2-2019

Does early active bilingualism enhance inhibitory control and monitoring? A propensity-matching analysis

Andree HARTANTO
Singapore Management University, andreeh@smu.edu.sg

Hwajin YANG
Singapore Management University, hjyang@smu.edu.sg

DOI: https://doi.org/10.1037/xlm0000581

Published in Journal of Experimental Psychology: Learning, Memory, and Cognition, 2018 July, Pages 1-20
https://doi.org/10.1037/xlm0000581

Does Early Active Bilingualism Enhance Inhibitory Control and Monitoring? A Propensity-Matching Analysis

Andree Hartanto and Hwajin Yang
Singapore Management University

Prior research suggesting that longer bilingual experience benefits inhibitory control and monitoring has been criticized for a lack of control over confounding variables. We addressed this issue by using a propensity-score matching procedure that enabled us to match early and late bilinguals on 18 confounding variables—for example, demographic characteristics, immigration status, fitness, extracurricular training, motivation, and emotionality—that have been shown to influence cognitive control. Before early and late bilinguals were matched ($N = 196$), we found early active bilingual advantages in flanker effects (in accuracy), global accuracy, and sensitivity (d') on the Attention Network Test for Interaction and Vigilance and global accuracy on the saccade task. After matching ($n = 113$), many of the early active bilingual advantages that had been identified before matching were either attenuated or disappeared. However, we observed that early active bilingual advantages in flanker effects (in response time) were strengthened after matching. These results stress robust early active bilingual advantages in inhibitory control and highlight the importance of matching language groups on nonlinguistic covariates.

Keywords: age of acquisition, early bilingualism, inhibitory control, monitoring, propensity matching

Research suggests that individual differences in various experiential factors—such as playing video games, musical training, and physical exercise—influence cognitive abilities over the life span (e.g., Colcombe & Kramer, 2003; Diamond, 2012; Hartanto, Toh, & Yang, 2016; Moreno et al., 2011; Valian, 2015). Among these experiential factors, the impact of challenging linguistic practices (i.e., bilingualism) on cognitive control has received the most notable empirical attention (e.g., Abutalebi et al., 2015; Bak, Long, Vega-Mendoza, & Sorace, 2016; Bialystok, Poarch, Luo, & Craik, 2014; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Hartanto, Toh, & Yang, 2018; Sorge, Toplak, & Bialystok, 2017; Woumans, Ceuleers, Van der Linden, Szmalec, & Duyck, 2015; Yang & Yang, 2016; see Paap & Greenberg, 2013, for an opposing view). In particular, numerous studies have demonstrated bilingual advantages in inhibitory control (i.e., the ability to deliberately resist or suppress the presence of potent internal and external distraction; Engle & Kane, 2004; Friedman & Miyake, 2004) or monitoring (i.e., the ability to detect the presence of conflict or a signal that demands a certain action; Costa, Hernández, & Sebastián-Gallés, 2008). However, these findings are challenged by a number of recent studies that have failed to find significant bilingual advantages in measures of inhibitory control and monitoring (e.g., Duñabeitia et al., 2014; Gathercole et al., 2014; Paap & Greenberg, 2013). In view of this inconsistency, we set out to revisit whether the duration of bilingual practice—early versus late bilingualism—moderates performance on inhibitory control and monitoring tasks, especially when participants are matched on 18 confounding variables by a propensity-score matching procedure (Rosenbaum & Rubin, 1983).

Theoretical Accounts

Given bilinguals' ability to speak two or more languages, observed bilingual advantages in inhibitory control and monitoring have been explained by two theoretical accounts—the inhibitory control model (Green, 1998) and the bilingual executive processing advantage hypothesis (Hilchey & Klein, 2011). The inhibitory control model (Green, 1998) postulates that when bilinguals' two languages are coactivated and compete with each other, bilinguals constantly recruit inhibitory control mechanisms to select the appropriate language while suppressing the irrelevant language, which is coactivated automatically. This theory suggests that bilinguals' routine practice of inhibiting the irrelevant language likely facilitates their ability to inhibit even nonlinguistic distractors (i.e., inhibitory control). On the other hand, Hilchey and Klein's (2011) bilingual executive processing advantage hypothesis postulates that bilinguals are better at global conflict-monitoring than are monolinguals because bilinguals' constant practice of coordinating two competitive languages should be beneficial for the monitoring aspects of cognitive control.

In contrast to the findings of many previous studies that support the theoretical predictions of Green's inhibitory control model (e.g., Abutalebi et al., 2015; Sorge et al., 2017; Yang, Yang, &

Andree Hartanto and Hwajin Yang, School of Social Sciences, Singapore Management University.

This study was supported by a grant awarded to Hwajin Yang by Singapore Management University through research Grant 14-C242-SMU-031 from the Ministry of Education Academy Research Fund Tier 1. We are grateful to Gaurav Singh, Lester Sim, Quanxiang Loo, Lu Yizhen, and Toh Wei Xing for their assistance with data collection and coding.

Correspondence concerning this article should be addressed to Hwajin Yang, School of Social Sciences, Singapore Management University, 90 Stamford Road, Level 4, Singapore 178903. E-mail: hjyang@smu.edu.sg

Lust, 2011) and Hilchey and Klein's (2011) hypothesis (e.g., Costa et al., 2009; Woumans et al., 2015), a notable number of studies have not found significant differences between bilinguals and monolinguals on measures of inhibitory control and monitoring (e.g., Duñabeitia et al., 2014; Gathercole et al., 2014; Paap & Greenberg, 2013). For example, Duñabeitia et al. (2014) found that bilingual ($n = 252$) and monolingual ($n = 252$) children performed equally across all indices of inhibitory control and monitoring, as measured by verbal and nonverbal versions of the Stroop task. Similarly, Paap and Greenberg (2013) tested college students in a series of high-powered studies but did not find any group differences on Simon or flanker tasks. This inconsistency suggests that more empirical studies are needed to shed light on our theoretical understanding of bilingual advantages in inhibitory control and monitoring.

Duration of Bilingual Practice and Inhibitory Control

In light of this discrepancy, recent reviews have highlighted the critical need to investigate various bilingual experiences that might moderate the manifestation of bilingual advantages in inhibitory control and monitoring (Bak, 2015; Woumans & Duyck, 2015). One bilingual experience that has been widely expected to moderate performance on inhibitory control and monitoring tasks is the duration of bilingual practice, that is, the age of bilingual acquisition. In terms of existing theoretical accounts (Green, 1998; Hilchey & Klein, 2011), it is reasonable to expect that a longer experience of language control (i.e., a longer duration of bilingual practice) would likely cause observable changes in inhibitory control and monitoring abilities, because early bilingualism typically entails longer bilingual practice, which likely facilitates adaptive transfer from language control to domain-general cognitive control (e.g., inhibitory control and monitoring).

To date, only a few studies have investigated the effect of early bilingualism on inhibitory control, despite growing interest in the subject. Specifically, Luk, De Sa, and Bialystok (2011) found that early bilinguals exhibited smaller flanker effects (i.e., the difference between incongruent trials and congruent trials) in response time (RT) than did late bilinguals and monolinguals, which supports the beneficial effect of early bilingualism on inhibitory control (for counterarguments, see Paap, Johnson, & Sawi, 2014). Subsequent studies, however, have failed to replicate findings that early bilinguals performed better than late bilinguals on tasks that measure inhibitory control (e.g., Humphrey & Valian, 2012; Kalia, Wilbourn, & Ghio, 2014; Kapa & Colombo, 2013; Pelham & Abrams, 2014; Tao, Marzecová, Taft, Asanowicz, & Wodniecka, 2011). For instance, Paap et al. (2014) found no differences between early and late bilinguals ($n = 384$) in inhibitory control, as measured by antisaccade, flanker, and Simon tasks. This apparent discrepancy warrants further investigation of the relationship between early bilingualism and inhibitory control (Paap, Darrow, Dalibar, & Johnson, 2015; Yang, Hartanto, & Yang, 2016).

Duration of Bilingual Practice and Monitoring

Studies on the link between bilingualism and monitoring have also yielded somewhat mixed findings. Some have found evidence in favor of early bilingualism (Kapa, Colombo, 2013; Paap et al., 2014; see also Luk et al., 2011; Pelham & Abrams, 2014; and Yow & Li, 2015, for different outcomes). For instance, Kapa and Colombo (2013) demonstrated that early bilingual children who had acquired their second language (L2) before age 3 had faster global RT—a commonly used measure of monitoring—than late bilingual children, who acquired their L2 after age 3. In the same vein, Paap et al. (2014) found that bilinguals who had been exposed to both their mother tongue (L1) and L2 since birth exhibited smaller mixing costs on the switching task than bilinguals who had been exposed to L2 after age 6. Although the authors claim that these effects are likely spurious, their finding is noteworthy, since mixing cost is calculated based on a comparison of switch blocks—which place high demands on monitoring—and a baseline block, which does not involve monitoring; thus, mixing cost is argued to be a purer measure of monitoring than a mere index of global RT (Paap & Greenberg, 2013). Notably, however, some empirical studies have obtained inconsistent findings; they did not detect positive effects of early bilingualism on monitoring (Luk et al., 2011; Pelham & Abrams, 2014; Yow & Li, 2015). For instance, Yow and Li (2015) did not find any correlations between the age of acquisition and mixing costs in the number-letter switching task. Moreover, Luk et al. (2011) and Pelham and Abrams (2014) found that early and late bilinguals had comparable global RT in the flanker task and attention network task (ANT).

Methodological and Conceptual Issues of Previous Research

A large part of these discrepancies in the literature can be attributed to methodological and conceptual issues. In particular, two crucial aspects should be addressed. First, the most critical methodological issue is the need to control for preexisting differences between early and late bilinguals (Paap et al., 2015). Since random assignment is not applicable in bilingualism research, demographic and individual difference variables, such as extracurricular activities or motivational factors, may confound the true effect of early bilingualism on inhibitory control and monitoring. Bak (2015) argues that confounding variables can operate in two ways: by producing spurious effects or masking genuine effects. Hence, controlling for the influence of potential confounding factors may explain the mixed results regarding the effect of early bilingualism on inhibitory control and monitoring. Although some previous studies have attempted to statistically control for the effects of confounds by conducting analysis of covariance (ANCOVA; e.g., Kapa & Colombo, 2013), ANCOVA does not always resolve preexisting group differences—which may obscure true bilingualism advantages or produce spurious effects (Paap et al., 2015)—especially when its core assumption, that covariates and experimental groups (i.e., independent variables) must be statistically independent (Miller & Chapman, 2001), is violated. For instance, if early and late bilingual groups are significantly different in socioeconomic status (SES), it is not appropriate to include SES as a covariate in ANCOVA, as the core assumption of statistical independence between covariates and IV is violated. This assumption regarding ANCOVA is crucial, since regression adjustment may attenuate part of a group effect or produce a spurious group effect (Elashoff, 1969; Miller & Chapman, 2001). Furthermore, previous studies that have attempted to control for confounding variables are nevertheless limited by their narrow focus on only a few confounds, such as demographic variables and nonverbal intelligence (Luk et al., 2011). However, a number of important confounding variables demand consideration, such as immigration status, fitness level, physical exercise, motivation, specialized skills (e.g., playing video games, musical training), and personality (see Valian, 2015, for a review). Therefore, the true effect of early bilingualism on inhibitory control and monitoring can be estimated only when a host of potential confounding variables are well matched between the comparison groups.

