

ORIGINAL ARTICLE

Different measurements of bilingualism and their effect on performance on a Simon task

Marie-France Champoux-Larsson^{1*}  and Alexandra S. Dylman²

¹Mid Sweden University and ²Stockholm University

*Corresponding author. E-mail: mfclarsson@gmail.com.

(Received 31 January 2020; revised 16 September 2020; accepted 01 October 2020; first published online 11 December 2020)

Abstract

We investigated how operationalizing bilingualism affects the results on a Simon task in a population of monolingual and bilingual native English speakers ($N = 166$). Bilingualism was measured in different ways within participants, and the measurements were used both as dichotomous and continuous variables. Our results show that the statistical significance and effect size varied across operationalizations. Specifically, the Composite Factor Score (the Language and Social Background Questionnaire's general score), showed a bilingual disadvantage on reaction times regardless of how it was used (dichotomously or continuously). When dividing participants into monolinguals and bilinguals based on the Nonnative Language Social Use score (a Language and Social Background Questionnaire subscore), differences in accuracy and reaction times were found between the groups, but the Nonnative Language Social Use score did not predict accuracy when used as a continuous variable (only reaction times). Finally, earlier age of acquisition predicted faster reaction times, but only when used on a continuum. Effect sizes were between the small and medium range. No differences on the Simon effect were found. Our results call for cautiousness when comparing studies using different types of measurements, highlight the need for clarity and transparency when describing samples, and stresses the need for more research on the operationalization of bilingualism.

Keywords: bilingualism; measurements; methodology; Simon task

In recent years, there has been a lively debate regarding the so-called bilingual advantage in executive functions. It has been suggested that bilingualism causes structural and functional changes in the brain due to the constant control that bilinguals must have on their languages, which, because they are activated simultaneously (e.g., Marian & Spivey, 2003; Wu & Thierry, 2010), constantly compete for selection (e.g., Bialystok, 2015, 2017; Bialystok, Craik, & Luk, 2012; Li, Legault, & Litcofsky, 2014). The bilingual's task to continuously monitor the environment, choose the appropriate language, and inhibit the other, is suggested to lead to domain-general changes in the brain that extend beyond linguistic control (Bialystok, 2009, 2015, 2017, Bialystok et al., 2012; Li et al., 2014). Therefore, the rationale is that bilinguals should show an advantage in

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

nonlinguistic tasks recruiting the same executive functions that are used when they control their languages (Bialystok, 2009, 2015, 2017; Bialystok et al., 2012). While there is a large body of research that has found such a bilingual advantage in children (e.g., Blom et al., 2017; Martin-Rhee & Bialystok, 2008, see Barac, Bialystok, Castro, & Saez, 2014, for a review) and adults (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Costa, Hernández, & Sebastián-Gallés, 2008; Damian, Ye, Oh, & Yang, 2018, see Bialystok et al., 2012, for a review), several studies have failed to replicate said results (e.g., Costa, Hernández, Costa-Faidella, & Sebastián-Gallés 2009; Namazi & Thordardottir, 2010; Paap, Johnson, & Sawi, 2014; Paap, Anders-Jefferson, Mason, Alvarado, & Zimiga, 2018) or have even found a bilingual disadvantage in certain executive functions (e.g., Folke et al., 2016; Paap & Greenberg, 2013; Paap & Sawi, 2014; Paap et al., 2017; Papageorgiou, Bright, Tomas, & Filippi 2018). The quest to understanding which mechanisms could lay behind a potential bilingual advantage, and why it cannot always be found, has led to numerous investigations tackling the issue from different angles (see, for instance, Lehtonen et al., 2018, for a meta-analysis). Despite the extensive number of studies investigating various aspects of executive functioning and potential differences between monolinguals and bilinguals, for instance at the neurological level (see Grundy, Anderson, & Bialystok, 2017, for a review), the debate concerning the existence (and true nature) of the bilingual advantage in executive functions remains, leaving us to wonder why no clear conclusion can be found.

However, when looking closely at studies where the executive functions of bilinguals are investigated, it becomes clear that the way bilingualism itself is defined and measured varies greatly between different studies (de Bruin, 2019). In a study by Surrain and Luk, (2019), where the different labels used to describe bilinguals in the scientific literature in recent years are reviewed, it is clear not only that different studies use different labels or characteristics to describe their bilingual sample but also that those labels are operationalized and measured differently as well. Not only does the specific facet of bilingualism that is used vary greatly across studies, but the clarity with which the operationalization of bilingualism is described and the extent to which it is measured vary as well (Surrain & Luk, 2019). Clearly, then, there is an enormous lack of consistency in the definition, operationalization, and measurement of bilingualism across studies, which makes the comparison of different results from independent studies problematic, and even impossible in some cases.

This variation may not be surprising considering the complexity of the concept of bilingualism. Important aspects such as proficiency, home versus society language use, language history, and the specific sociolinguistic context (Surrain & Luk, 2019) all are relevant facets of bilingualism that create variability even within bilinguals. For instance, there is a large body of research showing that the type and frequency of code switching in particular (see, for instance, the dynamic restructuring model, Pliatsikas, 2020; and the adaptive control hypothesis, Abutalebi & Green, 2016) lead to functional and structural neurocognitive variations within a given group of bilinguals. Because there are neurological differences between bilingual individuals depending on how they use their languages, it cannot be assumed that all types of bilingualism will potentially affect executive functions in the same way. De Bruin (2019) showed in a review that performance on tasks tapping into executive functions can vary greatly between different types of bilinguals depending on how they are defined, and how their bilingualism is measured. Bilingualism is not a homogenous concept and differences within bilinguals is an

important issue that impedes the direct comparison of results across studies. In addition, if the central characteristic of the participants is operationalized in disparate ways without knowing whether these are equivalent and can be used interchangeably, it could be methodologically problematic to compare different “types” of bilinguals with each other. This has led to an appeal to both define and operationalize bilingualism more rigorously in bilingualism research (e.g., de Bruin, 2019; Poarch & Krott, 2019; Surrain & Luk, 2019).

