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS1041.pdf

*Delcenserie, A. & Genesee, F. (2016). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism*, 1-17. http://doi.org/10.1177/1367006916639158

*DePape, A-M. R. (2007). *Musical expertise, bilingualism and executive functioning* (Master's thesis). York University, Toronto, Ontario, Canada. Retrieved from ProQuest Dissertations and Theses Database.

Deslauries, G. (2008). *The role of selective attention in foreign accented speech perception*. (Master's thesis). Retrieved from ProQuest. (1466563)

Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition*, 56(2), 165-176.

Donnelly, S. (2016). *Re-examining the bilingual advantage on interference-control and task-switching tasks: A meta-analysis.* (Doctoral Dissertation). *CUNY Academic Works.* http://academicworks.cuny.edu/gc_etds/762

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*(3), 234-251. http://doi.org/10.1027/1618-3169/a000243

*Duncan, H. D. (2010). *The effects of aging and bilingualism on language-specific attention control* (Master's thesis). Concordia University, Montreal, Quebec, Canada. Retrieved from ProQuest Dissertations and Theses Database.

*Duncan, H. D., Segalowitz, N., & Phillips, N. A. (2016). Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 1–16. http://doi.org/10.1017/S136672891400025X

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315(7109)*, 629–634.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305-307.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16(*1*)*, 143-149.

Fan, J., McCandliss, B., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340-347.

*Feng, X. (2008). *Working memory and bilingualism: An investigation of executive control and processing speed* (Doctoral dissertation). York University, Canada. Dissertation Abstracts International: Section B: The Sciences and Engineering (Vol. 70).

*Fernandes, M. A., Craik, F., Bialystok, E., & Kreuger, S. (2007). Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. *Canadian Journal of Experimental Psychology – Revue Canadienne de Psychologie Experimentale*, *61*(2), 128–141. http://doi.org/10.1037/cjep2007014

*Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition*, 150, 119–132. http://doi.org/10.1016/j.cognition.2016.02.008

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A Latent-Variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. http://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science (Wiley-Blackwell), 17(2)*, 172–179. http://doi.org/10.1111/j.1467-9280.2006.01681.x

*Friesen, D. C., Luo, L., Luk, G., & Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, Cognition and Neuroscience*, 3798, 1–13. http://doi.org/10.1080/23273798.2014.918630

*Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, 53(4), 1272–1278. http://doi.org/10.1016/j.neuroimage.2010.05.078

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, *31*(3), 303–327. http://doi.org/10.1080/23273798.2015.1068944

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11. http://doi.org/10.3389/fpsyg.2014.00011

*Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience, 33*(2), 387–396. http://doi.org/10.1523/JNEUROSCI.3837-12.2013

Gollan, T. H., Montoya, R. I., Cera, C,,& Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814.

*Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16*(4), 562–576. http://doi.org/10.1037/0894-4105.16.4.562

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38(4)*, 404.

Green, D., W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cogntion*, *1*, 67-81.

Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual speakers. *Frontiers in Psychology*, *2*, 103.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25(5)*, 515–530. http://doi.org/10.1080/20445911.2013.796377

*Greene, M., R. (2015). *The bilingual advantage in switching: Are two better than one?* (Doctoral dissertation). University of Houston, TX. Retrieved from: https://uh-ir.tdl.org/uh-ir/bitstream/handle/10657/1273/GREENE-DISSERTATION-2015.pdf?sequence=1&isAllowed=y

Grundy, J. G., & Timmer, K. (2016). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*. http://doi.org/10.1177/0267658316678286

*Guido Mendes, C. (2015). *The impact of bilingualism on conflict control* (Doctoral dissertation). University of Otago, Dunedin, New Zealand. Retrieved from WorldCat Theses Database.

*Gutierrez, M. (2009). *A study of possible pre-cognitive advantages of bilingualism* (Master's thesis). The University of Texas at El Paso, TX. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 1473867)

*Gutierrez, M. (2013). *Strengthening cognitive development in minority populations: A study of the beneficial effects of bilingualism* (Doctoral dissertation). The University of Texas at El Paso, TX. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3609486)

Hakuta, J. (1986). *Mirror of Language: The debate of bilingualism.* New York, NY: Basic Books.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18(*10*)*, 1279-1296.