Second, inconsistent findings in the literature can also arise from different conceptualizations of the onset of bilingualism, which determines the duration of bilingualism. For instance, Luk et al. (2011) distinguished early and late bilinguals according to the onset of active bilingualism, whereas others did so according to the age of immersion in the L2 environment (Tao et al., 2011); age of L2 production (Kapa & Colombo, 2013); age of first exposure to L2 (Kalia et al., 2014); and age of L2 fluency (Pelham & Abrams, 2014). Consequently, the specific age criteria used to classify early and late bilinguals differ across those studies. Different conceptual nuances are critical, because numerous indices of early bilingualism may have varying degrees of sensitivity in capturing the extent of linguistic practice accumulated over time. Specifically, bilinguals who have been exposed to both L1 and L2 since birth likely acquired both languages at about the same time. However, depending on the language environment (e.g., school, neighborhood, community), some bilinguals may have used both languages actively since acquisition, while others may have relied on one language more than the other until a later age. Although these bilinguals could have acquired L2 at similar ages, they should have experienced different durations of active bilingualism. Given that not all bilinguals have actively used L2 since birth, the age of L2 acquisition by itself may not necessarily provide a reliable estimate that captures the amount of rigorous practice bilinguals have had with their two languages. As suggested by the adaptive control hypothesis (Green & Abutalebi, 2013), the key to an adaptive transfer from language control to domain-general cognitive control abilities is the longer duration during which a bilingual person has actively exercised language control (also see Hartanto & Yang, 2016a; Macnamara & Conway, 2014). This explains, in part, why Luk et al. found that the flanker effect was correlated with the age of active bilingualism, but not with the age of mere L2 acquisition.

The Present Study

In view of these methodological and conceptual limitations, we aimed to examine the following issues: First, we sought to address preexisting differences between early and late bilinguals by using a propensity-score matching method (Rosenbaum & Rubin, 1983) that tackles the inherent challenges in matching participants on a host of confounding variables (Kaushanskaya & Prior, 2015). The logic behind the propensity-score matching procedure is that covariate balance—which here refers to the distribution of observed covariates in early and late bilinguals—on a large set of observed covariates can be achieved through careful matching based on a single propensity score, defined as the conditional probability (ranging from 0 to 1) that expresses the likelihood of a unit of analysis (e.g., an individual) being assigned to a particular condition (e.g., early vs. late bilingual), given a set of observed covariates—for example, demographic characteristics, immigration status, fitness level, physical exercise, video game experience, musical training, task motivation, positive and negative affect, and state anxiety (Rosenbaum & Rubin, 1983; Thoemmes & Kim, 2011). When the propensity score of each participant is matched between the conditions, the matching procedure is expected to create balance between them on covariates that are used to estimate the propensity score. Because propensity scores allow us to systematically match participants on a large set of confounding variables, it is possible to evaluate the purer effects of bilingualism on inhibitory control and monitoring (Paap et al., 2015; Valian, 2015). Furthermore, propensity-score matching can improve causal inference in studies in which treatments (e.g., bilingual status) cannot be randomly assigned to participants (see Thoemmes & Kim, 2011, for a review). Previous research has consistently demonstrated that when relevant covariates are considered, propensity-score matching in quasi-experimental studies has produced a treatment effect estimate significantly closer to that of a true experiment (e.g., Shadish, Clark, & Steiner, 2008). Hence, to estimate an unbiased propensity score, we have assessed a myriad of nonlinguistic factors that have been argued to affect performance on inhibitory control and monitoring tasks—for example, demographic characteristics, immigration status, fitness level, physical exercise, video game experience, and musical training (see Valian, 2015, for a review). We have also included motivational and emotional variables that have been shown to influence executive control: task motivation (Pessoa, 2009); positive and negative affect (Yang, Yang, Ceci, & Isen, 2015; Yang, Yang, & Isen, 2013); and state anxiety (Eysenck, Derakshan, Santos, & Calvo, 2007; Hartanto & Yang, 2016b). By including directly relevant covariates, propensity-score matching can yield unbiased causal effect estimates of bilingualism duration on inhibitory control and monitoring.

Second, we aimed to address the conceptual issue in defining the onset of bilingualism. Consistent with Luk et al.'s (2011) conceptualization and classification, we focused on the onset age of active bilingualism, which is argued to provide a more reliable estimate regarding the amount of bilingual practice with coordinating coactivated languages, while simultaneously inhibiting irrelevant languages (Yang et al., 2016). Following Luk et al., we recruited college students and classified them as either early active or late active bilinguals, according to our cutoff age of 10 years for active bilingualism. We chose 10 as the cutoff age because our participants were around the age of 20, and therefore age 10 divided them into bilinguals who had used both languages for longer (or less) than one half of their lives (Luk et al., 2011). Given that various indices of bilingualism duration have been used in the literature, we also aimed to conduct exploratory analyses to directly compare the age of active bilingualism with other indices of age of L2 acquisition and age of L2 fluency; these will in turn elucidate the sensitivity of each index.

In line with Green's (1998) inhibitory control model and Hilchey and Klein's (2011) bilingual executive processing advantage hypothesis, we hypothesized that early active bilingualism will be significantly associated with improvement in inhibitory control and monitoring. We also expected that these advantages in inhibitory control and monitoring would remain significant after matching language groups on a host of nonlinguistic covariates. In addition, we hypothesized that the age of early active bilingualism would better predict inhibitory control and monitoring than age of L2 acquisition or age of L2 fluency, as the former more precisely captures the period during which bilinguals have experienced high demands on coordinating two activated languages and inhibiting the irrelevant language.

To test these hypotheses, we used the Attention Network Test for Interactions and Vigilance (ANTI-V; Roca, Castro, López-Ramón, & Lupiáñez, 2011) to assess both inhibitory control and monitoring and the saccade task to assess inhibitory control. We chose the ANTI-V for two reasons. First, the task has been shown to be a purer index of monitoring than global RT, which is sensitive to noncore, secondary processes such as motor speed, motivation, and response strategies (Paap & Greenberg, 2013). Another notable advantage of the ANTI-V, compared with the typical flanker task, is that the former assesses monitoring on the basis of a widely used measure of monitoring, that is, the Sustained Attention to Response Task paradigm (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997)—by requiring participants to detect infrequent stimuli during some of the trials while completing the flanker task. Through the use of signal detection theory (SDT) procedures, response bias can be partialed out from monitoring performance (See, Howe, Warm, & Dember, 1995), allowing us to obtain a more precise index of monitoring performance.

Despite this difference from other flanker tasks, we believe that the ANTI-V does not restrict the task's sensitivity in identifying bilingual advantages in flanker effects for the following reasons. Similar to the ANTI, which does not include a direct measure of vigilance (Roca et al., 2011), studies that used the ANTI-V demonstrate robust indices for three attentional networks (alertness, orienting, and executive control) and their interactions (Callejas, Lupiáñez, & Tudela, 2004). This supports the task's validity as a measure of flanker effects, despite its inclusion of a vigilance task. Moreover, it has been argued that the ANTI-V is thought to impose higher cognitive control than the typical ANT (Roca et al., 2011) because of the need to distinguish infrequent stimuli from central target stimuli while completing the flanker task, which requires a high level of vigilance and monitoring abilities throughout the task. Given that bilingual advantages in flanker effects are more pronounced, especially when the task imposes high monitoring demand (Costa et al., 2009; Jiao, Liu, Wang, & Chen, 2017), the ANTI-V, compared with the ANT, likely increases the task's sensitivities in detecting bilingual advantages by imposing high demand on cognitive control. For these reasons, therefore, we believe that the ANTI-V is as sensitive and effective as other typical flanker tasks in detecting bilingual advantages in flanker effects.

To establish convergent validity for inhibitory control (Paap & Greenberg, 2013), we used an additional measure of inhibitory control, the saccade task, which was adapted from Unsworth, Spillers, Brewer, and McMillan (2011). We chose the saccade task for the following reasons. First, previous studies have reliably demonstrated significant correlation between the saccade and flanker tasks (e.g., Unsworth, Brewer, & Spillers, 2012; Unsworth, Redick, Lakey, & Young, 2010; Unsworth & Spillers, 2010). Second, the literature has also documented that bilingualism modulates performance on the saccade task (e.g., Bialystok & Viswanathan, 2009). For instance, Bialystok, Craik, and Ryan (2006) found that young and older monolinguals had significantly larger antisaccade effects than their bilingual counterparts (for opposite results, see Paap & Greenberg, 2013). Third, given its nonlinguistic nature, the saccade task can be more useful than other commonly used tasks, such as the Stroop, for testing bilingual advantages in inhibitory control. In addition, compared to the Simon task, which has been plagued by ceiling effects among young adults (e.g., Bialystok, 2006) and low reliability (*r*s = .37–.43; Paap & Sawi, 2016; Soveri et al., 2016), the saccade task has been shown to be demanding and sufficiently reliable for young adults (Hutton & Ettinger, 2006; Unsworth et al., 2011).

It is notable, however, that although the ANTI-V and saccade tasks are regarded as measures of inhibitory control, each may tap into different aspects of inhibitory control. For instance, the saccade task has commonly been regarded as assessing prepotent response inhibition, whereas the flanker task has commonly been regarded as assessing resistance to distractor interference (Friedman & Miyake, 2004). However, as Friedman and Miyake (2004) have demonstrated using confirmatory factor analysis, prepotent response inhibition and resistance to distractor interference are closely related (r = .67) and can be collapsed into a single latent variable called *response-distractor inhibition*. Hence, we employed both the ANTI-V and the saccade task to establish convergent validity, at least for response-distractor inhibition.

Method

Participants

Bilingual students (N =220) from a local university in Singapore participated for either extra course credit or $15. Bilingual participants spoke a wide variety of languages in addition to English, which included Mandarin (n = 186), Malay (n = 7), Indonesian (n = 7), Vietnamese (n = 7), Tamil (n = 5), Hindi (n = 2), Korean (n = 2), Telugu (n = 1), Burmese (n = 1), Cantonese (n = 1), and Teochew (n = 1). On the basis of self-reported age of active bilingualism (Luk & Bialystok, 2013), participants were divided into either early or late bilinguals. The cutoff age for active bilingualism was 10 years; early active bilinguals had actively begun to use both languages before age 10, and late active bilinguals after age 10. Following the procedure outlined by Luk et al. (2011), we excluded 14 participants who reported that they had never actively used both languages in their daily lives. We excluded 7 participants who had used two languages actively at age 10 because they did not fit into either group. We also excluded 1 late active bilingual and 1 early active bilingual who had abnormal flanker effects (9.4 and 5.1 standard deviation above the overall mean, respectively). Last, 1 participant was excluded because of a technical error during data collection. This yielded a final prematched sample of 196 bilinguals, which included 142 early active bilinguals and 54 late active bilinguals. Participants' linguistic and nonlinguistic characteristics are presented in Tables 1 and 2. The study was approved by the Singapore Management University Institutional Review Board.