This methodological issue (of whether different operationalizations of bilingualism are comparable with each other), may be a contributing factor to the contradictory results found in bilingualism research. For instance, in a study by Gathercole et al. (2010), a bilingual advantage was found in elementary school children on a Stroop task. There, participants were classified as bilingual or monolingual based on an extensive background questionnaire where home and school language use was measured. However, in a different study, the participants (also elementary school children) were categorized as bilinguals based on exposure to and proficiency in Spanish and Basque (Duñabeitia et al., 2014). Duñabeitia et al. did not find that bilingual participants performed better on the Stroop task than the monolingual children. Nevertheless, whether these disparate findings are the result of the different ways bilingualism was defined and measured is unclear.

To make matters even more complex, bilingualism is often measured as a categorical variable. However, few people are either monolingual *or* bilingual, but rather, most fall somewhere on a range in between the two extremes. Thus, it is increasingly argued that bilingualism should be operationalized as a continuous variable in order to better reflect its true nature and thus increase the precision and sensitivity of the operationalization (e.g., Champoux-Larsson & Dylman, 2019; DeLuca, Rothman, Bialystok, & Pliatsikas, 2019; Edwards, 2012; Gullifer et al., 2018; Gullifer & Titone, 2020; Incera & McLennan, 2018; Jylkkä et al., 2017; Kaushanskaya & Prior, 2015; Luk & Bialystok, 2013; Sulpizio, Del Maschio, Del Mauro, Fedeli, & Abutalebi, 2020; Surrain & Luk, 2019, but see Kremin & Byers-Heinlein, 2020, for a suggestion on how to use both categorical and continuous approaches simultaneously based on a factor mixture model and on a grade-of-membership model). Yet, there are no systematic investigations to determine whether this can directly affect the outcomes of studies.

Therefore, in order to take the first step in systematically investigating the effect of defining and measuring bilingualism based on different factors, we conducted a study where we operationalized bilingualism in two different ways (i.e., based on different characteristics of bilingualism, both dichotomously and continuously) and investigated whether these different operationalizations would affect the results on a Simon task within the same participants. Because our aim was to illustrate that different types of operationalization can lead to different results, but not to systematically investigate all possible operationalizations of bilingualism, we chose to use the Language and Social Background Questionnaire (LSBQ; Anderson, Mak, Chahi, & Bialystok, 2018) to measure bilingualism. The LSBQ is a comprehensive and validated tool that suited the purpose of this study well as it measures several aspects of bilingualism (including proficiency, language use in different contexts, and code switching) and through which bilingualism can be operationalized categorically or on a continuum. The LSBQ has also the advantage of providing different

composite scores based on various aspects of bilingualism that cluster together, and where each item is weighted according to its relative contribution to the concept. Of particular interest here is the Composite Factor Score, an extensive and comprehensive estimate of bilingualism that comprises all the aspects of language use, including proficiency, that are covered in the LSBQ. Two other composite scores of interest can be found in the LSBQ, namely, the Nonnative Home Use and Proficiency score and the Nonnative Language Social Use score. As both language use at home and in society are two important aspects of bilingualism (Surrain & Luk, 2019), these two composite scores are highly relevant when operationalizing bilingualism. Although both proficiency and code switching are included in at least one of the composite scores from the LSBQ, code switching is particularly interesting on its own due to the evidence that it is associated with structural and functional neurological differences not only between monolinguals and bilinguals but also across different bilinguals (e.g., Abutalebi & Green, 2016; Pliatsikas, 2020), and we therefore chose to investigate it separately. However, an aspect of bilingualism that is relatively often used in research but that is not included in the composite scores of the LSBQ is age of acquisition. Because of its frequency in the bilingual literature, this is also an aspect that we chose to investigate on its own.

As for the Simon task, it is a paradigm that has been used in several studies where the executive functions of bilinguals were investigated (e.g., Bialystok et al., 2004, 2005; Paap et al., 2014). The Simon task consists of stimuli of two different types (e.g., a square or a circle, or two different letters) shown on the left or the right side of a screen, where the participant's task is to determine which type of stimulus is shown by pressing a key on the right or the left of the keyboard. Of importance, some trials are spatially congruent (e.g., the answer key for a circle is placed on the left side of the keyboard and the circle is shown on the left side of the screen), while other trials are spatially incongruent (e.g., the circle is shown on the right side of the screen). Incongruent trials usually lead to slower response times and more mistakes than congruent trials. The Simon effect is thus the difference in accuracy or reaction time between congruent and incongruent trials. While a bilingual advantage has been found in some studies, meaning that bilinguals showed a reduced Simon effect (e.g., Bialystok et al., 2004, 2005), this advantage is not always replicated (e.g., Paap et al., 2014). However, note that the terms advantage and disadvantage, even though they are used extensively in the bilingual-advantage debate may be misleading terms. For instance, while slower reaction times are often seen as a disadvantage, they may simply reflect adaptive cognitive mechanisms that can lead to equally good performance overall compared to a group that is faster (e.g., Gullifer & Titone, *in press*).

