Current Research in Behavioral Sciences 3 (2022) 100066

Contents lists available at ScienceDirect

Current Research in Behavioral Sciences

journal homepage: www.elsevier.com/locate/crbeha

Task-specific bilingual effects in Mandarin-English speaking high school students in China

Adam John Privitera[a,b,*], Mohammad Momenian[c], Brendan Weekes[b,d,e]

[a] College of Liberal Arts, Wenzhou-Kean University, 88 Daxue Road, Wenzhou, China
[b] Faculty of Education, University of Hong Kong, Pokfulam Road, Hong Kong
[c] Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hung Hom, Hong Kong
[d] Department of Psychology, University of Cambridge, Downing St, Cambridge CB2 3 EB, United Kingdom
[e] School of Psychology, University of Melbourne, 12th floor Redmond Barry Building, Parkville Campus, Melbourne, VIC 3010, Australia

ARTICLE INFO

Keywords:
Bilingual advantage
Executive function
Inhibitory control
Simon task
Flanker task
Mixed-effects modeling

ABSTRACT

The question of whether bilingual language experience confers a cognitive advantage is still open. Studies report that putative bilingual advantages can be accounted for by individual differences in socioeconomic class, immigration status, or culture. Such studies typically consider bilingual experience to be a categorical variable using parametric statistical analyses. However, bilingual experience is itself highly variable across individual participants in most studies reported to date. Here we test the hypothesis that bilingual experience has a direct effect on executive function by estimating the effect of L2 (English) experience on performance in the Simon and flanker tasks. Linear mixed-effects models were used to assess effects of bilingual experience on performance. Self-reported L2 proficiency was associated with reduced interference on the Simon task as well as faster global response times on the flanker task, suggesting some cognitive advantages during inhibitory control. We conclude that individual differences in bilingual language experience may explain the many contradictory findings in studies testing the veracity of the bilingual advantage.

1. Introduction

Speaking more than one language is believed to deliver benefits in multiple domains (Diamond, 2010). However, the hypothesis that bilingualism is associated with enhanced cognitive control is controversial (Antoniou, 2019; Paap et al., 2020). Cognitive control is typically measured with paradigms such as the Simon task (Simon and Rudell, 1967) and flanker task (Eriksen and Eriksen, 1974), both having in common a requirement to inhibit a prepotent response to task irrelevant stimulus features - such as color, spatial location, or orientation – when one feature is relevant to task response. The bilingual advantage hypothesis is controversial because it is argued that the evidence is flawed by limitations including sampling errors, measurement issues, and straightforward biases (De Bruin et al., 2015).

The bilingual advantage is linked to behavioral (e.g., Lowe et al., 2021; Ware et al., 2020) and neural evidence (e.g., Tao et al., 2021). However, the focus of the present study is on tasks that are reported to show a behavioral advantage (Simon and flanker). Some bilingual speakers show reduction in interference on these tasks resulting in a faster reaction time (RT) overall compared with monolingual peers. Reduced interference is assumed to reflect a better capacity to inhibit competing stimulus features and this ability is in turn considered to support the bilingual advantage hypothesis. One prediction of this hypothesis is faster RT on incongruent trials (with competing alternative responses), and another is faster RT for overall processing, manifesting as faster RTs across all trials (Hilchey and Klein, 2011). Most bilingual advantages are assumed to result from an (undisputed) requirement to inhibit a non-target language during bilingual language processing (Green, 1998), or an increased demand with conflict monitoring necessary in bilingual discourse (Hilchey and Klein, 2011). Despite evidence for these assumptions, the evidence for putative benefits, if observed at all, are considered to be artefactual and a result of design flaws by several authors (Lowe et al., 2021; Paap, 2019; Ware et al., 2020).

It is correct to assert that previous studies of the bilingual advantage are characterized by design flaws. Arguably the most troubling is control over bilingual language experience using a binary categorical variable classification (Luk and Bialystok, 2013). This approach does not capture evidence of variability in bilingual experience including within the participants reported in these previous studies. This leaves open the question whether bilingual language experience itself has any significant effect on inhibitory control independent of other correlated individual differences. Furthermore, binary classification diminishes natural

* Corresponding author at: College of Liberal Arts, Wenzhou-Kean University, 88 Daxue Road, Wenzhou, China.
E-mail addresses: aprivite@qq.com (A.J. Privitera), mohammad.momenian@polyu.edu.hk (M. Momenian), weekes@hku.hk (B. Weekes).

https://doi.org/10.1016/j.crbeha.2022.100066
Received 9 November 2021; Received in revised form 5 December 2021; Accepted 12 December 2021

variability in language experience, ignoring a rich source of extra data to test the bilingual advantage hypothesis. A different approach is to abandon bilingual versus monolingual classification in favor of within-group analyses of individual differences. Although rare, this approach is becoming more common (e.g., Soveri et al., 2011; Xie, 2018; Xie and Zhou, 2020). We contend, however, that those studies are relatively weak tests of the bilingual advantage hypothesis because they assume a parametric distribution of individual differences. Although necessary for traditional statistical models, such models are not rigorous tests of the hypothesis. One alternative is Linear Mixed-Effects Models (LMEM). Our goal is to test the effects of individual differences in language experience with LMEM.

In addition to our main goal, another aim of the study is to address some of the methodological weaknesses found in previous studies. One criticism of these studies is poor control over correlated variables: socioeconomic status (SES), culture, education, and immigration status, that necessarily differ between monolingual and bilingual groups (Paap et al., 2015) *and* also vary across bilingual samples. Inconsistent capture and reporting of these variables is troubling in light of the evidence that the bilingual advantage in executive function is malleable (i.e., reduced, eliminated, or reversed; a bilingual disadvantage) when such variables are manipulated (Van den Noort et al., 2019).