Tasks

ANTI-V. The ANTI-V, which was developed by Roca et al. (2011), was administered to assess inhibitory control and monitoring. For the ANTI-V, a row of five cars was presented in one of the parking lanes on each side of a two-lane road (see Figure 1). The central target car was flanked by four surrounding cars pointing either in the same direction (congruent condition) or the opposite direction (incongruent condition). Participants were in-

Table 1
Nonlinguistic Characteristics of Prematched and Postmatched Early and Late Bilinguals

Characteristic	Prematched: Early bilinguals (n = 142) M (SD) or %	Prematched: Late bilinguals (n = 54) M (SD) or %	Postmatched: Early bilinguals (n = 73) M (SD) or %	Postmatched: Late bilinguals (n = 40) M (SD) or %
Propensity score	.22 (.14)	.42 (.25)	.28 (.15)	.31 (.19)
Age	21.47 (1.56)	21.81 (1.60)	21.53 (1.70)	21.85 (1.63)
Sex (% males)	28.17%	38.89%	34.25%	42.50%
Immigration status (% immigrants)	9.15%	38.89%	16.44%	20.00%
Years of formal education	13.89 (1.70)	14.26 (1.49)	14.16 (1.60)	14.05 (1.47)
Paternal education level[a]	4.03 (1.15)	4.09 (1.17)	3.90 (1.15)	3.83 (1.17)
Maternal education level[a]	3.80 (.99)	3.91 (1.07)	3.67 (.94)	3.60 (1.01)
Monthly household income[b]	3.92 (2.24)	2.81 (1.94)	3.25 (1.82)	3.30 (2.02)
Nonverbal intelligence standard score (KBIT-2)	103.44 (18.31)	101.85 (17.98)	102.29 (18.70)	100.23 (17.42)
Health status[c]	2.91 (.82)	2.85 (.94)	2.93 (.79)	2.95 (.96)
Fitness level[c]	2.58 (.91)	2.54 (.93)	2.60 (.95)	2.58 (.93)
Hours of physical exercise per week	4.86 (6.44)	6.55 (15.11)	5.32 (8.02)	4.84 (4.00)
Years of musical training	3.12 (4.26)	2.17 (3.84)	2.92 (4.25)	2.18 (4.11)
Musical proficiency[d]	2.69 (1.13)	2.54 (1.06)	2.75 (1.14)	2.58 (1.11)
Hours of video game play per week	2.98 (7.01)	3.94 (8.05)	3.73 (8.46)	4.45 (8.73)
State anxiety[e]	10.83 (2.97)	11.04 (3.50)	10.85 (2.81)	10.85 (3.63)
Task motivation	3.61 (.62)	3.46 (.79)	3.56 (.62)	3.60 (.78)
Trait positive affect[f,g]	3.29 (.50)	3.16 (.56)	3.27 (.49)	3.18 (.57)
Trait negative affect[f,g]	2.38 (.51)	2.42 (.70)	2.44 (.51)	2.49 (.76)

Note. Standard deviations are shown in parentheses. KBIT-2 = Kaufman Brief Intelligence Test—Matrices subtest (2nd ed.; Kaufman & Kaufman, 2004).
[a] Parental education level was rated on a scale of 1 (*none*) to 6 (*master's or PhD*). [b] Household income was rated on a scale of 1 (*less than S$2,500*) to 9 (*more than S$20,000*), with intervals of S$2500. [c] Fitness level and health status were self-reported on a scale of 1 (*poor*) to 5 (*excellent*). [d] Musical proficiency was self-reported on a scale of 1 (*very poor*) to 5 (*very good*). [e] State anxiety was assessed by the State version of the State–Trait Anxiety Inventory (six items; Marteau & Bekker, 1992) on a scale of 1 (*not at all*) to 4 (*very much*). Scores for each response was summed to yield a score for state anxiety. Higher scores indicate a higher level of state anxiety. [f] Positive affect and negative affect were rated on a scale of 1 (*never*) to 5 (*always*) using the International Positive and Negative Affect Schedule Short Form (Thompson, 2007). Scores on the items were averaged, with a higher score indicating a higher level of positive or negative affect. [g] Data points from eight participants were lost due to a technical error in conducting the survey. The hot-deck imputation technique (Myers, 2011) was used to handle missing values.

structed to identify the direction in which the central target car was pointed as quickly and accurately as possible by pressing either *C* or *M* on the keyboard for left and right, respectively. The task's varying flankers, cues, and tone conditions allowed us to assess three attentional network scores: executive control, orienting, and phasic alerting.

In half of the trials, the central target and flankers were pointing in the same direction (i.e., congruent condition); in the other half, they were pointing in different directions (i.e., incongruent condition). An executive control efficiency score—which indicates conflict resolution between two competing responses (i.e., inhibitory control)—was calculated by mean RT difference between congruent and incongruent conditions.

Prior to presentation of the central target and flankers, the visual cue (asterisk) was briefly presented for 50 ms in one of two parking lanes, which were located either above or below the central fixation cross. There were three cue conditions: (a) no cue, in which no visual cue (asterisk) was presented; (b) valid cue, in which a visual cue appeared in the same location as the forthcoming central target car; and (c) invalid cue, in which the cue appeared in a location opposite to the target. The three cue conditions were equally probable. Because the invalid condition required participants to reorient to the actual target location, whereas the valid condition did not require reorienting, an orienting efficiency score was calculated by mean RT difference between valid and invalid cue conditions.

In addition, in half of the trials, an alerting tone was presented for 50 ms to signal presentation of the target stimuli (tone condition), but in the other half, it was not (no tone condition). Because the tone engaged alerting, an alerting network score was calculated by the difference in mean RT between tone and no tone conditions.

Finally, to assess attentional vigilance, the task asked participants to remain vigilant and detect infrequently occurring odd events, in which the central target was significantly displaced to either the right or the left. Participants were instructed to press an alternative response key (spacebar) when they detected a misplaced central target. The vertical and horizontal location of each displaced target car varied in each trial (i.e., ± 4 pixel). Because this oddity was embedded in 25% of the trials (vigilance condition), the vigilance component embedded in the ANTI-V required a sufficiently high level of monitoring effort; this resulted in an equal percentage of congruent (37.5%) and incongruent (37.5%) trials, which are roughly similar to the proportion of trials in Costa et al. (2008). The hit rate is the proportion of correct spacebar responses to infrequently occurring (dislocated) central targets in the vigilance condition, and the false alarm rate is the proportion of incorrect spacebar responses to normally located central targets in the nonvigilance condition. On the basis of SDT procedures, hit and false alarm rates were used to compute sensitivity (d'), which has been shown to reflect the ability to discern between displaced central targets and normally placed targets. In other words, it indicates monitoring performance that is not contaminated by

Table 2
Language Characteristics of Prematched and Postmatched Early and Late Bilinguals

Characteristic	Prematched: Early bilinguals (n = 142) M (SD)	Prematched: Late bilinguals (n = 54) M (SD)	T M (SD)	Postmatched: Early bilinguals (n = 73) M (SD)	Postmatched: Late bilinguals (n = 40) M (SD)	t
Age of active bilingualism	2.70 (3.25)	15.15 (3.10)	−24.27***	2.52 (3.22)	14.80 (3.23)	−19.37***
Age of L2 acquisition[a]	.77 (1.34)	3.95 (3.98)	−8.36***	.56 (1.14)	2.73 (3.34)	−5.03***
Age of L2 fluency[b]	7.74 (4.00)	12.94 (5.36)	−7.32***	7.88 (3.49)	12.08 (5.46)	−4.94***
PPVT-III standardized score	101.68 (8.20)	96.41 (7.42)	4.13***	100.49 (7.12)	97.40 (7.38)	2.18*
Recent L1 exposure (%)	63.92 (26.83)	56.37 (27.33)	1.75	60.89 (28.76)	61.48 (27.87)	−.10
Recent L2 exposure (%)	32.40 (25.55)	41.04 (26.20)	−2.10*	34.25 (26.91)	36.65 (27.06)	−.45
Daily usage of L1 (%)	64.76 (26.87)	58.19 (27.92)	1.51	60.56 (28.44)	63.55 (28.07)	−.54
Daily usage of L2 (%)	32.55 (25.90)	38.63 (26.68)	−1.46	35.45 (27.29)	34.15 (27.23)	.24
Self-reported L1 proficiency[c]						
Speaking	8.27 (1.39)	8.69 (1.34)	−1.87	8.40 (1.15)	8.40 (1.34)	−.01
Comprehension	8.43 (1.33)	8.91 (1.20)	−2.30*	8.53 (1.19)	8.65 (1.25)	−.49
Reading	8.02 (1.78)	8.46 (1.80)	−1.55	7.89 (1.88)	8.08 (1.90)	−.50
Self-reported L2 proficiency[c]						
Speaking	6.89 (1.73)	7.09 (1.89)	−.70	7.23 (1.70)	6.70 (1.86)	1.54
Comprehension	7.23 (1.70)	7.56 (1.91)	−1.15	7.41 (1.72)	7.18 (1.95)	.67
Reading	6.32 (2.26)	7.02 (2.23)	−1.95	6.48 (2.29)	6.50 (2.24)	−.05

Note. L2 = second language; PPVT-III = Peabody Picture Vocabulary Task (3rd ed); L1 = first language.
[a] Data from one participant were missing. [b] Data from three participants were missing. [c] Proficiency was rated on a scale of 1 (none) to 10 (perfect).
* $p < .05$. *** $p < .001$.

response bias (β), which reflects the degree of conservativeness in approaching the task (see See et al., 1995). Following the recommendation of Stanislaw and Todorov (1999), sensitivity (d') and response bias (β) were calculated as follows:

$$d' = \Phi^{-1}(\textit{Hit Rate}) - \Phi^{-1}(\textit{False Alarm})$$

$$\beta = \exp((\Phi^{-1}(\textit{Hit Rate})^{\wedge 2} - \Phi^{-1}(\textit{False Alarm})^{\wedge 2})/2)$$

For each trial, a fixation cross appeared in the center of a computer screen for an interval that varied from 400 ms to 1,600 ms and was followed by an alerting tone for 50 ms. After an interval of 350 ms, an orienting cue appeared for 50 ms and was followed by an interval of 50 ms. The target car and its flankers were then presented for 200 ms, and participants were required to respond within a 2,000 ms response window. The fixation cross remained throughout the entire study. The target car and its flankers appeared in one of two parking lanes, which were located above or below the fixation cross. The duration of the initial and final background scenes (i.e., parking lanes) were computed such that a trial's total duration would not exceed 4,100 ms. Participants completed seven blocks, each of which consisted of 48 nonvigilance and 16 vigilance trials. We based the number of nonvigilance trials on previous studies that employed the ANT (e.g., Costa et al., 2008; Pelham & Abrams, 2014). This was done to ensure the task's reliability while minimizing the possibility of a practice effect, which could attenuate the task's sensitivity to detect group differences. Visual feedback was provided only in the first block, which served as a practice block and thus was not included in data analysis. After the practice block, all test blocks were run consecutively without any rest periods in between (see Figure 1).

The saccade task. In the saccade task, which was adapted from Unsworth et al. (2011), participants were required to identify the masked target stimulus ("B," "P," or "R") on each trial as quickly and accurately as possible by pressing the corresponding key using the right index, middle, or ring finger, respectively. The task consisted of four blocks: a prosaccade practice block, a prosaccade experimental block, an antisaccade practice block, and an antisaccade experimental block. The order of prosaccade and antisaccade blocks was counterbalanced between participants to control for order effects, with the practice block preceding the experimental block (Kane, Bleckley, Conway, & Engle, 2001).

For each trial, a fixation cross appeared in the center of a computer screen for varying intervals (200 ms; 600 ms; 1,000 ms; 1,400 ms; 1,800 ms; or 2,200 ms). After this, a cue (i.e., a white "=") was flashed either to the left or right of the fixation cross for 100 ms, followed by a blank screen for 50 ms. The cue then appeared a second time for 100 ms, as if it had flashed briefly, which in turn strongly attracted participants' attention. After a blank screen of 100 ms, the target stimulus appeared on screen for 100 ms and was masked by a letter stimulus ("H") for 50 ms, and an "8" remained on the screen until a response was made.

The antisaccade and prosaccade blocks were identical, except for the spatial relation between the flashing cue and target. In the prosaccade block, the flashing cue always appeared in the same location as the target stimuli, whereas in the antisaccade block the flashing cue always appeared in a location opposite to the target stimuli. Participants were instructed to gaze at the flashing cue in the prosaccade block, while in the antisaccade block they were instructed to shift their attention from the flashing cue and gaze at the opposite side of the screen (see Engle & Kane, 2004, for detailed task descriptions).