Thus, with this study, we aimed to investigate two main questions. First, we investigated whether the definition of bilingualism based on different factors led to different results on the Simon task in terms of accuracy, reaction times, and the Simon effect (measured as the difference between congruent and incongruent trials). Second, we investigated whether the operationalization of bilingualism as a dichotomous versus as a continuous variable affected the interpretation of the results on a Simon task also in terms of accuracy, reaction times, and the Simon effect.

Method

Participants

A total of 166 participants ($M_{age} = 37.5$, $SD_{age} = 10.4$; 50% males) were recruited online via Prolific and MechanicalTurk. However, 8 participants failed to fill out the survey properly, making it impossible to evaluate their language profile accurately, and were therefore excluded. The final sample consisted of 158 participants ($M_{age} = 37.4$, $SD_{age} = 10.5$; 49.4% males). The participants' highest completed educational level (1 = elementary school or lower, 2 = high school, 3 = professional education, 4 = bachelor's degree, 5 = master's degree, 6 = PhD) was used as an indicator of their socioeconomic status (Mode = 4, $n = 79$). Forty-two participants reported speaking only English ($M_{age} = 38.3$, $SD_{age} = 9.1$; 57.1% males; Mode_{education} = 4, $n = 17$). The remaining 116 participants ($M_{age} = 37.1$, $SD_{age} = 11$; 46.6% males; Mode_{education} = 4, $n = 62$) reported having English as their first language, and a variety of second languages. The mean reported frequency of use of the most proficient second language on average for speaking, writing, listening, and reading (each on a scale from 0 to 4) was 1.29 ($SD = 0.89$). For a list of the reported second, third, fourth, and fifth languages, please see Table 1. See the sections below for a detailed description of the sample's language profile based on the different measurements.

Procedure

Members of Prolific or MechanicalTurk self-enrolled to the study in exchange for monetary compensation (1.70 GBP and 1.25 USD respectively). The Simon task was programmed and presented online via PsyToolkit (Stoet, 2010, 2017). In our version of the task, the letters Q and P were used as stimuli and presented on the left or the right side of the screen for 2000 ms or until an answer was provided. After 12 practice trials, a total of 48 trials were presented, of which half were spatially congruent, and half were spatially incongruent (equally distributed across the two stimuli). Participants were instructed to answer as quickly and accurately as possible by pressing the letter P or Q on the keyboard. A fixation cross appeared for 500 ms between each trial as soon as an answer was provided, or after 2000 ms if no answer had been provided. The task was programmed to detect the type of device used by the participants and only participants using a physical keyboard could complete the experiment.

After the Simon task was completed, participants were sent to an online survey where information about their gender, age, and highest completed level of education was collected. Furthermore, in order to measure the participants language profile, the LSBQ (Anderson et al., 2018) was used. The questionnaire was digitalized and presented in Qualtrics. The LSBQ is a comprehensive and validated instrument that allows measuring different facets of bilingualism and provides different types of operationalizations. It consists of three parts, the first one covering demographic and other background information (Part A), the second one covering language background (Part B), and the third part concerning community language use behavior (Part C). In this study, we replaced Part A with our own demographic questions (i.e., gender, age, and education) and used Part B and Part C in their entirety. More specifically, Part B consisted of questions on which language(s) the participant

Table 1. Reported second, third, fourth, and fifth languages

Language	<i>n</i>			
	L2	L3	L4	L5
Afrikaans	1	—	1	—
Cambodian	1	—	—	—
Cantonese	3	—	—	—
Chinese	1	1	1	—
Creole	—	—	1	—
Czech	—	1	—	1
Danish	1	1	1	—
Dutch	—	1	1	—
Esperanto	1	—	—	—
French	21	17	5	1
German	6	7	4	—
Ghanaian	—	—	1	3
Greek	1	—	—	—
Gujarati	1	—	—	—
Hindi	1	2	1	—
Hungarian	1	—	—	—
Indonesian	1	1	—	—
Irish	1	1	—	—
Italian	1	4	1	1
Japanese	1	4	—	—
Kannada	—	—	1	1
Korean	1	1	—	1
Latin	1	—	—	—
Luganda	1	—	—	—
Malay	2	—	—	—
Malayalam	—	—	1	1
Mandarin	2	1	—	1
Marathi	—	1	—	—
Norwegian	—	1	1	—
Polish	1	—	1	—
Portuguese	—	1	1	—
Punjabi	1	—	—	—

(Continued)

Table 1. (Continued)

Language	<i>n</i>			
	L2	L3	L4	L5
Romanian	1	—	—	—
Russian	1	—	—	1
Spanish	37	15	2	—
Swedish	15	2	1	2
Tagalog	1	—	—	—
Telugu	—	—	1	1
Twi	2	1	—	—
Tamil	4	—	—	—
Turkish	1	—	—	—
Urdu	—	1	—	—
Yoruba	2	—	—	—

speaks and understands (including English), in which context(s) each language had been acquired, and at what age. Proficiency questions for English followed for speaking, understanding, reading, and writing on a scale from 0 to 100, as well as frequency of use of English for speaking, understanding, reading, and writing on a scale from 0 to 4 (0 = *none of the time*, 1 = *a little of the time*, 2 = *some of the time*, 3 = *most of the time*, and 4 = *all of the time*). The proficiency and frequency of use questions were presented again, but for the participant's second language (or, in cases where several second languages were reported, the one that was the most fluent). Participants who only spoke English could skip those questions. As for Part C, questions about language use were first answered on a scale from 0 to 4 (0 = *only English*, 1 = *mostly English*, 2 = *both languages equally*, 3 = *mostly the second language*, and 4 = *only the second language*). Those questions covered which language(s) were heard during different life periods (infancy, preschool age, primary school age, and high school age), which language(s) were used to communicate with different people (parents, siblings, grandparents, other relatives, partner, roommates/other people the person lives with, neighbors, and friends), which language(s) were used in different settings (at home, school, work, social activities, religious activities, hobbies, shopping and other commercial activities, healthcare, or contact with various authorities), and which language(s) were used for various activities (reading, e-mailing, texting, on social media, for writing lists and notes, for watching TV and listening to the radio, watching movies, surfing on the internet, or praying). Finally, a last block of question asked the participant how often code switching occurred on a scale from 0 to 4 (0 = *never*, 1 = *rarely*, 2 = *sometimes*, 3 = *often*, and 4 = *always*) with family members, friends, and on social media respectively.