It is obvious that individual differences in language experience (age of acquisition, dominance, proficiency, usage domain) impact on performance (Donnelly et al., 2019; Tao et al., 2021; Van den Noort et al., 2019; Ware et al., 2020). However, these individual differences in language experience, as well as differences in correlated variables and baseline task performance, are ignored by the methodological practices employed in most previous investigations. These common practices prevent a complete analysis of bilingual experience as well as the advantage hypothesis itself (De Bruin, 2019; Incera and McLennan, 2018; Luk and Bialystok, 2013). Even a statistical critique highlights co-linearity and lack of power from binary variables in linear regression models (Weekes et al., 1998), and correlated continuous variables often explain more variance, thus questioning the veracity of the bilingual advantage (Mindt et al., 2008; Paap et al., 2015). Following the reasoning of Linck and Cunnings (2015), we expected that individual differences in language experience would predict any bilingual advantages found on measures of inhibitory control. We tested this prediction in a sample of culturally, linguistically, and socioeconomically homogenous bilingual (i.e., Mandarin-English speaking) participants residing in Shenzhen, People's Republic of China.

2. Methods

2.1. Participants

Participants were recruited from a private, international high school in Shenzhen via official school email and on-campus informational sessions. Participating students (n = 41, 31 females; M_{age} = 16.21 years, SD_{age} = 1.61 years) were all native Mandarin speakers (L1) enrolled on a full-time basis in an academic program using English (L2) as the primary medium of instruction. Participants received community service hours credited toward graduation. Written informed consent was collected from all participants. For participants below the age of 18, informed consent was granted by parents or legal guardians. Approval for this study was granted by the Human Research Ethics Committee of the University of Hong Kong (#EA200010).

Demographic and language history data are summarized in Table 1. Participants, on average had over 10 years of experience using English (M = 10.49, SD = 2.03). In addition to Mandarin and English, participants reported speaking Cantonese (n = 7), Spanish (n = 4), French (n = 2), and Japanese (n = 2). All participants reported higher levels of proficiency, on average, for L1 compared to L2. It is interesting to note that, while all participants self-reported as native Mandarin speakers, the range of L1 proficiency reported was similar to that of L2. This may be due, in part, to the experience of participants in an English immersive school environment, but may also reflect demand characteristics given the motivation of students to speak Mandarin. The sample may therefore be considered multilingual given the mother tongue could be different to officially used Mandarin (i.e., regional dialect).

Table 1
Demographic and language history data.

	M	*SD*	Range
Age (years)	16.21	1.61	13 - 19
Socioeconomic status (1–4 points)	2.43	0.70	1 - 4
PSS-10 score (0–40 points)	18.97	5.66	10 - 32
Weekly video game time (hours)	10.09	13.04	0 - 70
Weekly musical instrument time (hours)	1.54	2.25	0 - 9
Number of languages used	2.36	0.57	2 - 4
Frequency of language switching (1–7 points)	4.84	1.74	1 - 7
L2 experience (years)	10.49	2.03	5 - 15
L1 proficiency (0–1 point)	0.89	0.10	0.57 - 1
L1 dominance (0–1 point)	0.56	0.10	0.4 - 0.94
L2 immersion (0–1 point)	0.64	0.11	0.39 - 0.86
L2 proficiency (0–1 point)	0.72	0.12	0.5 - 1
L2 dominance (0–1 point)	0.43	0.08	0.28 - 0.59
L2/L1 ratio of dominance	0.79	0.15	0.46 - 1.23

2.2. Materials

2.2.1. Language history questionnaire

The Language History Questionnaire (LHQ-3) (Li et al., 2020) was administered to all participants in English online via the Gorilla online experiment builder (Anwyl-Irvine et al., 2020). Self-report measures of bilingual language experience including the LHQ-3 are correlated with objective measures (Gollan et al., 2011; Grant and Li, 2019; Li et al., 2020; Marian et al., 2007). The LHQ-3 was selected due to its wide usage as a self-report measure of language experience, and because it allows for the calculation of scores for language proficiency, immersion, and dominance for each language used as well as different modalities. Results for each of these aggregated scores reflect a different aspect of language experience as a continuous variable as opposed to the more typically used categorical classification of bilingualism (Luk and Bialystok, 2013). Data from the LHQ-3 was limited to aggregate scores for proficiency, immersion, dominance, and dominance ratio between L2 and L1 scores (L2/L1 dominance ratio). Given our focus, the LHQ-3 was limited to Mandarin and English. Although information about other languages was not collected, participants were asked to indicate which other languages they used, allowing us record the total number of languages spoken per participant. Additional questions were added to estimate language switching experience (Bhandari et al., 2020), and number of hours playing video games (Bialystok, 2006) and musical instruments (Jentzsch et al., 2014) each week. Finally, although we assumed our sample was drawn from a socioeconomically similar population, this was tested with family education level measured as proxy for SES (Wermelinger et al., 2017).

2.2.2. Performance tasks

Simon and flanker effects were elicited remotely using the Gorilla online experiment builder (Anwyl-Irvine et al., 2020). Given restrictions associated with in-person data collection in the Covid-19 pandemic, Internet-based platforms have become necessary. Comparisons between virtual and laboratory conditions are lacking. However, studies show that established behavioral effects can be replicated using an Internet platform (Anwyl-Irvine et al., 2020; Crump et al., 2013; Jylkkä et al., 2017).

2.2.3. Simon task

A two-color Simon task (Bialystok et al., 2004; Schroeder et al., 2016) was administered to all participants. Each trial began with presentation of a fixation cross (black; 2.54 cm line; 0.254 cm thick) in the center of a white background for 300 ms. Next, the fixation cross vanished and a target stimulus appeared on the white background. Depending on the trial, a blue or brown square (2.54 × 2.54 cm) was presented in one of 3 locations: center, left, or right, relative to the location of a central fixation cross. Using a computer keyboard, participants were instructed to press either the Q key with their left index finger, or the P key with their right index finger based on the color of the stimulus, irrespective of location. Considering the color of the stimulus, and the mapping of color to response key, one of three possible trial conditions resulted: congruent (match between color and hand used to respond), incongruent (mismatch between color and response hand), or neutral (stimulus in the center, random presentation of either color). Stimuli remained onscreen until a response was recorded via button press. A blank screen was presented for 500 ms after each response prior to the start of the next trial. Prior to the start of the experimental block, 6 practice trials were completed which included an example of each of the possible color (blue or brown) and location (center, left, or right) combinations used. Feedback was provided after each response during practice trials only. Each of the 6 possible color and location combinations was presented 7 times during the experimental block for a total of 42 trials. The experimental block contained equal numbers of congruent, incongruent, and neutral trials (14 for each condition). Trial presentation in both the practice and experimental blocks was randomized.