*Hamlet, A. (2015). *Executive functioning in mono vs. bilingual young adults: Assessment using language-based tools* (Doctoral dissertation). St. John's University, NY. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3662958)

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition, 150,* 10–19. http://doi.org/10.1016/j.cognition.2016.01.016

*Heidlmayr, K., Hemforth, B., Moutier, S., & Isel, F. (2015). Neurodynamics of executive control processes in bilinguals: Evidence from ERP and source reconstruction analyses. *Frontiers in Psychology*, *6*, 821. http://doi.org/10.3389/fpsyg.2015.00821

*Heidlmayr, K., Moutier, S., Hemforth, B., Courtin, C., Tanzmeister, R., & Isel, F. (2014). Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioral Stroop Colour Word task. *Bilingualism: Language and Cognition*, *17*(3), 630–645. http://doi.org/10.1017/S1366728913000539

*Hermans, S. (2012). *Attrition and language reversion in Dutch-English migrants in Australia: Working memory capacity and inhibitory functioning across languages and age groups* (Master's thesis). Utrecht University, Netherlands. Retrieved from: https://dspace.library.uu.nl/handle/1874/254053

*Hernández, M., Costa, A., Fuentes, L. J., & Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition, 13*(3), 315–325. http://doi.org/10.1017/S1366728909990010

*Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language, 69*(3), 257–276. http://doi.org/10.1016/j.jml.2013.06.004

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*(4), 625–658. http://doi.org/10.3758/s13423-011-0116-7

Hilchey, M. O., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In: J. H. Schwieter

(Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 586–613). Cambridge University Press.

*Houtzager, N., Lowie, W., Sprenger, S., & De Bot, K. (2015). A bilingual advantage in task switching? Age-related differences between German monolinguals and Dutch-Frisian bilinguals. *Bilingualism: Language and Cognition*, 1–11. http://doi.org/10.1017/S1366728915000498

*Incera Burkert, S. (2016). *Bilingualism across the adult life-span: Age and language usage are continuous variables* (Doctoral dissertation). Cleveland State University, OH. Retrieved from: http://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1873&context=etdarchive

*Incera, S., & McLennan, C.T. (2015). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, 1–11. http://doi.org/10.1017/S1366728915000218

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016) Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, 144, 84-97. doi: 10.1016/j.jecp.2015.11.009

*Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*(703), 1–11. http://doi.org/10.3389/fpsyg.2016.00703

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology Review*, *17*, 213–233.

Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals:

An Internet-based study. *Journal of Cognitive Psychology*, *29*(4), 1–16. http://doi.org/10.1080/20445911.2017.1282489

Kane, M.J., Hambrick, D.Z., & Conway, A.R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66-71

*Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705–10. http://doi.org/10.3758/PBR.16.4.705

*Kazemeini, T., & Fadardi, J. S. (2015). Executive Function: Comparing bilingual and monolingual Iranian university students. *Journal of Psycholinguistic Research, 45*(6), 1315–1326. http://doi.org/10.1007/s10936-015-9403-1

*Keijzer, M. (2013). Working memory capacity, inhibitory control and the Role of L2 proficiency in aging L1 Dutch-speakers of near-native L2 English. *Brain Sciences, 3*(3), 1261–1281. http://doi.org/10.3390/brainsci3031261

*Kerrigan, K., Bright, P., Thomas, M, S., C., Filippi, R. (2016). Evidence of an advantage in visuo-spatial memory for bilingual compared to monolingual speakers. *Bilingualism: Language and Cognition, 20*(3), 602-612.

*Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology, 26*(6), 640–648. http://doi.org/10.1080/20445911.2014.929580

Klein, R. M. (2016). What cognitive processes are likely to be exercised by bilingualism and does this exercise lead to extra-linguistic cognitive benefits? *Linguistic Approaches to Bilingualism*, *6(5)*, 549-564.

*Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research*, 1446, 71–90. http://doi.org/10.1016/j.brainres.2012.01.052

*Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, 8, 250. http://doi.org/10.3389/fnbeh.2014.00250

Kovelman, I. (2006). Bilingual and monolingual brains compared: A fMRI study of semantic processing (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3219722)

*Kramer, R., & Mota, M. B. (2015). Effects of bilingualism on inhibitory control and working memory: A study with early and late bilinguals. *Gragoatá, 20*(38), 309–331. Retrieved from http://www.gragoata.uff.br/index.php/gragoata/article/view/422

Kroll, J., & Gollan, T. H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), The Oxford Handbook of Language Production. doi: 10.1093/oxfordhb/9780199735471.013.001

*Ladas, A. I. (2013). *Disentangling the effect of bilingualism in attention from socioeconomic influences: A lifespan approach* (Doctoral dissertation). The University of Sheffield, UK. Retrieved from NDLTD Theses Database.