The prosaccade practice block consisted of 18 trials (six trials per each target letter—B, P, and R). Because antisaccade trials were deemed more difficult than prosaccade trials, the antisaccade practice block consisted of 36 trials, which included one trial for

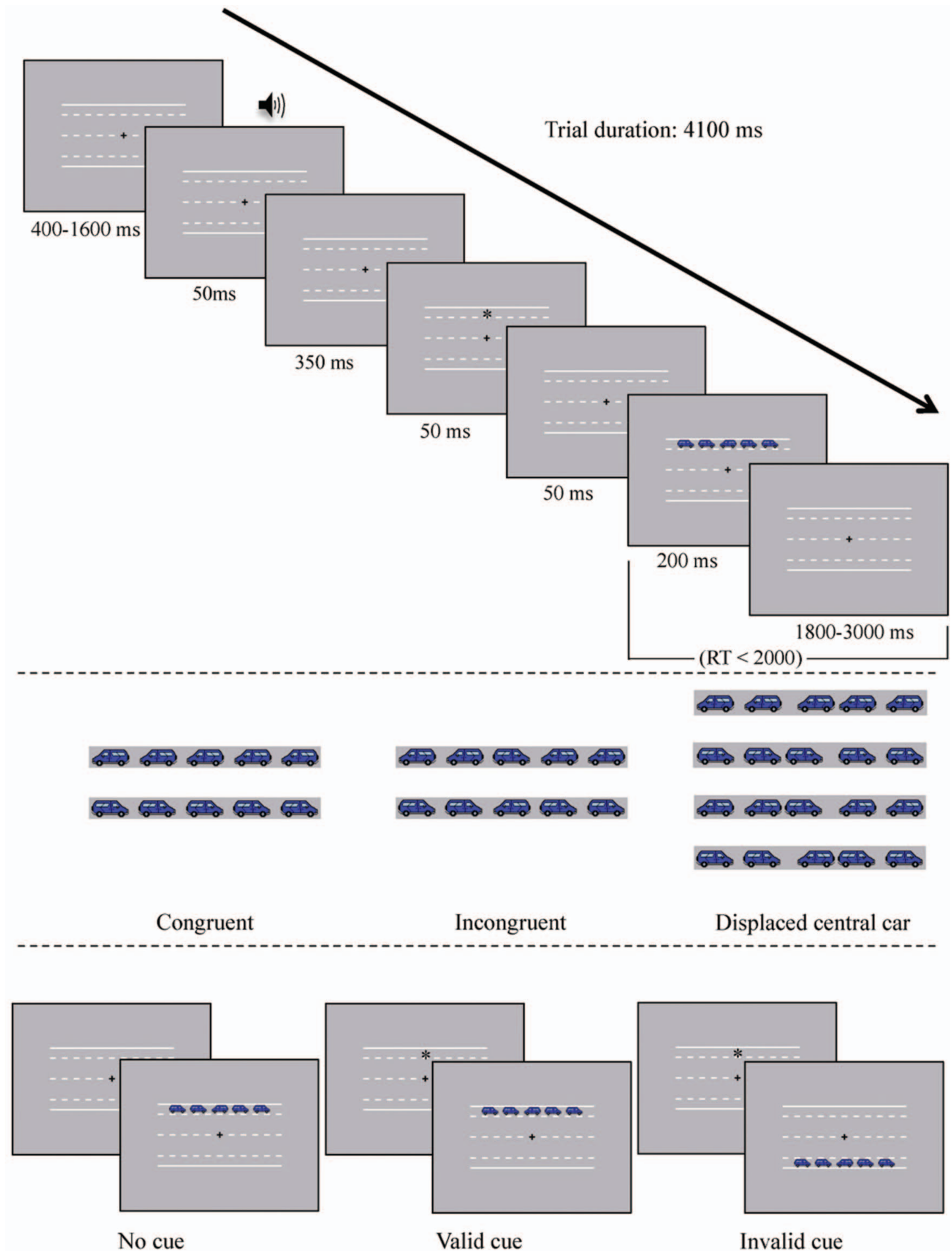

Figure 1. Schematic representation of the Attention Network Test for Interactions and Vigilance (ANTI-V) adapted from Roca et al. (2011). See the online article for the color version of this figure.

each of the combined conditions of the three target letters, six fixation durations, and two stimulus locations. Both prosaccade and antisaccade experimental blocks consisted of 108 trials, which included three trials for each of the combined conditions of the three targets, six fixation durations, and two stimulus locations.

Nonverbal intelligence. A computerized Kaufman Brief Intelligence Test—Matrices subtest (2nd ed. [KBIT-2]; Kaufman

& Kaufman, 2004) was used to measure nonverbal fluid intelligence. In the task, participants were presented with a series of images that depicted either concrete objects or abstract representations (e.g., designs or symbols) and asked to reason about the relationships between them. The KBIT-2 provides age-normed, standardized scores with a mean of 100 and a standard deviation of 15.

Receptive vocabulary. The Peabody Picture Vocabulary Task (3rd ed. [PPVT-III]; Dunn & Dunn, 1997) was used to assess participants' receptive vocabulary of English, which is the official instructional language used in schools in Singapore. In the PPVT-III, participants are presented with four pictures and asked to point to the word spoken by the experimenter. The PPVT-III provides age-normed, standardized scores with a mean of 100 and standard deviation of 15.

Procedure

Individual participants were seated in a cubicle and asked to complete the State version of the State–Trait Anxiety Inventory (Marteau & Bekker, 1992; $\alpha = .79$), along with an item that assessed their level of task motivation at that moment. Participants then completed the saccade task and ANTI-V, in that order. We chose this order to reduce potential fatigue, since the ANTI-V task took approximately 40 min and was considered to be exhausting. Afterward, participants completed the language background questionnaire adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), the International Positive and Negative Affect Schedule Short Form (I-PANAS-SF; Thompson, 2007; Positive Affect [PA] $\alpha = .75$; Negative Affect [NA] $\alpha = .80$), and a general questionnaire that assessed demographic characteristics, fitness level, health status, video game experience, and musical training. Finally, participants completed the computerized KBIT-2 and PPVT-III.

Results

Analysis of Prematched Samples

ANTI-V (RT). RTs that were either below 200 ms or 3 standard deviations above each participant's mean RT were excluded. For RT analysis in the nonvigilant condition, only accurate responses were included. Below, we compare early and late bilingual groups in terms of three attention network scores for executive control, orienting, and phasic alerting. For all of the significance tests, we performed two-tailed tests.

Executive control scores (i.e., incongruent—congruent) were submitted to a repeated-measures mixed factor ANOVA with bilingual group (early active vs. late active bilinguals) as a between-participants factor and flanker type (congruent vs. incongruent) as a within-participant factor. We found a significant main effect of flanker type, $F(1, 194) = 379.57$, $p < .001$, $\eta_p^2 = .662$, but not of bilingual group, $F(1, 194) = 0.16$, $p = .690$, $\eta_p^2 = .001$, which indicates that flanker types (congruent vs. incongruent) caused significant RT differences, but early active and late active bilinguals did not differ in their global RTs. We also found a marginally significant interaction between flanker type and bilingual group, $F(1, 194) = 3.37$, $p = .068$, $\eta_p^2 = .017$. However, further follow-up analyses found no significant group differences for either congruent or incongruent trials (Fs < 1; see Table 3).

Likewise, orienting scores calculated in RT were submitted to a similar repeated-measures mixed factor ANOVA with bilingual group (early active vs. late active bilinguals) as a between-participants factor and orienting cue (valid vs. invalid) as a within-participant factor. We found a significant main effect of orienting cue, $F(1, 194) = 269.89$, $p < .001$, $\eta_p^2 = .582$, indicating that RTs on valid-cue trials are faster than those on invalid-cue trials. However, neither the main effect of bilingual group, $F(1, 194) = 0.12$, $p = .730$, $\eta_p^2 = .001$, nor an interaction between orienting cue and bilingual group, $F(1, 194) = 0.98$, $p = .324$, $\eta_p^2 = .005$, was significant.

Table 3
Prematched Bilinguals' Response Latencies and Accuracy on the ANTI-V

	RT (ms)			Accuracy		
Condition	Early bilinguals ($n = 142$)	Late bilinguals ($n = 54$)	t	Early bilinguals ($n = 142$)	Late bilinguals ($n = 54$)	t
Global index	675 (103)	680 (98)	−.36	.92 (.09)	.88 (.12)	2.08*
Executive control	46 (31)	55 (36)	−1.83†	−.02 (.07)	−.06 (.15)	2.09*
Congruent	652 (102)	654 (94)	−.11	.93 (.10)	.91 (.10)	1.09
Incongruent	698 (106)	709 (106)	−.67	.91 (.10)	.86 (.18)	2.50*
Orienting	39 (26)	34 (32)	.99	.00 (.04)	−.01 (.04)	1.05
Valid	652 (106)	660 (102)	−.46	.92 (.10)	.89 (.12)	1.99*
Invalid	691 (100)	694 (94)	−.21	.92 (.09)	.88 (.12)	2.46*
Alerting	47 (47)	42 (42)	.65	−.04 (.07)	−.05 (.06)	1.08
Tone (no cue)	659 (107)	667 (105)	−.50	.93 (.08)	.90 (.12)	1.52
No tone (no cue)	705 (109)	709 (102)	−.22	.90 (.12)	.86 (.15)	1.67†
Sensitivity (d')				2.20 (.68)	1.84 (.89)	3.01*
Response bias (β)				14.21 (20.57)	11.11 (20.76)	.94
Hits				.53 (.21)	.49 (.23)	1.18
False alarms				.03 (.04)	.05 (.08)	−2.52*

Note. ANTI-V = Attention Network Test for Interactions and Vigilance; RT = response time.
† $p < .10$. * $p < .05$.

Alerting scores were submitted to a repeated-measures mixed factor ANOVA with bilingual group (early active vs. late active bilinguals) as a between-participants factor and alerting tone (no tone vs. tone in no cue condition) as a within-participant factor. We found a significant main effect of alerting tone, $F(1, 194) = 147.92$, $p < .001$, $\eta_p^2 = .433$, which indicates that the presence of an alerting tone facilitated RT. However, we found neither a main effect of bilingual group, $F(1, 194) = 0.14$, $p = .711$, $\eta_p^2 = .001$, nor an interaction between alerting tone and bilingual group, $F(1, 194) = 0.43$, $p = .514$, $\eta_p^2 = .002$.

ANTI-V (accuracy). A series of analyses similar to those described earlier was performed on accuracy data. To investigate group differences in executive control scores in terms of accuracy, a repeated-measures mixed factor ANOVA was conducted with bilingual group (early active vs. late active bilinguals) and flanker type (congruent vs. incongruent). We found two significant main effects of flanker type, $F(1, 194) = 25.00$, $p < .001$, $\eta_p^2 = .114$, and bilingual group, $F(1, 194) = 4.32$, $p = .039$, $\eta_p^2 = .022$, which suggests early active bilingual advantages in global accuracy. We also found a significant interaction between flanker type and bilingual group, $F(1, 194) = 4.37$, $p = .038$, $\eta_p^2 = .022$. Further analyses showed significant group differences for incongruent trials, $F(1, 194) = 6.24$, $p = .013$, $\eta_p^2 = .031$, but not for congruent trials, $F(1, 194) = 1.19$, $p = .276$, $\eta_p^2 = .006$, which suggests that early active bilingual advantages can be attributed to their better performance (in accuracy) on more difficult (incongruent) trials, which demand inhibitory control. In contrast, similar analyses of orienting and alerting network scores (in accuracy) did not yield any significant effects.

ANTI-V (vigilance). Indices for vigilance—which indicate aspects of monitoring—were computed only for the vigilance condition. There was no significant difference between early active and late active bilinguals in terms of hit rates, $F(1, 194) = 1.40$, $p = .238$, $\eta_p^2 = .007$. However, early active and late active bilinguals were significantly different in false alarm rates, $F(1, 194) = 6.34$, $p = .013$, $\eta_p^2 = .032$, and sensitivity (d'), $F(1, 194) = 9.06$, $p = .003$, $\eta_p^2 = .045$, which suggests that early active bilinguals were less likely to issue false alarms and more vigilant and discerning than late active bilinguals. We did not find any differences in response bias (β), $F(1, 194) = 0.89$, $p = .347$, $\eta_p^2 = .005$ (see Table 3).

Saccade tasks (RT). Only correct responses were used for analysis of RT data. RTs that were below 200 ms or 3 standard deviations above or below each participant's mean were excluded separately for prosaccade and antisaccade conditions. RT data were submitted to a repeated-measures mixed factor ANOVA, with bilingual group (early active vs. late active bilinguals) as a between-participants factor and task type (prosaccade vs. antisaccade) as a within-participant factor. We found a significant main effect of task type, $F(1, 194) = 117.01$, $p < .001$, $\eta_p^2 = .376$, which indicates that participants responded slower on the antisaccade task than on the prosaccade task. However, neither the main effect of bilingual group, $F(1, 194) = 0.19$, $p = .662$, $\eta_p^2 = .001$, nor its interaction with task type was significant, $F(1, 194) = 0.80$, $p = .371$, $\eta_p^2 = .004$. Follow-up analyses showed that the two bilingual groups did not differ in mean RT in either the antisaccade or prosaccade condition (see Table 4).