2.2.4. Flanker task

Participants completed an online version of the Attention Network Test (ANT; Fan et al., 2002). The task was split into three separate phases: (1) a no-cue practice phase; and, (2) a cued practice phase to familiarize participants with the format of the task and the different kind of trials they will encounter; and, (3) a testing phase containing both no-cue and cued trials. The phase order (i.e., no-cue practice, cued practice, testing) was identical for all participants. Prior to the start of a practice phase, participants were instructed to place their left index finger on the Q key and their right index finger on the P key of their computer keyboard and to focus on the fixation cross during the entire task (i.e., not to move their eyes to the target). A reminder of the stimuli-response mapping remained visible at the top of the screen during both practice phases. Only no-cue trials were included in our analyses, providing us with the equivalent of flanker task data for each participant.

Each trial began with the presentation of a black fixation cross in the center of a gray background for a random variable interval between 400 and 1600 ms. The fixation cross remained on screen during the entire task. In no-cue trials, the fixation interval was immediately followed by presentation of a row of five stimuli either above or below the fixation cross in equal proportions. Task stimuli consisted of a center target arrow surrounded by two flanking stimuli on either side. Flanking stimuli were either arrows (congruent and incongruent trials) or lines (neutral trials) identical in length and thickness to the target stimulus. Depending on the trial, arrow flankers either pointed in the same (congruent) or opposite direction (incongruent) as the target arrow. Participants were instructed to press the Q key with their left index finger or the P key with their right index finger based on the direction that the center arrow was pointing on each trial. Stimuli remained on screen until a response was given. After a response was registered, a second fixation period followed. The length of this fixation period was variable and was automatically adjusted so that the total duration between trials was 3000 ms. In total, 24 practice trials were completed (12 no-cue, 12 cued) which included samples of all possible trial types based on item congruency (congruent, incongruent, neutral) and cue condition (no cue, center cue, double cue, spatial cue). Feedback was provided after each practice trial in the form of a checkmark (correct) or "X" (incorrect) presented directly below the fixation cross after a response was registered. The testing phase consisted of 3 blocks each containing 96 trials for a total of 288 trials. Testing blocks each contained equal numbers of congruent, incongruent, and neutral trials, as well as cueing conditions. Trial presentation in both the practice and experimental phases was randomized. The inclusion of only no-cue trials in our analyses resulted in 72 trials per participant with 24 trials for each of the three item congruency conditions.

2.3. Administration procedures

While online data collection does not allow researchers to control the conditions in which tasks are completed, steps were taken to support uniform experimental conditions across participants. After entering the online testing system via a private link distributed through email, participants were automatically screened to ensure they were using a desktop computer or laptop. Accessing the testing system link with a smart phone or tablet resulted in automatic rejection. Participants were asked to find a quiet place in which to complete tasks, and to not engage in any distracting behaviors such as using their phone or opening websites. Similar directions were displayed before each task in the experimental design. All task directions were given in English only. Participants were allowed to take short breaks between each phase of the experiment. After giving informed consent, all participants completed the language questionnaire (LHQ-3), followed by one of two behavioral tasks (i.e., Simon or flanker task, order counterbalanced). In order to ensure that stress possibly caused by using an online task was not a critical determinant of RTs, and because stress is thought to have modulatory effects on executive function (Plieger and Reuter, 2020), participants were asked to complete the Perceived Stress Scale (PSS-10) (Cohen, 1988).

2.4. Data analysis

Behavioral data were analysed using LMEM via the lmer function from the lme4 package (Version 1.1–26; Bates et al., 2015b) using R software (Version 4.0.3; R Core Team, 2020). Excellent introductions to the use of LMEM can be found in the primary literature (Baayen et al., 2008), or in the form of an accessible instructional guide (Winter, 2019). The nlminb optimizer from the optimx package was used throughout all stages of model fitting. Analyses included data from all correct trials with the exception of those with RTs shorter than 150 ms or those resulting from suspected Internet connectivity issues. The decision not to further trim data prior to fitting was made in order to reduce the likelihood of obscuring the potential bilingual advantage (Zhou and Krott, 2016). Our experiment was determined to be sufficiently powered for LMEM analysis, exceeding the recommendation of 40 participants with 40 trials each for a total of 1600 observations (Brysbaert and Stevens, 2018).

Separate models were fit for Simon and flanker task data. Prior to model fitting, dependent variable RT data were log transformed due to non-normality, which resulted in residuals that were approximately normally distributed. All categorical variables were sum coded, and all continuous independent variables were standardized. Collinearity between variables was assessed using variance inflation factor (VIF) with variables above a value of 5 or 10 evaluated individually before inclusion in model fitting (Craney and Surles, 2002). While some researchers may adopt an arbitrary VIF threshold, a high VIF alone is not sufficient to justify the removal of variables prior to model fitting (O'Brien, 2007). Our fixed-effects structure included gender, task order, block (flanker data only), item congruency, age, SES, reported stress, number of languages spoken, hours per week playing video games, hours per week playing musical instruments, language switching frequency, L1 dominance, L1 proficiency, L2 dominance, L2 immersion, L2 proficiency and L2/L1 dominance ratio. The decision not to include L1 immersion in our model was based on our sample consisting of native Mandarin speakers who had never been outside of Mainland China for any extended period of time.

With the exception of gender, task order, age, reported stress, and number of languages spoken, interactions with item congruency were included for all fixed factors. This fixed effects structure was selected to capture predicted individual differences that may influence performance (Linck and Cunnings, 2015). Item congruency was initially sum

Table 2
Summary of Simon task performance by item congruency (average). Conflict Effect = Incongruent – Congruent.

	Item Congruency		
	Congruent	Incongruent	Neutral
Reaction Time (ms)	450 (58)	481 (50)	468 (49)
Error Rate (%)	.02 (0.03)	.11 (0.09)	.03 (0.05)
Conflict Effect: RT	31 (46)		
Conflict Effect: Error	.09 (0.09)		

coded (−1, 0, 1) during model fitting to assess main effects, and then dummy coded to compare the experimental conditions. With the "congruent" condition set as the reference level, a significant effect of congruency for the "incongruent" condition with a positive coefficient indicates a conflict effect. A bilingual advantage would be a significant main effect of L2 dominance, L2 immersion, L2 proficiency, or L2/L1 dominance ratio (negative coefficient), or else a significant interaction between these variables and the "incongruent" condition with a negative coefficient when the congruent condition was set as the reference level (Samuel et al., 2018).