Laine, M., & Lehtonen, M. (2018). *Cognitive consequences of bilingualism: Where to go from here?* Manuscript submitted for publication.

*Lee, T. M. C., & Chan, C. C. H. (2000). Stroop interference in Chinese and English. *Journal of Clinical and Experimental Neuropsychology, 22*(4), 465–471. http://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT465.

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia*, *50*(7), 1362–1371. http://doi.org/10.1016/j.neuropsychologia.2012.02.021

Lehtonen, M. & Laine, M. (2003). How word frequency affects morphological processing in mono- and bilinguals. *Bilingualism: Language and Cognition*, 6, 213 – 225.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018, January 8). *Bilingualism and Executive Functions*. Retrieved from osf.io/365v8

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*, 301–324. http://doi.org/10.1016/j.cortex.2014.05.001

*Limberger, B. (2014). *O desempenho de bilíngues e multilíngues em tarefas as de controle inbitirió e compreensão auditiva* [The performance of bilinguals and monolinguals in tasks of inhibitory control and auditory comprehension] (Master's thesis). Pontificia Universidade Catolica do Rio Grande, Porto Alegre, Brazil. Retrieved from: http://repositorio.pucrs.br/dspace/bitstream/10923/5658/1/000454131-Texto%2BCompleto-0.pdf

*Ljungberg, J.K., Hansson, P., Andre´s, P., Josefsson, M., & Nilsson, L.G. (2013). A longitudinal study of memory advantages in bilinguals. *PLoS ONE* 8(9). http://doi.org/10.1371/journal.pone.0073029

*Lourtie, M. & Coen, J. (2015). *Le profil cognitif de l'interprète simultané* [The cognitive profile of simultaneous interpreters]. (Master's thesis). Université Catholique Delouvain Université libre de Bruxelles, Belgium. Retrieved from:

https://dial.uclouvain.be/downloader/downloader.php?pid=thesis%3A2837&datastream=PDF_01

*Luk, G. C. (2008). *The anatomy of the bilingual influence on cognition: Levels of functional use and profiency of language* (Doctoral dissertation). York University, Toronto, Ontario, Canada. Retrieved from ProQuest Dissertations and Theses Database.

*Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition, 74*(3), 347–357. http://doi.org/10.1016/j.bandc.2010.09.004

*Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition, 14*(04), 588-595. https://doi.org/10.1017/S1366728911000010

*Luo, L., Craik, F. I. M., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a working memory task: Evidence from aging. *Psychology and Aging, 28*(1), 28–34. http://doi.org/10.1037/a0030875

*Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*(1), 29–41. http://doi.org/10.1016/j.cognition.2009.08.014

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), 4398-4403.

*Marian, V., Chabal, S., Bartolotti, J., Bradley, K., & Hernandez, A. E. (2014). Differential recruitment of executive control regions during phonological competition in monolinguals and bilinguals. *Brain and Language*, 139, 108–117. http://doi:10.1016/j.bandl.2014.10.005.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), S1366728903001068. http://doi.org/10.1017/S1366728903001068

*Martínez Reyes, D. A. (2013). *The effect of multilingualism on executive functioning and memory in a sample of college-educated older adults* (Doctoral dissertation). Fuller Theological Seminary, CA. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3604972)

*Martins, S., A. (2010). *Diferenças entre idosos bilíngues e monolíngues no desempenho de tarefas relacionadas às funções executivas, memória de trabalho e memória emocional de longo prazo* [Differences between elderly bilinguals and monolinguals in performance in tasks of executive functions, working memory and long-term emotional memory]. (Doctoral dissertation). Universidade Católicade Pelotas, Brazil. Retrieved from: http://tede.ucpel.edu.br:8080/jspui/handle/tede/111

*Marzecová, A., Asanowicz, D., Krivá, L., & Wodniecka, Z. (2012). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition*, 1–16. http://doi.org/10.1017/S1366728912000569

*Marzecová, A., Bukowski, M., Correa, Á., Boros, M., Lupiáñez, J., & Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: The role of flexibility in temporal preparation and category switching. *Journal of Cognitive Psychology, 25*(5), 586–604. http://doi.org/10.1080/20445911.2013.809348