Out of Part B and Part C, different scores can be calculated by using the provided LSBQ Factor Score Calculator (Anderson et al., 2018). We chose the LSBQ in particular as one of its factors is designed specifically to be used either as a continuous

variable or as a dichotomous variable, thus suiting the purpose of this study perfectly. Namely, the Composite Factor Score (CFS) is a measurement that includes all questions that are weighted according to the validation in Anderson et al. (2018). Another factor that can be computed is the Nonnative Home Use and Proficiency score (HUP), which includes a subset of questions related to second language use and proficiency only (for instance language used with grandparents, during infancy, proficiency in second language, etc.). Furthermore, another factor, the Nonnative Language Social Use score (LSU) includes a subset of questions related to second language use in the participant's social life (e.g., at work, when writing e-mails, frequency of code switching with friends, etc.). A last factor, namely, English proficiency, can be computed, but was not used in this study.

Data preparation

For this study, the three scores from the LSBQ described above were used. More specifically, the CFS (possible range: -6.582 to 32.32 , the higher the score, the more bilingual), the HUP (possible range: -13.9 to 24.163 , the higher the score, the more proficient the second language, and the more it is used in home settings), and the LSU (possible range: -7.5 to 80.304 , the higher the score the more frequently the second language is used in social settings). In addition, age of acquisition of the second language and code-switching frequency (CS, possible range: 0 to 4, based on the mean of three questions where answers ranged from *never* to *always*) were also used to operationalize the participants' language profile since they are frequently used to operationalize bilingualism in bilingual research.

Of importance, all the variables were used to divide the participants into two separate groups, but they were also all used as continuous variables. For the CFS, the LSBQ's guidelines were used to create a monolingual and bilingual group (monolinguals < -3.13 : $n = 60$, bilinguals > 1.23 : $n = 40$, those with a score falling in between these thresholds were excluded: $n = 58$). As for the HUP and LSU score, a median split (Mdn_{HUP} : -4.95 ; Mdn_{LSU} : -3.91) was used to create a group of monolinguals (HUP: $n = 79$, LSU: $n = 79$) and of bilinguals (HUP: $n = 79$, LSU: $n = 79$) as it is a practice that is frequently used in research despite its limitations (MacCallum, Zhang, Preacher, & Rucker, 2002). Participants who reported knowing more than one language ($n = 116$) were divided into two groups based on the age of acquisition of their L2. Participants who began to acquire their L2 before the age of 5 years were categorized as early bilinguals ($n = 40$) and those who began acquiring their L2 from the age of 5 years (or older) were categorized as late bilinguals ($n = 76$). Age 5 was chosen as the cutoff age based on findings suggesting neurological differences between bilinguals acquiring their L2 before age 5 and those acquiring it afterwards (e.g., Berken, Chai, Chen, Gracco, & Klein 2016; Bloch et al., 2009). Finally, for the participants who reported knowing more than one language, participants were divided into a group of nonswitchers for those who on average based on the three code-switching questions (with family, friends, on social media) reported code switching less frequently than "sometimes" (i.e., mean values below 2: $n = 86$) and between switchers (i.e., mean values of 2 and above: $n = 30$). For descriptive statistics for each independent variable when treated categorically and continuously, please see Table 2 and Table 3, respectively.

Table 2. Demographics based on language groups

Variable	Group	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max	Age <i>M (SD)</i>	Education Mode (<i>n</i>)	Gender % males
CFS	Monolinguals	60	-5.39	0.98	-6.58	-3.41	40.1 (11.4)	4 (32)	51.7
	Bilinguals	40	8.45	5.07	1.23	20.29	33.8 (8.8)	4 (18)	45
HUP	Monolinguals	79	-10.22	2.71	-13.9	-5.01	39.8 (10.7)	4 (43)	46.8
	Bilinguals	79	1.45	4.83	-4.9	11.89	35.1 (9.8)	4 (36)	51.9
LSU	Monolinguals	79	-6.73	1.02	-7.5	-3.5	38.6 (11.1)	4 (40)	58.2
	Bilinguals	79	12.87	16.21	-4.32	57.3	36.3 (9.9)	4 (39)	40.5
AoA	Early	40	0.7	1.4	0	4	32.3 (8.1)	4 (23)	57.5
	Late	76	17.2	10.6	5	58	39.6 (11.6)	4 (39)	40.8
CS	Nonswitcher	86	0.66	0.58	0	1.67	38.2 (11)	4 (46)	48.8
	Switcher	30	2.6	0.68	2	4	34.1 (10.8)	4 (16)	40

Note: CFS, Composite Factor Score. HUP, Nonnative Home Use score. LSU, Nonnative Language Social Use score. AoA, age of acquisition of the second language. CS, frequency of code switching.