Random effects structure model fitting began with a maximal model (Barr et al., 2013). Based on previous work using LMEM with RT data (Bates et al., 2015a; Momenian et al., 2021), random effects structures initially included random intercepts for participants, and by-participant random slopes for item congruency. Random effects correlation parameters were not included during model fitting. Principal component analysis (PCA) was used in order to determine the variance accounted for by each of the random factors in order to reduce the model to the most parsimonious random effects structure. Models were compared after the removal of each random factor using likelihood ratio tests (LRT). A random factor was only removed if the resultant model was not significantly different from a model that included that variable. The model with the most parsimonious random effects structure was compared to an identical model that included random effects correlation parameters using LRT. Correlation parameters were retained only if their inclusion resulted in a model that was significantly different from a model that did not include them. Finally, absolute standardized residuals exceeding 2.5 standard deviations were removed in order to address non-normal residual distribution (Baayen and Milin, 2010). Accuracy on both the Simon and flanker tasks was high across all participants and trial types (Tables 2, 4). For this reason, it was considered redundant to analyze accuracy data given possible low levels of validity.

3. Results

3.1. Modeling: Simon task

For the RT analyses, 95 incorrect trials, 1 trial with a recorded RT of 17,502 ms due to a suspected Internet connectivity issue, and 4 trials with RTs < 150 ms were removed prior to analyses. In total, 100 trials (5.81% of data) were removed, resulting in the inclusion of 1622 trials from 41 participants for analyses. After measuring VIF, the variables L2 dominance (VIF = 35.07), L2/L1 dominance ratio (VIF = 24.30), and L1 dominance (VIF = 18.12) were found to have VIF values indicative of high multicollinearity. Removal of L2 dominance greatly reduced the VIF for L2/L1 dominance ratio (VIF = 7.05) and L1 dominance (VIF = 6.60), and therefore a decision was made to include both variables given our *a priori* motivation to capture heterogeneity in our modeling. Finally, trimming extreme residuals prior to final model fitting resulted in the removal of 35 data points (2.03% of data removed) and a significantly improved final model fit (ΔAIC = −540.0; ΔBIC = −540.8).

Simon task performance is summarized in Table 2. Model fit was not influenced by the removal of item congruency from the random effects structure $x^2(6) = 9.64$, $p = 0.141$. Because random intercepts for participants were the only remaining component of our random effects structure, correlation parameters could not be included in the final model.

Model results of interest are summarized in Table 3. The presence of a significant effect of incongruent trial condition with a positive coefficient confirmed a significant Simon effect on RT. A significant interaction between L2 proficiency and RT on incongruent trials indicates that there was a decreased Simon effect associated with higher levels of L2 proficiency (Fig. 1). A significant interaction was also observed between L2 proficiency and performance on neutral trials, indicating that higher levels of L2 proficiency are associated with a general improvement in performance relative to congruent trials. Finally, a significant interaction between time playing a musical instrument and incongruent trial RT was found, with more time playing an instrument associated with faster RTs on incongruent trials relative to congruent trials (Fig. 1).

3.2. Modeling: flanker task

Due to technical issues, data for one participant was not available for analysis. We present results from $n = 40$ participants (30 females; $M_{age} = 16.19$ years, $SD_{age} = 1.57$ years). After the removal of 67 incorrect trials (2.40% of data) and 8 trials with RTs < 150 ms (< 1% of data), a total of 2711 trials from 40 participants were included in our analysis. In total, 75 trials (2.69% of data) were removed. After measuring VIF, the variables L2 dominance (VIF = 37.87), L2/L1 dominance ratio (VIF = 23.53), and L1 dominance (VIF = 18.16) were found to have VIF values indicative of high multicollinearity. Removal of L2 dominance greatly reduced the VIF for L2/L1 dominance ratio (VIF = 6.77) and L1 dominance (VIF = 6.37), and the decision was made to include both variables in our model. Trimming extreme residuals prior to final model fitting resulted in the removal of 55 data points (1.97% of data removed) and a significantly improved final model fit (ΔAIC = −1104.9; ΔBIC = −1105.7).

Flanker task performance is summarized in Table 4. Model fit was not influenced by the removal of item congruency from the random effects structure $x^2(6) = 8.96$, $p = 0.176$. Because random intercepts for participants were the only remaining component of our random effects structure, correlation parameters could not be included in the final model.

Model results of interest are summarized in Table 5. The presence of a significant effect of incongruent trial condition with a positive coefficient confirmed the presence of a flanker effect. A significant main effect of L2 proficiency indicated that higher levels of reported L2 proficiency were associated with faster performance in *all* experimental conditions (Fig. 2). Furthermore, significant main effects of both L2/L1 dominance ratio and L1 proficiency on RTs were identified, with higher levels of both variables associated with slower RTs in all experimental conditions (Table 5). There were no significant interactions between language experience and congruency.

4. Discussion

We predicted that individual differences in bilingual language experience would significantly explain inhibitory control performance on tasks used to measure the putative bilingual advantage. As predicted, we found significant effects on the performance of Mandarin-English speaking high school students when tasks were administered virtually. To the best of our knowledge, the present study is the first to report a bilingual advantage in Mandarin-English speaking high school students in a context different from most of the previous studies. Specifically, self-reported L2 proficiency was associated with faster RTs on incongruent trials relative to congruent trials on the Simon task – a bilingual inhibitory control advantage - and faster RTs overall for all flanker task conditions, suggesting a bilingual executive processing advantage or improved monitoring (Hilchey and Klein, 2011). We also observed that time playing musical instruments was associated with an inhibitory control advantage (reduced interference on the Simon task), and that

Table 3
Summary of Simon task RT effects and interactions of interest.