Meuter, R. F. I., & Ehrich, J. F. (2012). The acquisition of an artificial logographic script and bilingual working memory: Evidence for L1-specific orthographic processing skills transfer in Chinese–English bilinguals. *Writing Systems Research*, *4*, 8–29. http://doi.org/10.1080/17586801.2012.665011

Miyake, A., & Friedman, N. P. (2012). The nature and organisation of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21(1)*, 8–14. http://doi.org/10.1177/0963721411429458.The

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

*Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2015). Musical training, bilingualism, and executive function: A closer look at task switching and dual-task performance. *Cognitive Science, 39*, 992–1020. http://doi.org/10.1111/cogs.12183

*Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia, 66*, 157–169. http://doi.org/10.1016/j.neuropsychologia.2014.11.014

*Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics, 23*(6), 564–579. http://doi.org/10.1016/j.jneuroling.2010.05.002

Moreno, S. G., Sutton, A. J., Ades, A. E., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Medical Research Methodology*, *9*(2).

*Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific effects. *PLoS ONE 9(*4). http://doi.org/10.1371/journal.pone.0094169

*Morini, G. (2014). *Interactions between language experience and cognitive abilities in word learning and word recognition* (Doctoral dissertation). University of Maryland, MD.

Retrieved from ProQuest Dissertations and Theses Database. (Accession Order No. AAT 3644000)

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719-726. http://doi.org/10.1111/j.1467-7687.2007.00623.x

Mukadam, N., Sommerlad, A., & Livingston, G. (2017). The relationship of bilingualism compared to monolingualism to the risk of cognitive decline or dementia: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, *58*, 45-54. doi: 10.3233/JAD-170131.

Münte, T., Altenmüller, E. & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473-478.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2013). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12(2)*: 241–268. doi:10.3758/s13415-011-0083-5.

*Nikolova, A. G. (2007). *Bilingualism, inhibition, and executive processing: Evidence from Stroop color naming, proactive interference, task-switching, and working memory* (Doctoral dissertation). Claremont Graduate University, CA. Retrieved from WorldCat Theses Database. (UMI No. 3272607)

Oberauer, K., Schulze, R., Wilhelm, O., & Süss, H.M. (2005). Working memory and intelligence--their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61-65. doi: 10.1037/0033-2909.131.1.61

*Olsen, R. K., Pangelinan, M. M., Bogulski, C., Chakravarty, M. M., Luk, G., Grady, C. L., & Bialystok, E. (2015). The effect of lifelong bilingualism on regional grey and white matter volume. *Brain Research, 1612*, 128–139. http://doi.org/10.1016/j.brainres.2015.02.034

*Oschwald, J. (2014). *The shifting facet of executive functions : Testing the bilingual advantage* (Master's thesis). Philosophischen Fakultät der Universität Zürich, Switzerland. Retrieved from WorldCat Theses Database.

*Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6), 615–639. http://doi.org/10.1080/20445911.2014.944914

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. http://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R. & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29(2)*, 89-112. DOI: 10.1080/20445911.2016.1248436

*Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology 5*, 1–15. http://doi.org/10.3389/fpsyg.2014.00962

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *1*, 81-93. doi: 10.1016/j.jneumeth.2016.10.002

Pelham, S. D. (2012). Evaluating lexical access and executive function in monolinguals, late bilinguals, and early bilinguals* (Master's thesis). University of Florida, FL. Retrieved from WorldCat Theses Database.

*Pelham, S. D. (2014). *Monolinguals' and bilinguals' attentional control in the presence of cognitive and emotional distraction* (Doctoral dissertation). University of Florida, FL. Retrieved from ProQuest Dissertations and Theses Database.

*Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 313–325. http://doi.org/10.1037/a0035224

Peters, J. L., Sutton, A. J., Jobes, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *31*(10), 991–996.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, *26*(25), 4544–4562.

*Pinto, L. M. (2009). *O papel do bilinguismo na preservacao dos processos executivos da memoria de trabalho no envelhecimento*. [The role of bilingualism in preserving executive processes of working memory during aging]. (Master's thesis). Universitário Ritter dos Reis, Porto Alegre, Brazil. Retrieved from: http://livros01.livrosgratis.com.br/cp117472.pdf.

*Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, *22*(3), 415–422. http://doi.org/10.1016/j.acn.2007.01.015

*Prior, A., & Gollan, T. H. (2011). Good language-switchers are Good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*(4), 682–691. http://doi.org/10.1017/S1355617711000580

*Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology, 25*(5), 622–645. http://doi.org/10.1080/20445911.2013.821993

*Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253. http://doi.org/10.1017/s1366728909990526

*Rabkina, I. (2014). *Examining the intersection of the cognitive advantages and disadvantages of the bilingual brain* (Senior thesis). Scripps College, CA.

R Core Team. (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ramos, S., Fernández García, Y., Antón, E., Casaponsa, A., & Duñabeitia, J. A. (2017). Does learning a language in the elderly enhance switching ability? *Journal of Neurolinguistics, 43*, 39–48. http://doi.org/10.1016/j.jneuroling.2016.09.001

*Ransdell, S., Barbier, M.-L., & Niit, T. (2006). Metacognitions about language skill and working memory among monolingual and bilingual college students: When does multilingualism matter? *International Journal of Bilingual Education and Bilingualism, 9*(6), 728–741. http://doi.org/10.2167/beb390.0

*Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*(1), 1–11. http://doi.org/10.1080/20445911.2014.976226

*Rietbergen, M. (2014). *An investigation of neuroplasticity and cognitive performance.* (Master's thesis). University of Utrecht, The Netherlands. Retrieved from: https://dspace.library.uu.nl/handle/1874/298035

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8(3), 271-276.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2017, September 28). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/xlm0000450

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*, 747-758.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The test of everyday attention. *Journal of the International Neuropsychological Society : JINS, 2(6)*, 525–534. http://doi.org/10.1017/S1355617700001697

Robertson, I. H., Ward, T., Ridgeway, V., Nimmo-Smith, I., & McAnespie, A. W. (1994). The test of everyday attention (TEA). Bury St. Edmonds, United Kingdom: Thames Valley Test Company.

*Rodríguez-Pujadas, A., Sanjuán, A., Ventura-Campos, N., Román, P., Martin, C., Barceló, F., … Cesar Ávila (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS ONE 8*(9). https://doi.org/10.1371/journal.pone.0073028

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. http://doi.org/10.1037/0096-3445.124.2.207

*Romano, J. C. (2009). *Bilingualism, Cognition, and Successful Aging* (Doctoral dissertation). The Catholic University of America. Retrieved from WorldCat Theses Database. (UMI No. 3379737)

*Rosselli, M., Ardila, A., Araujo, K., Weekes, V. a, Caracciole, V., Padilla, M., & Ostrosky-Solis, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*, 17–24. http://doi.org/10.1207/S15324826AN0701

*Rosselli, M., Ardila, A., Lalwani, A, N., & Vélez-Uribe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *19*(193), 489–503. http://doi.org/10.1017/S1366728915000309

*Rosselli, M., Ardila, A., Santisi, M. N., Del, M., Arecco, R., Salvatierra, J., & Lenis, B. (2002). Stroop effect in Spanish-English bilinguals. *Journal of the International Neuropsychological Society,* 8, 819–827. https://doi.org/10.1017/S1355617702860106

Roth, D. (2003). *Effect of language status on neuropsychological test performance in elderly nursing home residents.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3098138)

*Rutkoski-Rodrigues, L. (2013). *Cognitive differences between monolinguals and bi/multilinguals: Executive functions boosted by code-switching?* (Master's thesis). Universidade Católica de Pelotas, Brazil. Retrieved from http://tede.ucpel.edu.br:8080/jspui/bitstream/tede/284/1/lisandra.pdf

*Rutkoski Rodrigues, L. & Zimmer, M. C. (2015). Bilingualism and inhibitory control: Possible confounds with the variables "profession" and "level of education". *Calidoscópio 13*(1), 104-112. http://doi.org/10.4013/cld.2015.131.10

*Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language, 74*, 46–76. http://doi.org/10.1016/j.jml.2014.04.003

*Sabourin, L., & Vinerte, S. (2015). The bilingual advantage in the Stroop task: Simultaneous vs. early bilinguals. *Bilingualism: Language and Cognition, 18*(2), 350-355.

*Salvatierra, J. L., & Rosselli, M. (2010). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism 15(*1), 1–12. http://doi.org/10.1177/1367006910371021

*Samuel, S. (2015). *An investigation of a proposed bilingual advantage in aspects of executive function: Evidence from visual perspective taking and Simon tasks* (Doctoral dissertation). University of Essex, UK. Retrieved from NDLTD Theses Database.

Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., & Adrover-Roig, D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *PLoS One*, *12*(4), e0176151. http://doi.org/10.1371/journal.pone.0176151

*Sandoval, T. (2010). *The role of control in bilingual verbal fluency: Evidence from aging and Alzheimer's disease* (Doctoral dissertation). San Diego State University, CA. Retrieved from WorldCat Theses Database. (UMI No. 3412317)

*Scaltritti, M., Peressotti, F., & Miozzo, M. (2015). Bilingual advantage and language switch: What's the linkage? *Bilingualism: Language and Cognition*, *20*(1), 80-97. http://doi.org/10.1017/S1366728915000565

*Schroeder, S. R., Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology, 24*(5), 591-601. http://doi.org/10.1080/20445911.2012.669367

*Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2015). Bilingualism and musicianship enhance cognitive control. *Neural Plasticity*, vol. *2016*. http://dx.doi.org/10.1155/2016/4058620

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 1–10. http://doi.org/10.3389/fpsyg.2014.00772

*Shulley, L. J., & Shake, M. C. (2016). Investigating the relationship between bilingualism, cognitive control, and mind wandering. *Journal of Cognitive Psychology*, *28*(3,) 257-274. http://doi.org/10.1080/20445911.2015.1128438

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51(3)*, 300

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-Curve and effect size. Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*(6), 666–681.

*Smithson, L., & Nicoladis, E. (2015). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 934–944. http://doi.org/10.1017/S1366728913000175

Soveri, A., Antfolk, J., Karlsson, L., Salo, B. & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, doi: 10.3758/s13423-016-1217-0

*Soveri, A., Laine, M., Hämäläinen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition, 14*(3), 371–378. http://doi.org/10.1017/S1366728910000118

Soveri, A., Lehtonen, M., Karlsson, L.C., Lukasik, K., Antfolk, J., & Laine, M. (2016). Test–retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology: Adult*, doi: 10.1080/23279095.2016.1263795

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011a). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology*, *2*. http://doi.org/10.3389/fpsyg.2011.00183

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143(2)*, 850-886. doi: 10.1037/a0033981

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643.

*Suárez, P. (2013). *The role of bilingualism on neuropsychological test performance among Spanish speakers tested in their native language* (Doctoral dissertation). University of California, San Diego, CA. Retrieved from http://escholarship.org/uc/item/92f4233q

Sullivan, M. D., Janus, M., Moreno, S., Astheimer, L., & Bialystok, E. (2014). Early stage second-language learning improves executive control: Evidence from ERP. *Brain and Language, 139*, 84–98. http://doi.org/10.1016/j.bandl.2014.10.004

Tabares, J.G. (2012). *Phonological influences in verbal working memory in monolinguals and bilinguals* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3494340)

*Taler, V., Johns, B. T., Young, K., Sheppard, C., & Jones, M. N. (2013). A computational analysis of semantic structure in bilingual verbal fluency performance. *Journal of Memory and Language, 69*(4), 607–618. http://doi.org/10.1016/j.jml.2013.08.004

*Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, *2*, 1–19. http://doi.org/10.3389/fpsyg.2011.00123

*Tao, L., Taft, M., & Gollan, T. H. (2015). The bilingual switching advantage: Sometimes related to bilingual proficiency, sometimes not. *Journal of the International Neuropsychological Society*, *21*(7), 531–544. http://doi.org/10.1017/S1355617715000521

*Teubner-Rhodes, S. E. (2014). *The neural bases of the bilingual advantage in cognitive control – An investigation of conflict adaptation phenomena* (Doctoral dissertation). University of Maryland, College Park, MD. Retrieved from http://hdl.handle.net/1903/15283

*Tran, C. D., Arredondo, M. M., & Yoshida, H. (2015). Differential effects of bilingualism and culture on early attention: A longitudinal study in the U.S., Argentina, and Vietnam. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.00795

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition, 18(1)*, 3-24. http:// doi:10.1017/S1366728914000522

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45(2)*, 576–594. http://doi.org/10.3758/s13428-012-0261-6

*Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition, 137*, 40–46. http://doi.org/10.1016/j.cognition.2014.12.008

Verreyt, N., Woumans, E., Vandelanotte, D. & Szmalec, A. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*(*1*), 181-190. https://doi.org/10.1017/S1366728914000352

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

*Vinerte S., & Sabourin, L. (2015). Bilingualism and cognitive control: The ANT in a Canadian context. *Proceedings of the Annual Conference of the Canadian Linguistics Association*, Ottawa, Canada, 1-15.