Table 3. Demographics based on continuous variables

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	Min	Max	Age <i>M (SD)</i>	Education Mode (<i>n</i>)	Gender % males
CFS	158	-0.38	6.09	-6.58	20.29	37.5 (10.5)	4 (79)	49.4
HUP	158	-4.39	7.04	-13.9	11.89			
LSU	158	3.07	15.09	-7.5	57.31			
AoA	116	11.5	11.6	0	58	37.1 (11)	4 (62)	46.6
CS	116	1.17	1.06	0	4			

Note: CFS, Composite Factor Score. HUP, Nonnative Home Use score. LSU, Nonnative Language Social Use score. AoA, age of acquisition of the second language. CS, frequency of code switching.

Both accuracy (number of correct answers) and reaction times (in milliseconds) were analyzed for congruent and incongruent trials. Furthermore, a Simon effect score was calculated for accuracy (difference between congruent and incongruent trials) and reaction times (difference between incongruent and congruent trials), where larger scores represent a larger Simon effect. Given that event-related potentials show that focusing spatial attention on a stimulus and preparing motor action occurs around the time period of the N2-wave (Luck, 2012), answers that occurred within 200 ms were too quick for the participant to have had time to process the stimulus and were considered as mistakes. Trials where the participants did not answer within the time limit of 2000 ms were also considered as mistakes. Only correct answers were included in reaction time analysis.

The different dependent variables described above were tested in individual analyses with each predictor (CFS, HUP, LSU, age of acquisition, and frequency of code switching). Thus, for each independent variable, accuracy for congruent trials,

accuracy for incongruent trials, reaction times for congruent trials, reaction times for incongruent trials, the Simon effect based on accuracy, and the Simon effect based on reaction times were analyzed. For the dichotomous independent variables (where two groups were created), mixed-model analyses of variance were performed for the accuracy measurements (congruent and incongruent trials), as well as for the reaction times (congruent and incongruent trials). For increased readability, the main effects of condition (congruent or incongruent), which were not the principal interest of this study, are reported in supplementary materials only. Furthermore, *t* tests for independent samples were performed for each of the Simon effect measurements (based on accuracy and based on reaction times). In addition, for the categorical variables, *t* tests were performed on age, Mann–Whitney *U* tests were performed on education level, and chi-square test for independence were performed on gender to control whether the groups differed on these variables. As for the continuous independent variables, simple linear regression analyses were conducted. The analyses were performed using JASP version 0.10.2. A summary of the means, standard deviations, *p* values and effect sizes for the different variables, groups, and conditions for all analyses is presented in Table 4 at the end of the Results section.

Results

Dichotomous independent variables

CFS

The main effect of group for accuracy was not significant (monolinguals: $M = 44.5$, $SD = 3.1$; bilinguals: $M = 44.9$, $SD = 3.3$; $F < 1$), and neither was the interaction between group and condition (congruent, incongruent: $F < 1$). As for the reaction times, there was a significant difference between the groups where monolinguals ($M = 449$ ms, $SD = 77$) were faster than the bilinguals ($M = 484$ ms, $SD = 90$), $F(1, 98) = 4.33$, $p = .04$, $\eta^2 = .04$. The interaction between group and condition was not significant for reaction times either ($F < 1$). None of the analyses for the Simon effect (accuracy or reaction times) were significant (both $ts < 1$).

As for background variables, the *t* test for age showed that the monolingual group ($M = 40.1$, $SD = 11.4$) was significantly older than the bilingual group ($M = 33.8$, $SD = 8.8$), $t(98) = 2.98$, $p = .004$, $d = 0.61$. In addition, although both groups had a median of 4 for education, the bilingual group ($M = 4.2$, $SD = 0.99$) had a higher level of education than the monolingual group ($M = 3.5$, $SD = 0.89$), $U = 728$, $p < .001$. The chi-square for gender was not significant, $\chi^2(1, n = 100) = 0.43$, $p = .51$.

HUP

The main effect of group for accuracy was not significant (monolinguals: $M = 44.8$, $SD = 2.8$; bilinguals: $M = 44.8$, $SD = 3.5$; $F < 1$), and neither was the interaction between group and condition (congruent or incongruent: $F < 1$). Neither main effect of group for reaction times (monolinguals: $M = 457$ ms, $SD = 75$; bilinguals: $M = 475$, $SD = 81$; $F < 1$), nor the interaction between group and condition was significant (congruent or incongruent: $F < 1$). Further, none of the *t* tests for the Simon effect (accuracy or reaction times) were significant (both $ts < 1$).

Table 4. Summary of results

Group		Dichotomous				Continuous					
		Accuracy <i>M (SD)</i>	RT <i>M (SD)</i>	Simon		Accuracy		RT		Simon	
				Acc. <i>M (SD)</i>	RT <i>M (SD)</i>	Congr. <i>M (SD)</i>	Incongr. <i>M (SD)</i>	Congr. <i>M (SD)</i>	Incongr. <i>M (SD)</i>	Acc. <i>M (SD)</i>	RT <i>M (SD)</i>
CFS	Monolinguals	44.5 (3.1)	449 (77)*	2.4 (2.7)	47 (39)	23.5 (0.9)	21.2 (2.8)	439 (83)*	486 (89)†	2.3 (2.7)	47 (41)
	Bilinguals	44.9 (3.3)	484 (90)*	2.2 (2.7)	47 (44)						
HUP	Monolinguals	44.8 (2.8)	457 (75)	2.2 (2.5)	44 (37)	23.5 (0.8)	21.3 (2.8)	443 (78)	489 (82)	2.2 (2.7)	46 (37)
	Bilinguals	44.8 (3.5)	475 (81)	2.3 (2.9)	48 (37)						
LSU	Monolinguals	44.3 (3.3)*	446 (71)**	2.6 (2.9)	47 (36)	23.5 (0.8)	21.3 (2.8)	443 (78)*	489 (82)†	2.2 (2.7)	46 (37)
	Bilinguals	45.4 (2.9)*	486 (80)**	1.9 (2.4)	45 (38)						
AoA	Early	44.5 (3.7)	470 (87)	2.7 (3)	44 (36)	23.1 (1.1)	20.7 (3.1)	423 (83)*	469 (79)†	2.5 (3.1)	46 (42)
	Late	45.5 (2.4)	475 (72)	1.9 (2.2)	47 (35)						
CS	Nonswitcher	45.3 (2.9)	465 (75)	2 (2.5)	47 (37)	23.1 (1.1)	20.7 (3.1)	423 (83)	469 (79)	2.5 (3.1)	46 (42)
	Switcher	44.9 (3.1)	484 (81)	2.5 (2.6)	40 (29)						