Fixed effects	*t* value	Std. error	*p* value	95% CI
Intercept	419.77	0.006	< 0.001	2.651, 2.676
Condition: Incongruent	7.50	0.005	< 0.001	0.025, 0.043
Condition: Neutral	4.49	0.004	< 0.001	0.011, 0.029
L2 proficiency	−0.76	0.012	0.452	−0.032, 0.014
L2 immersion	1.11	0.006	0.275	−0.005, 0.019
L2/L1 ratio of dominance	−0.43	0.014	0.668	−0.034, 0.021
L2 proficiency X Incongruent	−2.15	0.009	0.032	−0.039, −0.002
L2 proficiency X Neutral	−2.45	0.009	0.014	−0.041, −0.005
Instrument Time X Incongruent	−2.09	0.006	0.037	−0.023, −0.001
Random effects	*Variance*	*SD*		
Subject (intercept)	0.001	0.032		
Residual	0.005	0.074		

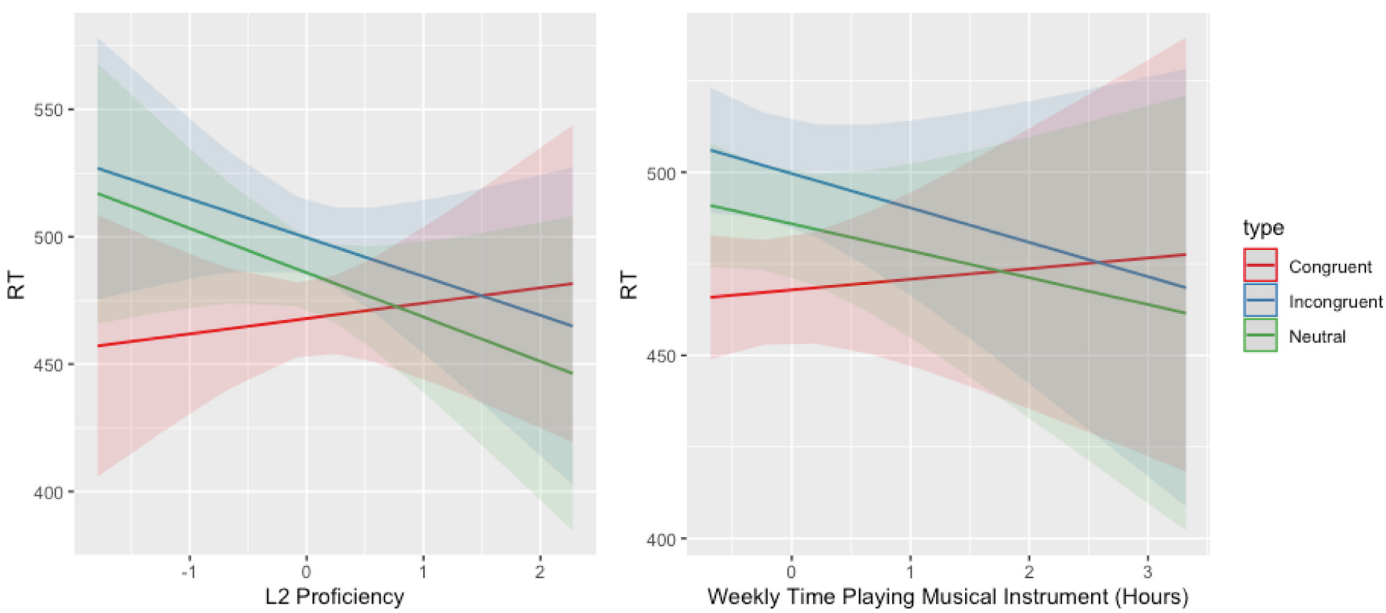

Fig. 1. Congruency interactions with reported L2 proficiency and reported weekly time playing a musical instrument on the Simon task (95% confidence interval). Predicted RT is plotted on its original scale for display purposes.

Table 4
Summary of flanker task performance by item congruency (average). Conflict Effect = Incongruent – Congruent.

	Item Congruency		
	Congruent	Incongruent	Neutral
Reaction Time (ms)	486 (42)	521 (49)	487 (44)
Error Rate (%)	.01 (0.10)	.04 (0.19)	.02 (0.13)
Conflict Effect: RT	34 (23)		
Conflict Effect: Error	.03 (0.15)		

Table 5
Summary of flanker task RT effects and interactions of interest
Condition: Incongruent estimate from Congruent condition reference level (simple effect).

Fixed effects	*t* value	Std. error	*p* value	95% CI
Intercept	524.43	0.005	< 0.001	2.688, 2.708
Condition: Incongruent	10.26	0.003	< 0.001	0.024, 0.035
L2 proficiency	−2.48	0.010	0.017	−0.042, −0.005
L2/L1 dominance ratio	2.10	0.011	0.042	0.002, 0.045
L1 proficiency	2.22	0.009	0.032	0.002, 0.040
Random effects	*Variance*	*SD*		
Subject (intercept)	0.001	0.026		
Residual	0.004	0.061		

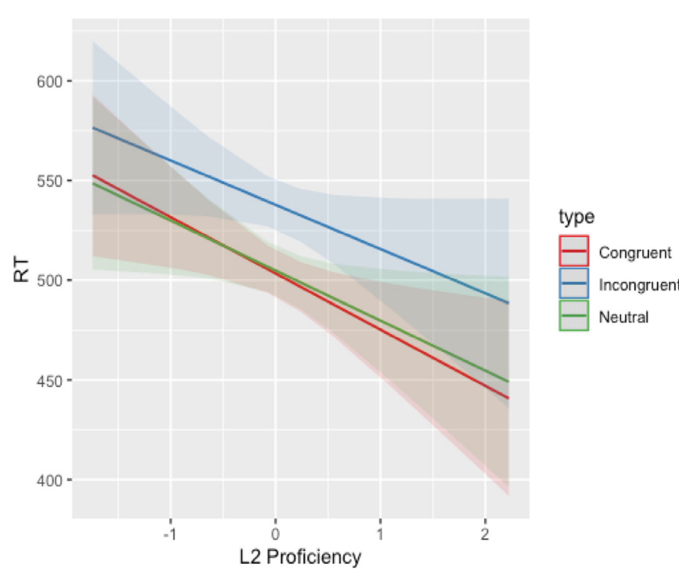

Fig. 2. Main effect of reported L2 proficiency on flanker task performance (95% confidence interval). Predicted RT is plotted on its original scale for display purposes.