Saccade tasks (accuracy). A similar analysis was conducted on accuracy data in the saccade task, which is more commonly used as a dependent variable (e.g., Unsworth & Spillers, 2010). A repeated-measures mixed factor ANOVA was performed with bilingual group (early active vs. late active bilinguals) as a between-participants factor and task type (prosaccade vs. antisaccade) as a within-participant factor. We found significant main effects of task type, $F(1, 194) = 556.51$, $p < .001$, $\eta_p^2 = .742$, and bilingual group, $F(1, 194) = 8.34$, $p = .004$, $\eta_p^2 = .041$, but we did not find a significant interaction between task type and bilingual group, $F(1, 194) = 0.10$, $p = .751$, $\eta_p^2 = .001$. When follow-up analyses were performed to compare the two language groups in prosaccade and antisaccade conditions separately, we found a significant group difference in both prosaccade, $F(1, 194) = 5.63$, $p = .019$, $\eta_p^2 = .028$, and antisaccade conditions, $F(1, 194) = 6.49$, $p = .012$, $\eta_p^2 = .032$. This shows that early active bilinguals had significantly higher accuracy than late bilinguals in both prosaccade and antisaccade conditions.

In sum, we found early active bilingual advantages in inhibitory control and monitoring as evidenced by significant group differences in favor of early active bilingualism for the executive control network in accuracy and RT (marginally significant), global accuracy, sensitivity (d'), false alarms, and accuracy in prosaccade and antisaccade tasks.

Table 4
Prematched Bilinguals' Response Latencies and Accuracy on Saccade Tasks

Latency/accuracy	Early bilinguals ($n = 142$) M (SD)	Late bilinguals ($n = 54$) M (SD)	t
Response latencies			
Antisaccade	789 (253)	759 (218)	.73
Prosaccade	574 (167)	578 (138)	−.17
Antisaccade effect	214 (228)	181 (230)	.90
Accuracy			
Antisaccade	.54 (.15)	.48 (.16)	2.55*
Prosaccade	.88 (.17)	.81 (.22)	2.37*
Antisaccade effect	−.34 (.17)	−.33 (.20)	−.32

* $p < .05$.

Propensity-Score Matching

We used PS Matching in SPSS (Thoemmes, 2012) to perform propensity-score matching. The PS Matching program conducts all analyses in R through the SPSS Essentials for R plugin. We used a logistic regression method to estimate propensity scores and a nonparsimonious approach to include all relevant covariates that have been shown to influence higher order cognitive performance. Thoemmes and Kim (2011) suggest that variables that are theoretically important confounders should be included in the model, regardless of statistical significance (see Beal & Kupzyk, 2014, for similar advice), and therefore we included the covariates of age, gender, immigration status, years of formal education, paternal education, maternal education, household income, fluid intelligence, fitness level, health status, engagement in physical exercise or sports per week, years of musical training, musical proficiency, hours of video game play per week, task motivation, state anxiety, and trait-level PA and NA. A detailed description of these covariates is included in Appendix Table A1. It is considered advisable to exclude variables that are directly related to the treatment variable (age of active bilingualism) as covariates in the calculation of propensity scores (Caliendo & Kopeinig, 2008). Therefore, since onset age of active bilingualism likely influences language-related variables (Luk & Bialystok, 2013), we excluded these (language proficiency, language exposure, and language usage) and did not match early and late bilinguals on them.[1]

After we estimated propensity scores, matching was used as a conditioning strategy. Early active and late active bilinguals who had the closest propensity scores were matched by using the nearest-neighbor matching algorithm. To ensure a good match, we set the caliper width (the maximum distance in propensity scores that two matched units can be apart from each other) to .20. Since we had more early active bilinguals than late active bilinguals, we used a 2:1 matching strategy without replacement; matching with replacement violates the independence-of-cases assumption for typical statistical analyses (Beal & Kupzyk, 2014). The 2:1 matching allowed us to match one late active bilingual with up to two early active bilinguals with similar propensity scores. As demonstrated by Ming and Rosenbaum (2000), 2:1 matching is less biased than 1:1 matching when sample sizes are unequal.

The matching algorithm successfully matched 73 early active bilinguals to 40 late active bilinguals.[2] As shown in Figure 2, covariate balance was improved in the matched sample, especially in terms of immigration status and household income. Also, covariate balance for each covariate was achieved, since none of the covariates had a Cohen's *d* larger than 0.25 (Thoemmes, 2012). Large areas of common support—that is, the degree of overlapping in the distribution of the two bilingual groups—were found in the matched sample (see Figure 3), which supports the generalizability of the sample (Thoemmes & Kim, 2011).

Analysis for Postmatched Samples

ANTI-V (RT). A series of analyses similar to those performed for prematched samples was undertaken for each of the attention network scores in RTs. When executive control network scores were submitted to a repeated-measures mixed factor ANOVA, we found a significant main effect of flanker type, $F(1, 111) = 270.18$, $p < .001$, $\eta_p^2 = .709$, and its interaction with bilingual group, $F(1, 111) = 5.79$, $p = .018$, $\eta_p^2 = .050$ (see Figure 4). However, the main effect of bilingual group, $F(1, 111) = 0.20$, $p = 658$, $\eta_p^2 = .002$, was not significant. Similar to the results of prematched analyses, we did not find significant group differences on either congruent or incongruent trials (Fs < 1), although early active bilinguals ($M = 41$, $SD = 29$) had relatively smaller executive control network scores than late active bilinguals ($M = 55$, $SD = 32$). For orienting network scores, we found a significant main effect of orienting cue, $F(1, 111) = 180.38$, $p < .001$, $\eta_p^2 = .619$. However, neither the main effect of bilingual group, $F(1, 111) = 0.18$, $p = .677$, $\eta_p^2 = .002$, nor the interaction between orienting cue and bilingual group was significant, $F(1, 111) = 0.17$, $p = .683$, $\eta_p^2 = .002$. For alerting network scores, we found a significant main effect of alerting tone, $F(1, 111) = 104.02$, $p < .001$, $\eta_p^2 = .484$, but neither the main effect of bilingual group, $F(1, 111) = 0.26$, $p = .615$, $\eta_p^2 = .002$, nor an interaction between alerting tone and bilingual group was significant, $F(1, 111) = 0.34$, $p = .559$, $\eta_p^2 = .003$. Finally, consistent with prematched analyses, we did not find a global RT advantage for early active bilinguals, $F(1, 111) = 0.21$, $p = .651$, $\eta_p^2 = .002$ (see Table 5).

ANTI-V (accuracy). We conducted further analyses of accuracy data. For executive control network scores in accuracy, we found a significant main effect of flanker type, $F(1, 111) = 8.32$, $p = .005$, $\eta_p^2 = .070$. However, in contrast to prematching results, neither the main effect of bilingual group $F(1, 111) = 0.05$, $p = .823$, $\eta_p^2 = .000$, nor an interaction between flanker type and bilingual goup was significant, $F(1, 111) = 0.47$, $p = .495$, $\eta_p^2 = .004$.

ANTI-V (vigilance). Similarly, we conducted further analyses of indices of vigilance, which indicate aspects of monitoring. We did not find any group differences in hit rates, $F(1, 111) = 0.33$, $p = .568$, $\eta_p^2 = .003$, false alarm rates, $F(1, 111) = 0.48$, $p = .492$, $\eta_p^2 = .004$, sensitivity (d'), $F(1, 111) = 0.25$, $p = .619$, $\eta_p^2 = .002$, or response bias (β), $F(1, 111) = 0.10$, $p = .750$, $\eta_p^2 = .001$.

In sum, we found a significant group difference in executive control network scores (in RT). However, we did not observe early active bilingual advantages over late active bilinguals in global accuracy, accuracy-based executive network scores, sensitivity, or response bias when both early and late bilingual groups were matched on a wide array of nonlinguistic covariates. Similar to prematched analyses, we did not find significant differences between early active and late active bilinguals in terms of orienting and alerting network scores in accuracy.

Saccade tasks. Analyses similar to those performed on prematched samples were undertaken for analysis of RT data (see Table 6). We found a significant main effect of task type, $F(1, 111) = 98.69$, $p < .001$, $\eta_p^2 = .471$. However, neither the main effect of bilingual group, $F(1, 111) = 0.31$, $p = .580$, $\eta_p^2 = .003$, nor the interaction between task type and bilingual group was significant, $F(1, 111) = 0.63$, $p = .428$, $\eta_p^2 = .006$. Also, early active and late active bilinguals did not differ in mean RTs on either the antisaccade or prosaccade tasks.

When accuracy data were submitted to the same analysis, we found significant main effects of task type, $F(1, 111) = 312.22$, $p < .001$, $\eta_p^2 = .738$, and bilingual group, $F(1, 111) = 4.97$, $p = .028$, $\eta_p^2 = .043$; however, the two variables did not interact, $F(1, 111) = 0.14$, $p = .708$, $\eta_p^2 = .001$. Our follow-up analyses showed only marginally significant group differences in prosaccade, $F(1, 111) = 3.76$, $p = .055$, $\eta_p^2 = .033$, and antisaccade conditions, $F(1, 111) = 3.16$, $p = .078$, $\eta_p^2 = .028$.

Importantly, given that our sample includes bilinguals with a wide variety of different language combinations, we limited our analyses to postmatched bilinguals who only spoke English and Mandarin. This yielded a matched sample of 97 bilinguals, which included 63 early active bilinguals and 34 late active bilinguals. Similar to our original results, we found that early active bilinguals outperformed late active bilinguals in executive control network scores ($p < .05$) but in neither

[1] Although PPVT was not matched between the groups, we addressed the imbalance in PPVT by measuring nonverbal IQ scores (assessed by KBIT-2). In doing so, we ensured that early active and late active bilinguals were matched in nonverbal intelligence, and therefore any group differences in PPVT may simply reflect differences in English receptive vocabulary, but not in intelligence.

[2] It is noteworthy that a 2:1 matching ratio may not necessarily result in a sample size with an exact 2:1 ratio of early to late bilinguals, because the matching algorithm applies 1:1 matching when there is only one matched candidate in each group.

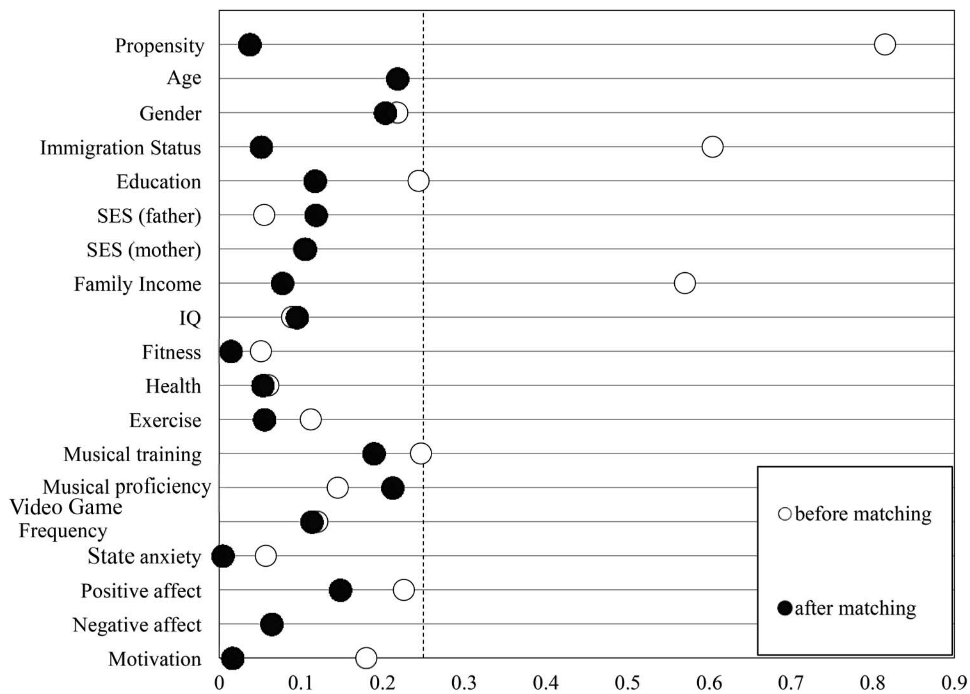

Figure 2. Absolute standard mean differences (Cohen's *d*) of covariates before and after matching. As recommended by Ho, Imai, King, and Stuart (2007), Cohen's *d* was calculated instead of a standard parametric test to avoid bias in balance diagnostics for covariates. Following Austin (2008), Cohen's *d* was based on the absolute weighted standardized mean difference between the two language groups. Imbalance is defined as occurring when Cohen's *d* is larger than .25 (Thoemmes, 2012).

sensitivity (p = .375) nor antisaccade effect (p = .466). The result suggests that variations in language combinations in our bilinguals are less likely to influence our overall findings.