Note: Acc., Accuracy. RT, reaction times (in ms). CFS, Composite Factor Score. HUP, Nonnative Home Use score. LSU, Nonnative Language Social Use score. AoA, age of acquisition of the second language. CS, frequency of code switching. Values in bold indicate a significant difference between the groups (dichotomous) or a significant (or approaching significance) predictor (continuous). * $p < .05$. ** $p < .01$. †approaching significance.

As for background variables, the t test for age showed that the monolingual group ($M = 39.8$, $SD = 10.7$) was significantly older than the bilingual group ($M = 35.1$, $SD = 9.8$), $t(156) = 2.9$, $p = .004$, $d = 0.46$. In addition, although both groups had a median of 4 for education, the bilingual group ($M = 4$, $SD = 1.1$) had a significantly higher level of education than the monolingual group ($M = 3.6$, $SD = 0.91$), $U = 2406$, $p = .008$. The gender distribution was equal, $\chi^2(1, n = 158) = 0.41$, $p = .52$.

LSU

The main effect of group for accuracy was significant, where bilinguals ($M = 45.4$, $SD = 2.9$) were more accurate than the monolinguals ($M = 44.3$, $SD = 3.3$), $F(1, 156) = 4.96$, $p = .03$, $\eta^2 = .03$. The interaction between group and condition was not significant for accuracy ($F < 1$). As for reaction times, the model was significant, showing that the monolinguals ($M = 446$ ms, $SD = 71$) were faster than the bilinguals ($M = 486$ ms, $SD = 80$), $F(1, 156) = 10.64$, $p = .001$, $\eta^2 = .06$. The interaction between group and condition was not significant for reaction times ($F < 1$). None of the t tests for the Simon effect (accuracy or reaction times) were significant: accuracy, $t(156) = 1.51$, $p = .13$, $d = .24$, reaction times: $t < 1$.

As for the background variables, the t test for age showed no significant difference between the monolingual group ($M = 38.6$, $SD = 11.1$) and the bilingual group ($M = 36.3$, $SD = 9.9$), $t(156) = 1.4$, $p = .16$, $d = 0.22$. In addition, while both groups had a median of 4 for education, the bilingual group ($M = 4.1$, $SD = 0.98$) had a higher level of education than the monolingual group ($M = 3.5$, $SD = 0.97$), $U = 2,034$, $p < .001$. The gender distribution between the groups was unequal, with 58.2% of the monolinguals being male (41.8% females), while 40.5% of the bilinguals were male (59.5% females), $\chi^2(1, n = 158) = 4.96$, $p = .03$.

Age of acquisition

There was no main effect of group for accuracy (early: $M = 44.55$, $SD = 3.7$; late: $M = 45.51$, $SD = 2.45$), $F(1, 114) = 2.79$, $p = .098$, $\eta^2 = .02$, and no significant interaction effect between group and condition, $F(1, 114) = 2.26$, $p = .135$, $\eta^2 = .01$. As for reaction times, there were no differences between the groups (early: $M = 470$ ms, $SD = 87$; late: $M = 475$ ms, $SD = 72$; $F < 1$). The interaction between group and condition was not significant for reaction times either ($F < 1$). None of the t tests for the Simon effect (accuracy or reaction times) were significant: accuracy, $t(114) = 1.5$, $p = .14$, $d = .29$, reaction times: $t < 1$.

Here again, we took a closer look at the background variables. Age was significantly lower in the early bilinguals group ($M = 32.3$, $SD = 8.1$) than in the late bilinguals group ($M = 39.6$, $SD = 11.6$), $t(114) = 3.56$, $p < .001$, $d = 0.7$. Education level (both groups with a median of 4) and gender did not differ significantly.

Code switching

The main effect of group was not significant for accuracy (nonswitchers: $M = 45.3$, $SD = 2.9$; switchers: $M = 44.9$, $SD = 3.1$; $F < 1$) or for reaction times (nonswitchers: $M = 465$ ms, $SD = 75$; switchers: $M = 484$, $SD = 81$; $F < 1$). None of the

interactions between group and condition for accuracy or reaction times were significant (both $F_s < 1$). None of the t tests for the Simon effect (accuracy or reaction times) were significant: accuracy, $t < 1$; reaction times, $t(114) = 1.03$, $p = .306$, $d = 0.22$. In addition, no significant differences were found for age ($t < 1$), education (both medians = 4), or gender.

Continuous independent variables

CFS

The model predicting reaction times on the congruent trials was significant, $F(1, 156) = 4.62$, $p = .03$, $R^2 = .03$, where the higher the CFS was, the longer the reaction times were. As for the model predicting reaction times on the incongruent trials, it approached significance, $F(1, 156) = 3.43$, $p = .066$, $R^2 = .02$, where the higher the CFS was, the longer the reaction times tended to be. The models for accuracy on the congruent and incongruent trials, as well as the models for the Simon effects (accuracy or reaction times) were not significant (all $F_s < 1$).