L2/L1 dominance ratio and rated L1 proficiency were associated with an overall *increase* in RT on flanker performance across all conditions. The implications of the results are: (1) within-participant variability in bilingual language experience has significant effects on task performance; (2) bilingual language processing has an impact on cognitive control (inhibition and monitoring); (3) bilingual advantages are *robust* i.e., can be elicited online and are not more susceptible to stress. (4) LMEM analyses of nontrivial individual differences are more revealing than exclusively fixed-effects models.

It is not at all surprising that detailed, multi-dimensional reporting of *language experience* can explain variability in bilingual advantages (see De Bruin, 2019; Gullifer et al., 2021; Li et al., 2020; Luk and Bialystok, 2013; Marian et al., 2007), as do non-linguistic variables (e.g., Bak, 2016; Naeem et al., 2018; Samuel et al., 2018; Van den Noort et al., 2019; Ware et al., 2020). However, the latter variables are taken as support for arguments against an effect of bilingualism on cognition that could not be explained by the individual differences in SES, culture, or immigration status typical in most comparisons of monolinguals and bilinguals (Paap et al., 2015). Indeed, these confounding variables can reduce, eliminate, and reverse reported bilingual advantages on cognition (Van den Noort et al., 2019). Our results argue for an effect of bilingualism on cognition that can be directly verified by second language experience itself. However, we cannot reject further interpretation of our results based on individual differences that were not measured. Our sample was educationally, culturally, linguistically, and socioeconomically homogenous and SES had no significant impact on performance. It is possible that immigration status, income, modalities of language use (speech, reading, writing) are relevant variables. However, these did not vary systematically in our sample.

These variables should be analysed using LMEM in further tests of the bilingual advantage. The use of mixed-effects models allows for consideration of any nontrivial individual differences using a fixed and random effects structure (Linck and Cunnings, 2015). However, we contend that the categorization of bilingual language experience as a unidimensional fixed effect and the exclusion of random participant intercepts capturing individual differences are methodological errors. Our finding that significant main effects of L1 proficiency and L2/L1 dominance ratio can differentially influence performance on behavioral tasks has been observed previously (e.g., Jylkkä et al., 2017; Luk et al., 2011; Soveri et al., 2011; Tse and Altarriba, 2012). The positive effect of reported time playing musical instruments on the Simon task simply underscores the importance of considering a range of variables in tests of the bilingual advantage on aspects of cognitive function.

Previous work has investigated the impact of bilingualism on cognitive function across a wide range of ages (Giovannoli et al., 2020; Lowe et al., 2021; Paap, 2019; Ware et al., 2020). To date, little work has focused on high school age pupils. One reason may be that younger adults are assumed to be at their developmental peak in executive function (Anderson, 2002) making it likely that ceiling effects would mask bilingual advantages (Bialystok, 2016). Our results argue against the peak performance hypothesis given sufficient variability in performance (see also Moradzadeh et al., 2015; Paap, 2019; Paap et al., 2014; Samuel et al., 2018).

We observed some unexpected findings that are not consistent with a bilingual advantage. First, there was no evidence of a bilingual inhibitory control advantage on the flanker task, although there was evidence of improved monitoring or 'zooming' of attention (Ong et al., 2017). Furthermore, performance on the Simon task provided only partial support for improved monitoring (i.e., improved performance on neutral and incongruent trials). This finding is not surprising since the emergence of bilingual advantages is known to differ based on the task (Ware et al., 2020) and the nature of conflict within a task (Xia et al., 2021). Second, overall RT reductions across conditions are among the most commonly reported findings in samples of bilinguals (Hilchey and Klein, 2011). The absence of a global RT reduction on the Simon task could also relate to the number of trials we used. Although some studies presented fewer trials (e.g., Bialystok et al., 2004), others used more trials (e.g., Bialystok et al., 2005). We reasoned that too many trials could result in practice effects that mask bilingual advantages in inhibitory control, but support the emergence global reductions in RT, as we identified in our flanker task data. Global increases in RT associated with higher levels of reported L1 proficiency suggest that cognitive advantages may be limited to participants' level of L2 proficiency, as supported by previous studies (e.g., Xie, 2018; Xie and Pisano, 2019). It is worth noting that most previous studies do not explore the impact of L1 proficiency on task performance (but see Tse and Altarriba, 2012). Finally, the global increase in RT associated with higher levels of reported L2/L1 dominance ratio may reflect a decreased cognitive benefit in bilinguals who are less balanced (Yow and Li, 2015). Alternatively, this finding could have resulted from the way in which dominance was operationalized in the present study, a factor known to influence results (Anthony and Blumenfeld, 2019). The influences of L1 proficiency and L2/L1 dominance ratio on task performance merit further investigation.

Self-reporting of language experience is susceptible to demand characteristics that may be biased toward unrealistic assessments of second language proficiency in a relatively naïve sample. Although the validity of self-report measures of ability has been questioned (Zell and Krizan, 2014), these correlate with objective measures (Gollan et al., 2011; Jia et al., 2002; Li et al., 2020; Marian et al., 2007; Schrauf, 2009), and comparable effects emerge with objective and self-report measures (Zahodne et al., 2014). We chose self-report measures because these are less limiting in bilinguals with similar language experience (Tomoschuk et al., 2019). Another limitation of our study is that, although we collected data on key tasks that have been validated for use in virtual environments, additional measures of executive function would have broadened the impact of our results beyond inhibition (Miyake et al., 2000). Although it may be true that individual differences in shifting and updating are also partially captured through random participant intercepts in our model, including a direct measure of these dimensions would improve the power of our methodology. It is possible that performance on Internet-administered behavioral tasks is impacted by differences in attention more than in the laboratory-controlled environment and we are hopeful of face-to-face testing soon. Our findings suggest the bilingual advantage is graded (Ong et al., 2019). However, this conjecture requires replication across virtual tasks and in vivo. Until then we submit that the consequences of individual differences in bilingual language experience merits further investigation using LMEM.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project was supported by a Gorilla grant that provided free online task hosting, and a Postgraduate Scholarship from the University of Hong Kong, both awarded to AJP. The authors would like to acknowledge the organizational and administrative support provided by Mr. Greg Edwards and Ms. Wanyi Xie at our partner school.

Data and code availability statements

The data that support the findings of this study, and all code used in analyses are available from the corresponding author (aprivite@qq.com). All tasks used are freely available online through the Gorilla Experiment Builder.