Taken together, when early active and late active bilinguals were matched on nonlinguistic covariates, many of the significant results that were found on the ANTI-V and saccade tasks prior to matching were either attenuated (accuracy on the saccade task) or disappeared (the executive control network in accuracy, global accuracy, sensitivity, and response bias). However, we found a significant effect of early active bilingualism on executive control network in RT.

Correlational Analyses

Before we conducted any correlational analyses using information on age of acquisition, we observed that some bilinguals in our sample did not report their onset age of L1 (i.e., mother tongue) acquisition as zero. This was problematic, because L1 acquisition should logically have started at birth, and an incorrectly identified age of L1 acquisition may have subsequently affected the onset of L2 acquisition. For instance, suppose a simultaneous bilingual has used both languages since birth, but reported his or her ages of L1 and L2 acquisition as 3, while another simultaneous bilingual reported the ages of L1 and L2 acquisition as zero. In the former case, the bilingual's ages of L1 and L2 acquisition were inadvertently shifted to later ages. To resolve this issue, we adjusted for potential variability in age of L2 acquisition by subtracting the age of L2 acquisition from the age of L1 acquisition; thus, the age of L2 acquisition was standardized relative to the age of L1 acquisition. This adjustment is consistent with our conceptualization of the onset age of active bilingualism, which was assessed by an item that states, "I have actively used two languages since birth" on the language background questionnaire.

Correlation among bilingualism duration variables. Given the high prevalence of bilinguals in Singapore—due to its bilingualism policy, which promotes L2 acquisition—most of our bilingual participants had likely been exposed to L2 since birth; they reported that their onset ages of active bilingualism (n = 77) and L2 acquisition (n = 117) were zero (i.e., exposed to both languages since birth). As a result, the indices of bilingualism duration were positively skewed. To deal with violation of the normality assumption, we performed a nonparametric Spearman correlation by operationalizing various indices of bilingualism duration as continuous variables.[3] We found that the onset age of active bilingualism was highly but not perfectly correlated with age of L2 acquisition (r = .53, $p < .001$) and age of L2 fluency (r = .45, $p < .001$), suggesting that despite early acquisition of L2, not all bilinguals have actively used L2 since infancy (see Table 7).

Correlation between bilingualism duration and inhibitory control and monitoring. To roughly compare the predictability of multiple indices of bilingualism duration, we performed Spearman

[3] Although Luk et al. (2011) mitigated the normality assumption by converting zero responses to missing values, this technique may not be feasible for our data set, since 77 bilinguals reported zero as their onset age of active bilingualism (compared with Luk et al.'s 19 participants). Therefore, performing a Spearman correlation was more appropriate, although the result can be more conservative.

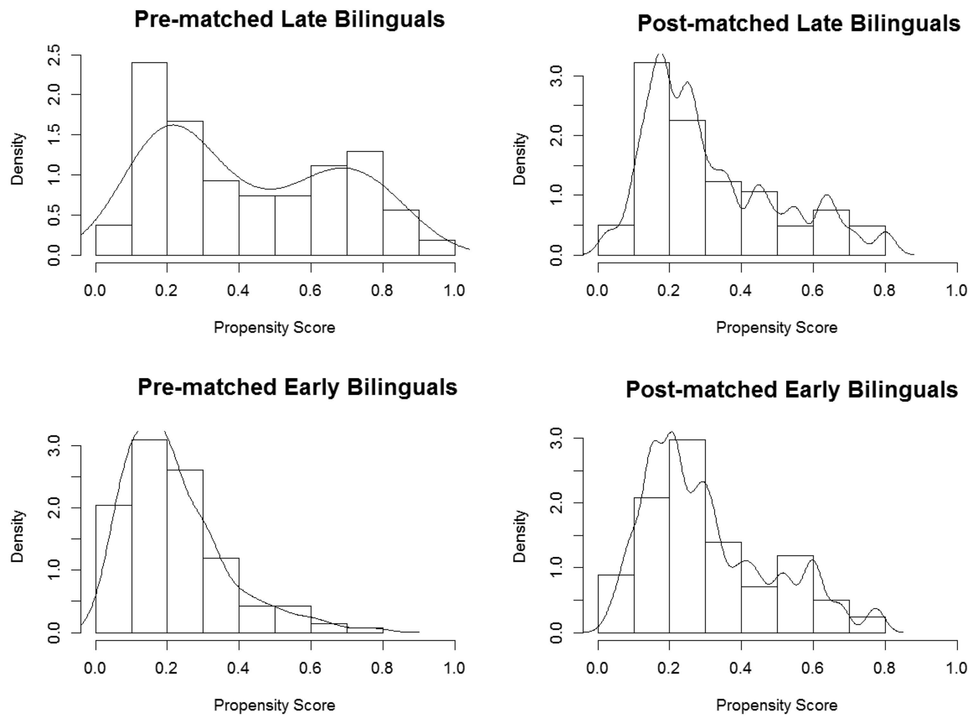

Figure 3. Histogram of propensity scores for early and late bilinguals before and after matching, with an overlaid kernel density estimate. The degree of overlapping in the distribution of early and late bilinguals indicates a large area of common support, which signifies the degree of overlapping in the distribution of the two language groups.

correlation analyses between the three indices of bilingualism duration (onset ages of active bilingualism, L2 acquisition, and L2 fluency) and various performance indices of the ANTI-V and saccade tasks for both prematched and postmatched samples. Following Luk et al. (2011), for correlational analysis in the prematched sample, we included the seven bilinguals who reported their onset age of active bilingualism as 10. These participants had previously been excluded, since they did not fit into either the early- or late-bilingual group; note that our results (see Table 7) do not change, even with inclusion of these participants. Overall, we found that onset age of active bilingualism, age of L2 acquisition, and age of L2 fluency showed similar patterns in predicting various indices of performance on ANTI-V and saccade tasks. Notably, in postmatched samples, the ages of both active bilingualism and L2 acquisition were significantly correlated with executive control network scores and accuracy in prosaccade and antisaccade tasks. However, given that covariates were matched based on the categorical variable of age of active bilingualism (but not on the other two indices of age of L2 acquisition and L2 fluency), the results of this exploratory analysis should be interpreted with caution.

Discussion

We examined the effects of early active bilingualism on inhibitory control and monitoring abilities, as assessed by the ANTI-V and saccade tasks before and after matching early and late bilingual groups on a large number of nonlinguistic covariates, including demographic characteristics, immigration status, fitness level, physical exercise, video game experience, musical training, task motivation, positive and negative affect, and state anxiety (see Appendix Table A1). Before matching, we found that early active bilinguals showed significantly greater executive control network scores in accuracy via their outperformance on incongruent trials; similarly, early active bilinguals showed marginally greater executive control network scores in RT. Early active bilinguals also outperformed late active bilinguals in global accuracy, sensitivity, and response bias measures on the ANTI-V. In terms of saccadic performance, we observed that early active bilinguals outperformed late active bilinguals in accuracy on both the prosaccade and antisaccade trials. Critically, however, when the two language groups' preexisting differences in all of the covariates were balanced via propensity-score matching, the two groups' marginal difference in executive control network scores (in RT) was strengthened. In contrast, previously significant group differences in global accuracy and sensitivity measures on the ANTI-V were not upheld. Similarly, early active bilingual advantages that had been found in accuracy scores for both the antisaccade and prosaccade tasks were also weaker. These results stress the importance of assessing and matching nonlinguistic covariates in studies of bilingual advantages (Valian, 2015). Previous studies have warned of the potential danger of confounding variables that might restrain conclusions about bilingual advantages. Our findings, however, point to another possibility, which has been neglected by many investigators: Failing to match confounding variables may also counteract and veil true bilingual advantages (i.e., the executive network score in RT). This is in line with Bak's (2015) suggestion that confounding variables can either produce spurious effects or mask genuine ones.

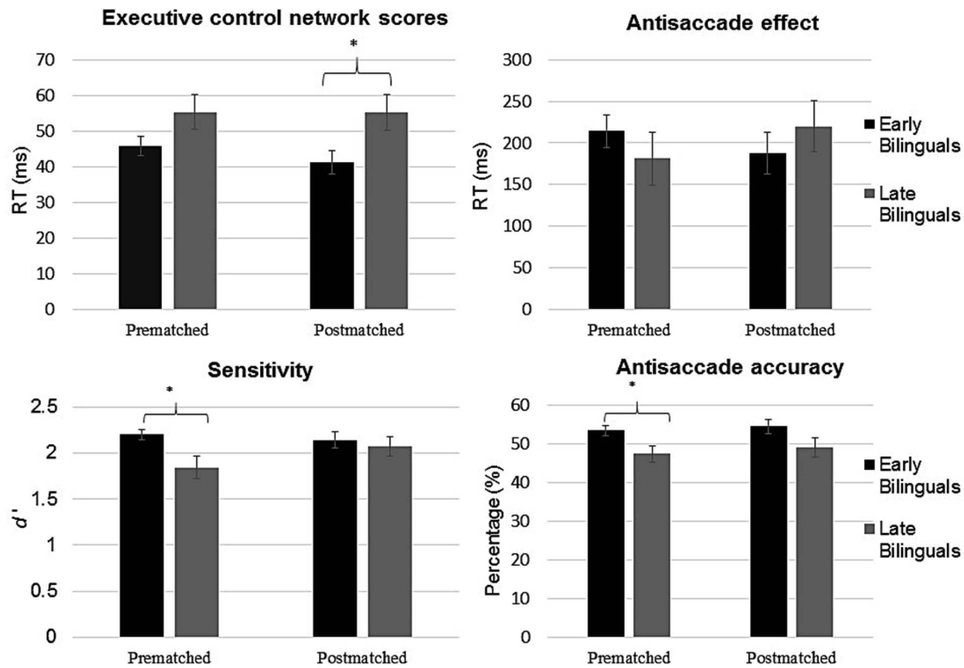

Figure 4. Early active and late active bilinguals' executive control network scores, sensitivity, antisaccade effects, and antisaccade accuracy before and after propensity-score matching. Error bars denote standard error of the mean. * $p < .05$

In terms of early active bilingual advantages for inhibitory control, our results are consistent with the Luk et al. (2011) finding that early active bilinguals had smaller flanker effects than late bilinguals. Notably, this result was still evident when participants were matched on 18 nonlinguistic covariates, which suggests that the earlier onset age of active bilingualism confers benefits on inhibitory control, as assessed by the ANTI-V. In contrast, when propensity-score matching was performed, early active bilinguals showed marginally greater accuracy on both the prosaccade and antisaccade tasks than late active bilinguals, but the two groups did not differ in antisaccade effects. Given this discrepancy on two measures of inhibitory control, our study does not provide evidence of convergent validity (Paap & Greenberg, 2013); hence, it remains unclear whether our finding of early active bilingual advantages on the ANTI-V can be attributed solely to better inhibitory control.