HUP

None of the models with the HUP score as predictor were significant: congruent reaction times, $F(1, 156) = 2.13$, $p = .146$, $R^2 = .01$; incongruent reaction times, $F(1, 156) = 1.59$, $p = .209$, $R^2 = .01$; all other $F_s < 1$.

LSU

The model for the reaction times for congruent trials was significant, $F(1, 156) = 4.69$, $p = .032$, $R^2 = .03$, but only approached significance for reaction times for incongruent trials, $F(1, 156) = 3.47$, $p = .064$, $R^2 = .02$. In both cases, the higher the LSU score, the longer the reaction times were. The models for accuracy for congruent and incongruent trials, as well as the models for the Simon effects (accuracy or reaction times) were all nonsignificant (all $F_s < 1$).

Age of acquisition

The model for reaction times for congruent trials was significant, $F(1, 114) = 4.41$, $p = .04$, $R^2 = .04$, and the model for reaction times for incongruent trials, $F(1, 114) = 3.27$, $p = .073$, $R^2 = .03$, approached significance, where a higher age of acquisition of a second language lead (or tended to lead) to slower reaction times on the Simon task. The models for both types of accuracy and for both types of Simon effects (accuracy and reaction times) were not significant (all $F_s < 1$).

Code switching

None of the models with frequency of code switching as the predictor were significant (all $F_s < 1$).

Discussion

In this study, we operationalized bilingualism in different ways and analyzed performance on a Simon task (where a bilingual advantage has at times been found in previous studies) based on various possible independent variables in order to investigate whether operationalizing bilingualism in different ways can affect the results. All predictors were used both as dichotomous variables and on a continuous scale in order to investigate whether dividing participants into distinct groups or treating bilingualism as a continuous variable would affect the results.

Our results showed that one of the predictors (the CFS), was almost constantly associated with slower reaction times for bilinguals. Namely, the more bilingual participants were according to the CFS, the slower they were on congruent trials on the Simon task, and there was a tendency toward a significant effect for the incongruent trials as well. When participants were divided into groups of monolinguals and bilinguals based on the CFS, the bilingual group was significantly slower.

Of interest, while a bilingual advantage on accuracy was found when the groups were based on the LSU score, this effect was not found when accuracy was predicted by the LSU scores on a continuous scale. Furthermore, when the LSU was used as a dichotomous variable, monolinguals were significantly faster than bilinguals for both congruent and incongruent trials. However, when LSU was treated as a continuous variable, slower reaction times were predicted by bilingualism for congruent trials only (the model only approached significance for the incongruent trials). This suggests that, in our sample, the more bilingual participants were based on the LSU score, the slower they responded to congruent trials. However, this effect was not found for incongruent trials.

Finally, while age of acquisition of the second language did not show effects on accuracy or reaction times when used dichotomously (early vs. late bilinguals), age of acquisition did significantly predict reaction times for congruent trials when bilingualism was measured on a continuum. Furthermore, the prediction of age of acquisition on incongruent trials approached significance. As for the HUP score and code switching, they did not predict any differences regardless of how they were used.

Markedly, a variation on the Simon effect was not predicted by any of the variables, whether it was measured based on accuracy or on reaction times. None of the effect sizes found in the significant results were impressively large, however, most of them being small or halfway toward medium in size. This implies that the results in this study should be interpreted with caution. It is also worth mentioning that, although we label slower reaction times as being a disadvantage, as we pointed out earlier, these slower reaction times may actually reflect a speed–accuracy trade-off as accuracy was higher for the bilinguals. It could be that the bilinguals use different cognitive strategies when performing the task. As Gullifer and Titone (*in press*) suggest, it could be that the bilinguals use active goal maintenance to a higher degree than monolinguals do when managing conflict, which could manifest itself behaviorally in terms of slower reaction times. This is not necessarily a disadvantage per se, however, especially when accuracy is improved.

Nonetheless, the results of this study demonstrate clearly that operationalization can have an effect on the results. For instance, while we found a bilingual advantage in accuracy and a disadvantage in speed when bilingualism was operationalized

dichotomously for the LSU, the disadvantage only appeared for congruent (and not incongruent) trials when the LSU was used as a continuous variable. Even more interesting, while no effects whatsoever were detected when participants were divided into early and late bilinguals, an effect on reaction times for congruent trials was found when age of acquisition was used on a continuum. A possible explanation is that the effect of LSU was found for accuracy when groups were used only due to a Type I error, since dividing a continuous variable into categories increases the risk of a Type I error (Cohen, 1983). As for the effect of LSU on reaction times, it could also be that the effect was driven by the congruent trials only, although no significant interaction effects were found. In contrast, the advantage for bilinguals having acquired their second language earlier (when using age of acquisition as a continuum on congruent trials) disappeared completely when participants were divided into early and late bilinguals. This difference may reflect the more subtle and fine-grained differences that can be found when using a variable on a continuum. Dividing participants into groups based on a continuous variable may lead to as much, if not more, variability within the group as between the groups.

The measurement that was arguably the more consistent was the CFS, where no effects were found on accuracy regardless of whether the variable was used dichotomously or continuously, but where significant effects were found for reaction times regardless of how the variable was used. This consistency may be due to several reasons. A possibility is that the CFS covers several facets of bilingualism, thus making it a robust measurement including many of the aspects of bilingualism that are of importance for performance on a Simon task. In contrast, measurements such as the LSU cover social aspects of language use only, regardless of the participant's reported language skills. This could be a possible explanation as to why a more complete measurement such as the CFS yielded more stable results than, for instance, the LSU. Another possibility is that the groups that are created based on the recommendations of the LSBQ exclude participants in a "gray zone" between monolingualism and bilingualism, thus creating groups with clear boundaries and decreasing the variability within those groups. At the same time, when used continuously, the CFS includes the "gray zone" participants and the advantage of using the variable as a continuum to detect fine-grained effects is preserved.