References

Anderson, P., 2002. Assessment and development of executive function (EF) during childhood. Child Neuropsychol. 8 (2), 71–82.

Anthony, J.J.R., Blumenfeld, H.K., 2019. Language dominance predicts cognate effects and inhibitory control in young adult bilinguals. Biling. Lang. Cognit. 22 (5), 1068–1084.

Antoniou, M., 2019. The advantages of bilingualism debate. Ann. Rev. Linguist. 5, 395–415.

Anwyl-Irvine, A.L., Massonnié, J., Flitton, A., Kirkham, N., Evershed, J.K., 2020. Gorilla in our midst: an online behavioral experiment builder. Behav. Res. Methods 52 (1), 388–407.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412.

Baayen, R.H., Milin, P., 2010. Analyzing reaction times. Int. J. Psychol. Res. (Medellin) 3 (2), 12–28.

Bak, T.H., 2016. The impact of bilingualism on cognitive ageing and dementia: finding a path through a forest of confounding variables. Linguist. Approaches Biling. 6 (1–2), 205–226.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68 (3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015b. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. doi:10.18637/jss.v067.i01.

Bhandari, P., Prasad, S., Mishra, R.K., 2020. High proficient bilinguals bring in higher executive control when encountering diverse interlocutors. J. Cult. Cognit. Sci. 4 (2), 201–215.

Bialystok, E., 2006. Effect of bilingualism and computer video game experience on the Simon task. Can. J. Exp. Psychol. 60 (1), 68 Revue Canadienne de Psychologie Expérimentale.

Bialystok, E., 2016. The signal and the noise: finding the pattern in human behavior. Linguist. Approaches Biling. 6 (5), 517–534.

Bialystok, E., Craik, F.I.M., Grady, C., Chau, W., Ishii, R., Gunji, A., Pantev, C., 2005. Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. Neuroimage 24 (1), 40–49.

Bialystok, E., Craik, F.I.M., Klein, R., Viswanathan, M., 2004. Bilingualism, aging, and cognitive control: evidence from the Simon task. Psychol. Aging 19 (2), 290.

Brysbaert, M., Stevens, M., 2018. Power analysis and effect size in mixed effects models: a tutorial. J. Cognit. 1 (1).

Cohen, S., 1988. Perceived stress in a probability sample of the United States. In: The Social Psychology of Health. Sage Publications, Inc, pp. 31–67.

Craney, T.A., Surles, J.G., 2002. Model-dependent variance inflation factor cutoff values. Qual. Eng. 14 (3), 391–403.

Crump, M.J.C., McDonnell, J.V., Gureckis, T.M., 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS ONE 8 (3), e57410.

De Bruin, A., 2019. Not all bilinguals are the same: a call for more detailed assessments and descriptions of bilingual experiences. Behav. Sci. 9 (3), 33.

De Bruin, A., Treccani, B., Della Sala, S., 2015. Cognitive advantage in bilingualism: an example of publication bias? Psychol. Sci. 26 (1), 99–107.

Diamond, J., 2010. The benefits of multilingualism. Science 330 (6002), 332–333.

Donnelly, S., Brooks, P.J., Homer, B.D., 2019. Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. Psychon. Bull. Rev. 26 (4), 1122–1147.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16 (1), 143–149.

Fan, J., McCandliss, B.D., Sommer, T., Raz, A., Posner, M.I., 2002. Testing the efficiency and independence of attentional networks. J. Cognit. Neurosci. 14 (3), 340–347.

Giovannoli, J., Martella, D., Federico, F., Pirchio, S., Casagrande, M., 2020. The impact of bilingualism on executive functions in children and adolescents: a systematic review based on the PRISMA method. Front. Psychol. 11, 2398.

Gollan, T.H., Salmon, D.P., Montoya, R.I., Galasko, D.R., 2011. Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. Neuropsychologia 49 (14), 3826–3830.

Grant, A., Li, P., 2019. Proficiency affects intra-and inter-regional patterns of language control in second language processing. Lang. Cognit. Neurosci. 34 (6), 787–802.

Green, D.W., 1998. Mental control of the bilingual lexico-semantic system. Biling. Lang. Cognit. 1 (2), 67–81.

Gullifer, J.W., Kousaie, S., Gilbert, A.C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., Titone, D., 2021. Bilingual language experience as a multidimensional spectrum: associations with objective and subjective language proficiency. Appl. Psycholinguist. 42, 1–34.

Hilchey, M.D., Klein, R.M., 2011. Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. Psychon. Bull. Rev. 18 (4), 625–658.

Incera, S., McLennan, C.T., 2018. Bilingualism and age are continuous variables that influence executive function. Aging Neuropsychol. Cognit. 25 (3), 443–463.

Jentzsch, I., Mkrtchian, A., Kansal, N., 2014. Improved effectiveness of performance monitoring in amateur instrumental musicians. Neuropsychologia 52, 117–124.

Jia, G., Aaronson, D., Wu, Y., 2002. Long-term language attainment of bilingual immigrants: predictive variables and language group differences. Appl. Psycholinguist. 23 (4), 599.

Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., Laine, M., 2017. Relationship between language switching experience and executive functions in bilinguals: an Internet-based study. J. Cognit. Psychol. 29 (4), 404–419.

Li, P., Zhang, F., Yu, A., Zhao, X., 2020. Language History Questionnaire (LHQ3): an enhanced tool for assessing multilingual experience. Biling. Lang. Cognit. 23 (5), 938–944.

Linck, J.A., Cunnings, I., 2015. The utility and application of mixed-effects models in second language research. Lang. Learn. 65 (S1), 185–207.

Lowe, C.J., Cho, I., Goldsmith, S.F., Morton, J.B., 2021. The bilingual advantage in children's executive functioning is not related to language status: a meta-analytic review. Psychol. Sci., 0956797621993108.

Luk, G., Bialystok, E., 2013. Bilingualism is not a categorical variable: interaction between language proficiency and usage. J. Cognit. Psychol. 25 (5), 605–621.

Luk, G., De Sa, E., Bialystok, E., 2011. Is there a relation between onset age of bilingualism and enhancement of cognitive control? Biling. Lang. Cognit. 14 (4), 588–595.

Marian, V., Blumenfeld, H.K., Kaushanskaya, M., 2007. The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. J. Speech Lang. Hear. Res..