Table 5
Postmatched Bilinguals' Response Latencies and Accuracy on ANTI-V

	RT (ms)			Accuracy		
Condition	Early bilinguals (n = 73) M (SD)	Late bilinguals (n = 40) M (SD)	t	Early bilinguals (n = 73) M (SD)	Late bilinguals (n = 40) M (SD)	t
Global index	677 (107)	687 (93)	−.45	.91 (.11)	.90 (.10)	−.45
Executive control	41 (29)	55 (32)	−2.41*	−.02 (.09)	−.03 (.04)	.68
Congruent	658 (109)	659 (90)	−.09	.91 (.12)	.91 (.10)	−.17
Incongruent	699 (109)	715 (98)	−.78	.90 (.12)	.89 (.10)	.45
Orienting	39 (27)	37 (32)	.41	.00 (.04)	.00 (.04)	.75
Valid	654 (110)	663 (96)	−.46	.91 (.12)	.90 (.09)	.23
Invalid	694 (105)	701 (92)	−.37	.91 (.11)	.90 (.10)	.51
Alerting	47 (47)	42 (37)	.59	−.04 (.06)	−.04 (.06)	.32
Tone (no cue)	662 (113)	675 (100)	−.61	.92 (.10)	.92 (.10)	−.16
No tone (no cue)	708 (113)	716 (98)	−.38	.88 (.13)	.88 (.14)	.21
Sensitivity (d')				2.14 (.77)	2.07 (.66)	.50
Response bias (β)				13.79 (19.95)	12.46 (23.34)	.32
Hits				.52 (.21)	.54 (.19)	−.57
False alarms				.03 (.05)	.04 (.04)	−.69

Note. ANTI-V = Attention Network Test for Interactions and Vigilance; RT = response time.
* $p < .05$.

Table 6
Postmatched Bilinguals' Response Latencies and Accuracy on Saccade Task

Task/effect	Early bilinguals (n = 73) M (SD)	Late bilinguals (n = 40) M (SD)	t
Response latency			
Antisaccade	747 (215)	778 (199)	−.75
Prosaccade	560 (116)	558 (127)	.06
Antisaccade effect	187 (116)	220 (193)	−.80
Accuracy			
Antisaccade	.55 (.16)	.49 (.16)	1.78†
Prosaccade	.88 (.15)	.81 (.23)	1.94†
Antisaccade effect	−.33 (.18)	−.32 (.20)	−.38

† $p < .10$.

Given that a saccade task has been shown to assess prepotent response inhibition—while a flanker task assesses resistance to interference with distractors (Friedman & Miyake, 2004)—it is possible that early active bilingual advantages may be attributed relatively more to inhibitory control for interference from distractors but less to the suppression of prepotent responses. Considering that early active bilingual advantages in antisaccadic accuracy were weaker after propensity-score matching, further research is warranted to examine the generalizability of our findings to other measures of inhibitory control. Moreover, it is critical that future research endeavor aims to examine the specific mechanism that underlies early active bilingual advantages in inhibitory control by comparing them with a matched monolingual sample.

It is notable that although we found significant group differences in executive network scores (in RT) between postmatched early active and late active bilinguals, further analyses showed that early active bilinguals (M = 699, SD = 109) did not significantly differ from late active bilinguals (M = 715, SD = 98) on incongruent trials, (p = .440). Although Paap et al. (2015) argued that bilingual advantages in inhibitory control should be substantiated by their significant outperformance on incongruent trials—which demand inhibitory control—this claim should be interpreted with caution. It is noteworthy that the measure of RT on purely incongruent trials is subject to various influential processes that implicate not only inhibitory control, but also motor speed, motivation, response strategies, and so on (Paap & Greenberg, 2013). These noncore, secondary processes may explain, in part, the presence of huge variability (i.e., standard deviation) in RT, which in turn increased error terms and contributed to insignificant differences when early active and late active bilinguals' mean RTs for incongruent trials were compared. Indeed, when systematic errors caused by noncore secondary processes were partialed out by controlling for global RT, we found a significant group difference between early active and late active bilinguals after matching (p = .028). These results are similar to those from our analyses of executive network scores, which were also computed by subtracting RTs on congruent trials from those on incongruent trials (for further discussion, see Linck, 2015).

With regard to monitoring, we failed to detect significant differences between early active and late active bilinguals on various measures of monitoring (sensitivity, global RT, and global accu-

Table 7
Correlations Between Variables of Bilingualism Duration and ANTI-V

Bilingualism duration	ANTI-V									Saccade task					
	EC (RT)	Alert (RT)	Orient (RT)	Global RT	Global ACC	Hits	False alarm	d'	β	Anti ACC	Pro ACC	Anti RT	Pro RT	Anti effect (RT)	Anti effect (ACC)
Prematched sample															
Age of active bilingualism (n = 203)	.117†	−.107	−.077	.051	−.191*	−.043	.195*	−.166*	−.241*	−.183*	−.170*	−.053	−.006	−.021	.031
Age of L2 acquisition (n = 202)	.130†	−.108	−.134†	−.030	−.085	−.048	.123†	−.123†	−.164*	−.257*	−.191*	−.015	−.024	−.034	−.015
Age of L2 fluency (n = 200)	.149*	−.044	−.094	−.036	−.120	−.047	.052	−.082	−.111	−.063	−.038	−.044	.008	−.052	.063
Postmatched sample															
Age of active bilingualism (n = 113)	.231*	−.090	−.094	.062	−.165†	.059	.216*	−.086	−.214*	−.203*	−.117	.069	−.067	.147	−.021
Age of L2 acquisition (n = 112)	.187*	−.136	−.223*	.029	−.085	.008	.149	−.090	−.123	−.279*	−.097	.083	−.048	.138	−.154
Age of L2 fluency (n = 111)	.131	−.029	−.088	−.087	−.160	−.032	−.016	−.059	−.060	−.127	−.061	.050	−.017	.042	−.007

Note. ANTI-V = Attention Network Test for Interactions and Vigilance; RT = response time; d' = sensitivity, β = response bias.
† $p < .10$. * $p < .05$.

racy of ANTI-V), especially when preexisting differences in covariates were matched. These findings conflict with previous studies that have demonstrated a link between bilingualism and monitoring performance (Kapa & Colombo, 2013; Paap et al., 2014), which suggests that early bilingual advantages in monitoring may arise from preexisting differences in nonlinguistic covariates between early and late bilinguals. Alternatively, it is possible that the null effect of early bilingualism on monitoring could be due to the possibility that inhibitory control and monitoring abilities are not similarly malleable to bilingual training, and instead may require varying intensity and duration of training for improvement. In support of this notion, some evidence suggests that training-associated improvement in resisting interference to distractors does indeed differ from the abilities required for monitoring. For instance, Strobach, Liepelt, and Schubert (2012) found that mixing cost (which is an index of monitoring) was eliminated on two different variants of the task-switching paradigm after eight practice sessions of about 7,000 trials. However, flanker effects were not entirely eliminated, even after ten 1-hr practice sessions (Ishigami & Klein, 2010). Several training studies have also failed to find any short-term practice effect on flanker effect, despite improved speeds for processing information on both congruent and incongruent trials (Brown & Fera, 1994, Experiment 3; Chen, Tang, & Chen, 2013; and Lin, 2010; however, see Costa et al., 2009 and Stasenko, Matt, & Gollan, 2017 for evidence that bilingual advantages can diminish with practice). Given the relative malleability of monitoring, as opposed to that of inhibitory control, it is possible that bilingualism, compared with monolingualism, readily confers benefits on monitoring performance, regardless of early- or late-active bilingualism. In this regard, although we did not test monolinguals, we expect that both early and late bilinguals likely outperform monolinguals in monitoring. However, more extensive training in early active bilingualism may be required to produce notable improvement in resisting interference from distractors—in our case, the executive attention network score (i.e., flanker effect). Hence, we expect that bilingual advantages in resisting interference from distractors would be more likely in early active bilinguals than either late active bilinguals or monolinguals. This would explain why Luk et al. (2011) found early active bilingual advantages in flanker effects, but not in global RT, which is a simple index of global monitoring.

Although we found early active bilingual advantages in inhibitory control, our findings should be interpreted with caution, especially when being generalized to a comparison with monolinguals or different aspects of inhibitory control. Future studies should, therefore, replicate our findings with other measures of inhibitory control and a matched monolingual sample. Future studies should also consider using tasks with greater working memory demand, which may produce stronger results (Costa et al., 2009; Jiao et al., 2017), and using more objective measures of L1 and L2 to overcome the potential limitations of self-reported measure of linguistic proficiency (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2011). Moreover, given that bilinguals' interactional contexts of conversational exchanges moderate bilingual advantages in task switching (e.g., Hartanto & Yang, 2016a), future research should also examine whether bilinguals' interaction contexts would qualify the effect of early active bilingualism on other aspects of executive function (e.g., inhibitory control). Although we argue that bilinguals' interactional contexts and early active bilingualism do not necessarily entail similar advantages in all aspects of executive functions, investigating bilinguals' interactional context and their complex experiences in conjunction with early active bilingualism will shed further light on the mechanism underlying bilingual advantages in executive functions.

In contrast to the Luk et al. (2011) results, we did not find strong evidence to suggest that the onset age of active bilingualism is more sensitive than other indices of bilingualism duration (i.e., age of L2 acquisition and age of L2 fluency) for predicting performance on tasks of inhibitory control; for instance, our study suggests that the onset age of L2 acquisition can still be used as an index of bilingualism duration. It is noteworthy, however, that the correlation between onset age of active bilingualism and that of L2 acquisition was still not perfect ($r = .53$, $p < .001$), suggesting that onset ages of active bilingualism and L2 acquisition should not be used interchangeably to predict performance on measures of cognitive control.

Finally, our use of the propensity-score matching method is noteworthy, as it tackles the inherent difficulty in controlling for a wide range of confounding factors in bilingualism research, which have been shown to influence inhibitory control and monitoring (Paap et al., 2015; Valian, 2015). Some may argue that propensity-score matching reduces statistical power due to the removal of unmatched participants; this is not necessarily true, however, because of the improved covariate balance (see Smith, 1997, for a demonstration). Moreover, if propensity-score matching drastically reduces sample size, other conditioning strategies are available to mitigate the sample-size issue (see Austin, 2011, for other propensity-scoring techniques). Therefore, we believe that using the propensity-scoring matching method, which substantially enhances equivalence between language groups, can be an important method for resolving ongoing debates in the field about confounding factors and strengthening causal inferences regarding bilingualism's overall effect on cognitive control.