Waris, O., Soveri, A., Ahti, M., Hoffing, R. C., Ventus, D., Jaeggi, S. M., … & Laine, M. (2017). A latent factor analysis of working memory measures using large-scale data. *Frontiers in Psychology, 8*, 1–14. http://doi.org/10.3389/fpsyg.2017.01062

*Waldie, K., Badzakova-Trajkov, G., Milivojevic, B., & Kirk, I. (2009). Neural activity during Stroop colour-word task performance in late proficient bilinguals: A functional magnetic resonance imaging study. *Psychology and Neuroscience*, *2*(2), 125–136. http://doi.org/10.3922/j.psns.2009.2.004

*Wierzbicki, S. (2014). *An investigation into the relationship between bilingualism and attentional networks with a focus on linguistic distance and language proficiency* (Master's thesis). University of the Witwatersrand, Johannesburg, South Africa. Retrieved from http://mobile.wiredspace.wits.ac.za/handle/10539/13400

*Wodniecka, Z., Craik, F. I. M., Luo, L., & Bialystok, E. (2010). Does bilingualism help memory? Competing effects of verbal ability and executive control. *International Journal of Bilingual Education and Bilingualism*, *13*(5), 575–595. http://doi.org/10.1080/13670050.2010.488287

*Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(5), 1579.* http://dx.doi.org/10.1037/xlm0000107

Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The Longitudinal Effect of Bilingual Immersion Schooling on Cognitive Control and Intelligence. *Language Learning*, 66: 76–91. doi:10.1111/lang.12171

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *Journal of Neuroscience*, *30*(22), 7646–7651. http://doi.org/10.1523/JNEUROSCI.1602-10.2010

*Wu, Y. J., Zhang, H., & Guo, T. (2016). Does speaking two dialects in daily life affect executive functions? An event-related potential study. *PLoS ONE*, *11*(3), 1–12. http://doi.org/10.1371/journal.pone.0150492

*Xie, Z., & Dong, Y. (2015). Contributions of bilingualism and public speaking training to cognitive control differences among young adults. *Bilingualism: Language and Cognition*, 20(1), 55-68. http://doi.org/10.1017/S1366728915000474

*Yamasaki, B. L., & Prat, C. S. (2014). The importance of managing interference for second language reading ability: An individual differences investigation. *Discourse Processes, 51*, 5-6, 445-467. http://doi.org/10.1080/0163853X.2014.913415

*Yang, E. (2015). *Cognitive effects of bilingualism: Executive function and language practice* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No.3714695)

*Yang, H., & Yang, S. (2015). Are all interferences bad? Bilingual advantages in working memory are modulated by varying demands for controlled processing. *Bilingualism: Language and Cognition*, *20*(1), 1–13. http://doi.org/10.1017/S1366728915000632

*Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology*, *146*, 121–136. http://doi.org/10.1016/j.jecp.2016.01.011

Yang, S., Yang, H. & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14(3)*, 412-422. https://doi.org/10.1017/S1366728910000611

*Yoo, J. (2014). *Domain-general working memory and syntactic processing in monolinguals vs. bilinguals* (Doctoral dissertation). University of Wisconsin-Madison, WI. Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3669010)

*Yoo, J., & Kaushanskaya, M. (2012). Phonological memory in bilinguals and monolinguals. *Memory & Cognition*, *40*(8), 1314–1330. http://doi.org/10.3758/s13421-012-0237-x

*Zavaleta, K. L. (2014). *The role of executive control in language learning* (Doctoral dissertation). The University of Arizona, AZ. Retrieved from NDLTD Theses Database.

Zhang, H., Kang, C., Wu, Y., Ma, F., & Guo, T. (2015). Improving proactive control with training on language switching in bilinguals. *Neuroreport*, *26*(6), 354–9. http://doi.org/10.1097/WNR.0000000000000353

Zhou, B., & Krott, A. (2016). Data trimming procedure can eliminate bilingual cognitive advantage. *Psychonomic Bulletin & Review*, *23*(4), 1221–1230. http://doi.org/10.3758/s13423-015-0981-6