Mindt, M.R., Arentoft, A., Germano, K.K., D'Aquila, E., Scheiner, D., Pizzirusso, M., Sandoval, T.C., Gollan, T.H, 2008. Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. Neuropsychol. Rev. 18 (3), 255–268.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cognit. Psychol. 41 (1), 49–100.

Momenian, M., Bakhtiar, M., Chan, Y.K., Cheung, S.L., Weekes, B.S., 2021. Picture naming in bilingual and monolingual Chinese speakers: capturing similarity and variability. Behav. Res. Methods 1–12.

Moradzadeh, L., Blumenthal, G., Wiseheart, M., 2015. Musical training, bilingualism, and executive function: a closer look at task switching and dual-task performance. Cognit. Sci. 39 (5), 992–1020.

Naeem, K., Filippi, R., Periche-Tomas, E., Papageorgiou, A., Bright, P., 2018. The importance of socioeconomic status as a modulator of the bilingual advantage in cognitive ability. Front. Psychol. 9, 1818.

O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.

Ong, G., McKague, M., Weekes, B., Sewell, D.K., 2019. Diffusing the bilingual lexicon: task-based and lexical components of language switch costs. Cognit. Psychol. 114, 101225.

Ong, G., Sewell, D.K., Weekes, B., McKague, M., Abutalebi, J., 2017. A diffusion model approach to analysing the bilingual advantage for the Flanker task: the role of attentional control processes. J. Neurolinguist. 43, 28–38.

Paap, K., 2019. The bilingual advantage debate: quantity and quality of the evidence. Handb. Neurosci. Multiling. 701–735.

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.

Paap, K.R., Mason, L., Zimiga, B., Ayala-Silva, Y., Frost, M., 2020. The alchemy of confirmation bias transmutes expectations into bilingual advantages: a tale of two new meta-analyses. Q. J. Exp. Psychol. 73 (8), 1290–1299.

Paap, K.R., Wagner, S., Johnson, H., Bockelman, M., Cushing, D., & Sawi, O. (2014). 20,000 Flanker trials: are the effects reliable, robust, and stable? Poster Presented at the Association for Psychological Science, San Francisco, CA.

Plieger, T., Reuter, M., 2020. Stress & Executive Functioning: a review considering moderating factors. Neurobiol. Learn. Mem., 107254.

R Core Team. (2020). R: a language and environment for statistical computing. Vienna, Austria.

Samuel, S., Roehr-Brackin, K., Pak, H., Kim, H., 2018. Cultural effects rather than a bilingual advantage in cognition: a review and an empirical study. Cognit. Sci. 42 (7), 2313–2341.

Schrauf, R.W., 2009. English use among older bilingual immigrants in linguistically concentrated neighborhoods: social proficiency and internal speech as intracultural variation. J. Cross Cult. Gerontol. 24 (2), 157–179.

Schroeder, S.R., Marian, V., Shook, A., Bartolotti, J., 2016. Bilingualism and musicianship enhance cognitive control. Neural Plast. 2016.

Simon, J.R., Rudell, A.P., 1967. Auditory SR compatibility: the effect of an irrelevant cue on information processing. J. Appl. Psychol. 51 (3), 300.

Soveri, A., Rodriguez-Fornells, A., Laine, M., 2011. Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. Front. Psychol. 2, 183.

Tao, L., Wang, G., Zhu, M., Cai, Q., 2021. Bilingualism and domain-general cognitive functions from a neural perspective: a systematic review. Neurosci. Biobehav. Rev. 125, 264–295.

Tomoschuk, B., Ferreira, V.S., Gollan, T.H., 2019. When a seven is not a seven: self-ratings of bilingual language proficiency differ between and within language populations. Biling. Lang. Cognit. 22 (3), 516–536.

Tse, C.S., Altarriba, J., 2012. The effects of first-and second-language proficiency on conflict resolution and goal maintenance in bilinguals: evidence from reaction time distributional analyses in a Stroop task. Biling. Lang. Cognit. 15 (3), 663–676.

Van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S.H., Lim, S., 2019. Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? A systematic review. Behav. Sci. 9 (3), 27.

Ware, A.T., Kirkovski, M., Lum, J.A.G., 2020. Meta-analysis reveals a bilingual advantage that is dependent on task and age. Front. Psychol. 11 (1458).

Weekes, B.S., Chen, M.J., Qun, H.C., Lin, Y.B., Yao, C., Xiao, X.Y., 1998. Anomia and dyslexia in Chinese: a familiar story? Aphasiology 12 (1), 77–98.

Wermelinger, S., Gampe, A., Daum, M.M., 2017. Bilingual toddlers have advanced abilities to repair communication failure. J. Exp. Child Psychol. 155, 84–94.

Winter, B., 2019. Statistics for linguists: An Introduction using R. Routledge.

Xia, L., Bak, T.H., Sorace, A., Vega-Mendoza, M., 2021. Interference suppression in bilingualism: stimulus-Stimulus vs. Stimulus-Response conflict. Biling. Lang. Cognit. 1–13.

Xie, Z., 2018. The influence of second language (L2) proficiency on cognitive control among young adult unbalanced Chinese-English bilinguals. Front. Psychol. 9, 412.

Xie, Z., Pisano, T.S., 2019. Second language (L2) proficiency, socioeconomic status (SES), and intelligence (IQ) are significant predictors of cognitive control differences among young adult unbalanced Chinese–English bilinguals. Biling. Lang. Cognit. 22 (4), 866–882.

Xie, Z., Zhou, S., 2020. Bilingualism, demographics, and cognitive control: a within-group approach. Front. Psychol. 11, 94.

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. Front. Psychol. 6, 164.

Zahodne, L.B., Schofield, P.W., Farrell, M.T., Stern, Y., Manly, J.J., 2014. Bilingualism does not alter cognitive decline or dementia risk among Spanish-speaking immigrants. Neuropsychology 28 (2), 238.

Zell, E., Krizan, Z., 2014. Do people have insight into their abilities? A metasynthesis. Perspect. Psychol. Sci. 9 (2), 111–125.

Zhou, B., Krott, A., 2016. Data trimming procedure can eliminate bilingual cognitive advantage. Psychon. Bull. Rev. 23 (4), 1221–1230. doi:10.3758/s13423-015-0981-6.