References

Abutalebi, J., Guidi, L., Borsa, V., Canini, M., Della Rosa, P. A., Parris, B. A., & Weekes, B. S. (2015). Bilingualism provides a neural reserve for aging populations. *Neuropsychologia, 69,* 201–210. http://dx.doi.org/10.1016/j.neuropsychologia.2015.01.040

Austin, P. C. (2008). Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. *Pharmacoepidemiology and Drug Safety, 17,* 1218–1225. http://dx.doi.org/10.1002/pds.1674

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399–424. http://dx.doi.org/10.1080/00273171.2011.568786

Bak, T. H. (2015). Beyond a simple "yes" and "no." *Cortex, 73,* 332–333. http://dx.doi.org/10.1016/j.cortex.2015.08.003

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), e0153485. http://dx.doi.org/10.1371/journal.pone.0153485

Beal, S. J., & Kupzyk, K. A. (2014). An introduction to propensity scores what, when, and how. *The Journal of Early Adolescence, 34,* 66–92. http://dx.doi.org/10.1177/0272431613503215

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental, 60,* 68–79. http://dx.doi.org/10.1037/cjep2006008

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,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.32.6.1341

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

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

Brown, P., & Fera, P. (1994). Turning selective attention failure into selective attention success. *Canadian Journal of Experimental Psychology, 48,* 25–57. http://dx.doi.org/10.1037/1196-1961.48.1.25

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22,* 31–72. http://dx.doi.org/10.1111/j.1467-6419.2007.00527.x

Callejas, A., Lupiáñez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain and Cognition, 54,* 225–227. http://dx.doi.org/10.1016/j.bandc.2004.02.012

Chen, A., Tang, D., & Chen, X. (2013). Training reveals the sources of Stroop and Flanker interference effects. *PLoS ONE, 8*(10), e76580. http://dx.doi.org/10.1371/journal.pone.0076580

Chen, C. C., & Tyler, C. W. (2008). Excitatory and inhibitory interaction fields of flankers revealed by contrast-masking functions. *Journal of Vision (Charlottesville, Va.), 8,* 1–14. http://dx.doi.org/10.1167/8.4.10

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science, 14,* 125–130. http://dx.doi.org/10.1111/1467-9280.t01-1-01430

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

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

Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science, 21,* 335–341. http://dx.doi.org/10.1177/0963721412453722

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. *Experimental Psychology, 61,* 234–251. http://dx.doi.org/10.1027/1618-3169/a000243

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York, NY: Elsevier.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6,* 383–401. http://dx.doi.org/10.3102/00028312006003383

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336–353. http://dx.doi.org/10.1037/1528-3542.7.2.336

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

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://dx.doi.org/10.3389/fpsyg.2014.00011

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2011). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition, 15,* 594–615. http://dx.doi.org/10.1017/S1366728911000332

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

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

Hartanto, A., Toh, W. X., & Yang, H. (2016). Age matters: The effect of onset age of video game play on task-switching abilities. *Attention, Perception & Psychophysics, 78,* 1125–1136. http://dx.doi.org/10.3758/s13414-016-1068-9

Hartanto, A., Toh, W. X., & Yang, H. (2018). Bilingualism narrows socioeconomic disparities in executive functions and self-regulatory behaviors during early childhood: Evidence from the Early Childhood Longitudinal Study. *Child Development*. Advance online publication. http://dx.doi.org/10.1111/cdev.13032

Hartanto, A., & Yang, H. (2016a). 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://dx.doi.org/10.1016/j.cognition.2016.01.016

Hartanto, A., & Yang, H. (2016b). Is the smartphone a smart choice? The effect of smartphone separation on executive functions. *Computers in Human Behavior, 64,* 329–336. http://dx.doi.org/10.1016/j.chb.2016.07.002

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,* 625–658. http://dx.doi.org/10.3758/s13423-011-0116-7

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15,* 199–236. http://dx.doi.org/10.1093/pan/mpl013

Humphrey, A. D., & Valian, V. V. (2012, November). *Multilingualism and cognitive control: Simon and flanker task performance in monolingual and multilingual young adults*. Paper presented at the 53rd Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology, 43,* 302–313. http://dx.doi.org/10.1111/j.1469-8986.2006.00403.x

Ishigami, Y., & Klein, R. M. (2010). Repeated measurement of the components of attention using two versions of the Attention Network Test (ANT): Stability, isolability, robustness, and reliability. *Journal of Neuroscience Methods, 190,* 117–128. http://dx.doi.org/10.1016/j.jneumeth.2010.04.019

Jiao, L., Liu, C., Wang, R., & Chen, B. (2017). Working memory demand of a task modulates bilingual advantage in executive functions. *International Journal of Bilingualism*. Advance online publication. http://dx.doi.org/10.1177/1367006917709097

Kalia, V., Wilbourn, M. P., & Ghio, K. (2014). Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology, 26,* 699–713. http://dx.doi.org/10.1080/20445911.2014.956748

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183. http://dx.doi.org/10.1037/0096-3445.130.2.169

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development, 28,* 233–246. http://dx.doi.org/10.1016/j.cogdev.2013.01.011

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaushanskaya, M., & Prior, A. (2015). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition, 18,* 27–28. http://dx.doi.org/10.1017/S1366728914000510

Lin, Y. H. (2010). *Flanker interference in younger and older adults: Does training influence focusing of attention?* (Unpublished master's thesis). Department of Psychology, University of Waterloo. Waterloo, Ontario, Canada.

Linck, J. A. (2015). Methods matter for critical reviews too. *Cortex, 73,* 354–355. http://dx.doi.org/10.1016/j.cortex.2015.07.011

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25,* 605–621. http://dx.doi.org/10.1080/20445911.2013.795574

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,* 588–595. http://dx.doi.org/10.1017/S1366728911000010

Macnamara, B. N., & Conway, A. R. (2014). Novel evidence in support of the bilingual advantage: Influences of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review, 21,* 520–525. http://dx.doi.org/10.3758/s13423-013-0524-y

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940–967. http://dx.doi.org/10.1044/1092-4388(2007/067)

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology, 31*(Part 3), 301–306. http://dx.doi.org/10.1111/j.2044-8260.1992.tb00997.x

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40–48. http://dx.doi.org/10.1037/0021-843X.110.1.40

Ming, K., & Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics, 56,* 118–124. http://dx.doi.org/10.1111/j.0006-341X.2000.00118.x

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22,* 1425–1433. http://dx.doi.org/10.1177/0956797611416999

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures, 5,* 297–310. http://dx.doi.org/10.1080/19312458.2011.624490

Paap, K. R., Darrow, J., Dalibar, C., & Johnson, H. A. (2015). Effects of script similarity on bilingual advantages in executive control are likely to be negligible or null. *Frontiers in Psychology, 5,* 1539. http://dx.doi.org/10.3389/fpsyg.2014.01539

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66,* 232–258. http://dx.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,* 615–639. http://dx.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://dx.doi.org/10.1016/j.cortex.2015.04.014

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, 274,* 81–93. http://dx.doi.org/10.1016/j.jneumeth.2016.10.002

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

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13,* 160–166. http://dx.doi.org/10.1016/j.tics.2009.01.006

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. http://dx.doi.org/10.1016/S0028-3932(97)00015-8

Roca, J., Castro, C., López-Ramón, M. F., & Lupiáñez, J. (2011). Measuring vigilance while assessing the functioning of the three attentional networks: The ANTI-Vigilance task. *Journal of Neuroscience Methods, 198,* 312–324. http://dx.doi.org/10.1016/j.jneumeth.2011.04.014

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55. http://dx.doi.org/10.1093/biomet/70.1.41

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin, 117,* 230–249. http://dx.doi.org/10.1037/0033-2909.117.2.230

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103,* 1334–1344. http://dx.doi.org/10.1198/016214508000000733

Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology, 27,* 325–353. http://dx.doi.org/10.1111/1467-9531.271030

Sorge, G. B., Toplak, M. E., & Bialystok, E. (2017). Interactions between levels of attention ability and levels of bilingualism in children's executive functioning. *Developmental Science, 20*(1), e12408. http://dx.doi.org/10.1111/desc.12408

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, 25,* 155–165. http://dx.doi.org/10.1080/23279095.2016.1263795

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31,* 137–149. http://dx.doi.org/10.3758/BF03207704

Stasenko, A., Matt, G. E., & Gollan, T. H. (2017). A relative bilingual advantage in switching with preparation: Nuanced explorations of the proposed association between bilingualism and task switching. *Journal of Experimental Psychology: General, 146,* 1527–1550. http://dx.doi.org/10.1037/xge0000340

Strobach, T., Liepelt, R., Schubert, T., & Kiesel, A. (2012). Task switching: Effects of practice on switch and mixing costs. *Psychological Research, 76,* 74–83. http://dx.doi.org/10.1007/s00426-011-0323-x

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,* 123. http://dx.doi.org/10.3389/fpsyg.2011.00123

Thoemmes, F. J. (2012). *Propensity score matching in SPSS*. Retrieved from arXiv:1201.6385

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46,* 90–118. http://dx.doi.org/10.1080/00273171.2011.540475

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38,* 227–242. http://dx.doi.org/10.1177/0022022106297301

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language, 67,* 1–16. http://dx.doi.org/10.1016/j.jml.2011.12.005

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence, 38,* 111–122. http://dx.doi.org/10.1016/j.intell.2009.08.002

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62,* 392–406. http://dx.doi.org/10.1016/j.jml.2010.02.001

Unsworth, N., Spillers, G. J., Brewer, G. A., & McMillan, B. (2011). Attention control and the antisaccade task: A response time distribution analysis. *Acta Psychologica, 137,* 90–100. http://dx.doi.org/10.1016/j.actpsy.2011.03.004

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

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey. *Medical Care, 34,* 220–233. http://dx.doi.org/10.1097/00005650-199603000-00003

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,* 1579–1586. http://dx.doi.org/10.1037/xlm0000107

Woumans, E., & Duyck, W. (2015). The bilingual advantage debate: Moving toward different methods for verifying its existence. *Cortex, 73,* 356–357. http://dx.doi.org/10.1016/j.cortex.2015.07.012

Yang, H., Hartanto, A., & Yang, S. (2016). The importance of bilingual experience in assessing bilingual advantages in executive functions. *Cortex, 75,* 237–240. http://dx.doi.org/10.1016/j.cortex.2015.11.018

Yang, H., Yang, S., Ceci, S. J., & Isen, A. M. (2015). Positive affect facilitates the effect of a warning on false memory in the DRM paradigm. *The Journal of Positive Psychology, 10,* 196–206. http://dx.doi.org/10.1080/17439760.2014.950177

Yang, H., Yang, S., & Isen, A. M. (2013). Positive affect improves working memory: Implications for controlled cognitive processing. *Cognition and Emotion, 27,* 474–482. http://dx.doi.org/10.1080/02699931.2012.713325

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://dx.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,* 412–422. http://dx.doi.org/10.1017/s1366728910000611

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology, 6,* 164. http://dx.doi.org/10.3389/fpsyg.2015.00164

Appendix

Table A1
Description of Covariates

Covariate	Description
Age	Self-reported chronological age
Gender	Self-reported gender (female = 0, male = 1)
Immigration status	Participant's country of origin. Participants not originally from Singapore were considered to be immigrants (Singapore = 0, country other than Singapore = 1)
Formal education	Number of years of formal education, starting from primary school
Paternal education level	Highest educational level obtained by male caregiver (father), rated on a scale of 1 to 6 (1 = *none*, 2 = *primary school*, 3 = *secondary school*, 4 = *high school or equivalent*, 5 = *bachelor's degree*, 6 = *master's or PhD*)
Maternal education level	Highest educational level obtained by female caregiver (mother), rated on a scale of 1 to 6 (1 = *none*, 2 = *primary school*, 3 = *secondary school*, 4 = *high school or equivalent*, 5 = *bachelor's degree*, 6 = *master's or PhD*)
Monthly household income	Combined income of all people sharing a household, rated on a scale of 1 (less than S$2,500) to 9 (more than S$20,000), with intervals of S$2,500
Fluid intelligence	Standardized nonverbal intelligence score measured using KBIT-2 (Kaufman & Kaufman, 2004)
Health status	Participant's perceived current overall health status; rated on a scale of 1 to 5 (1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*). Item was adapted from the Health Status Questionnaire (HSQ-12; Ware, Kosinski, & Keller, 1996)
Fitness level	Participant's current physical fitness, rated on a scale of 1 to 5 (1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*)
Hours of engaging in physical exercise per week	Number of hours (per week) spent engaging in physical exercise or sports for the last 6 months
Hours of playing video games per week	Number of hours (per week) spent engaging in video games for the last 6 months
Years of musical training	Number of years of formal musical training in any instrument(s)
Musical proficiency	Self-rated musical proficiency; rated on a scale of 1 to 5 (1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*)
State anxiety	Participants' state anxiety immediately before beginning the executive control tasks, measured by a validated state version of the State-Trait Anxiety Inventory (Marteau & Bekker, 1992)
Task motivation	Participant's motivation immediately before beginning the executive control tasks; rated on a scale of 1 to 5 (*1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = very good*)
Trait positive affect	Participant's trait-level positive affect, measured by a validated trait version of the International Positive and Negative Affect Schedule Short Form (Thompson, 2007)
Trait negative affect	Participant's trait-level negative affect, measured by a validated trait version of the International Positive and Negative Affect Schedule Short Form (Thompson, 2007)