Another interesting aspect of this study is that, when using a median split on the LSU and HUP (which is a practice that is not necessarily optimal but that is nonetheless used), arbitrary cutoff limits were created. Thus, a bilingual in our sample may have been a monolingual in a different sample since the median will inevitably vary across samples. Here, this became even clearer as some participants ended up being bilingual according to either the LSU or the HUP, but monolingual according to the other factor. Twenty-four participants were categorized as monolingual on the LSU and bilingual on the HUP, and an additional 24 participants were categorized as bilingual on the LSU and monolingual on the HUP. None of the LSU-monolinguals were categorized as bilinguals on the CFS (21 were excluded, 3 were monolinguals). However, 2 of the HUP-monolinguals were categorized as bilinguals according to the CFS (20 were excluded, 2 were monolinguals), which is clearly problematic. These discrepancies not only demonstrate the operationalization issue well and why creating categories based on arbitrary cutoffs such as a median split makes comparison across studies problematic but also illustrates

the complexity of the bilingual experience when it is measured based on different facets, as well as how important each and every of these facets are.

An important note to make is that, according to the CFS values (computed based on the LSBQ), our sample had very few “true” bilinguals, namely, participants that had a score above the suggested cutoff value for bilingualism (i.e., several participants fell in between the monolingual and bilingual categories or in the “gray zone”). Even if the LSBQ is a validated tool and although the CFS was computed using the calculator provided by Anderson et al. (2018), the values in Anderson et al. are from a North American population and the questionnaire has been validated in this population only. It may be that the different items in the questionnaire would load differently on the factors identified by Anderson et al. if they were validated in another population. Although it is beyond the scope of this paper to validate the LSBQ in a different population, we cannot reject the possibility that the LSBQ did not measure the language profile of our population as accurately as for the population used in Anderson et al. However, a large portion of our sample was recruited via MechanicalTurk, where a majority of users are also in North America and shared several similarities with Anderson et al.’s sample. The variety of second languages spoken by our participants was similar, as was the level of education and the gender distribution, even though our sample was slightly older. Thus, it is probable that our sample is similar to the sample used in the LSBQ validation study. However, since the average CFS, LSU, and HUP are not provided in Anderson et al. (2018), it is difficult to compare the samples directly, and the possibility remains that our sample was significantly different in terms of bilingual experience.

In addition, our participants were recruited exclusively online via crowd sourcing platforms (MechanicalTurk and Prolific). Although samples recruited via such services appear to be as reliable as other more traditional samples and have the advantage of including participants outside of the college and university students population, it may be that a sample recruited in a different context would behave differently. Hauser and Norbert (2016) showed that MTurkers (participants recruited via MechanicalTurk) pay better attention than non-MTurkers when performing cognitive tests. This could explain at least in part why the accuracy rates in our study were so high. Furthermore, in particular because accuracy rates in our study were so high, the results should be interpreted with caution. The use of statistical analysis such as analysis of variance can be problematic when a ceiling effect is present and can lead to Type I errors (Šimkovic & Träuble, 2019).

Furthermore, there were some demographic differences between the groups that were created, and we cannot rule out that they may have affected the significance, or nonsignificance, of the results. Namely, CFS- and HUP-based bilinguals were younger and reported higher levels of education, but responded slower (but more accurately) when grouped based on the CFS. Based on the LSU, monolinguals reported lower levels of education and were predominantly males. Thus, the better accuracy of the bilingual group could be an effect of level of education. As for early bilinguals, they were also significantly younger, but then again, no effect was found between the groups. Only the nonswitchers versus switchers did not differ when it came to age, education, and gender, and no differences were found in terms of performance between those groups.

However, we would like to stress that the primary purpose of the current study was not to add yet another data set to the already ongoing debate of the existence of

a bilingual advantage in executive functions. Rather, the main purpose was to explore whether operationalizing bilingualism in different ways could lead to different patterns of results, which is what we found. However, although this study highlights the fact that the specific way bilingualism is operationalized can affect the results and thus the conclusions that are drawn, it is not comprehensive enough to solve this methodological concern. First, the participants of the current study consisted of adults, while the bilingual advantage in executive functions is more consistently found in children (e.g., Bialystok & Viswanathan, 2009; Janus & Bialystok, 2018; Kovács, 2009; Yow et al., 2017) or elderly populations (e.g., Bialystok, Poarch, Luo, & Craik, 2014; Borsa et al., 2018; Cox et al., 2016; Gold, Kim, Johnson, Kryscio, & Smith, 2013). Replicating the current results with additional age groups such as children or elderly populations might therefore lead to further insight in this matter, as will replicating the current study but incorporating additional tasks, particularly tasks that recruit other types of executive functions such as switch cost.

Second, there are other facets of bilingualism that we did not measure in this study. For instance, there are other questionnaires available for measuring bilingualism both for adults (e.g., the Language Experience and Proficiency Questionnaire; Marian, Blumfeld, & Kaushanskaya, 2007) and children (e.g., the Bilingual Language Experience Calculator; Unsworth, 2013), and other facets such as L2 proficiency or frequency of use of the L2 that were not explored in the current study (although they are included as part of the different LSBQ scores). In addition, there are tests beyond self-reports that may more objectively measure knowledge of a language, or at least some aspects of it such as receptive vocabulary (e.g., LexTALE: Lemhöfer & Broersma, 2012; Peabody Picture Vocabulary Test; Dunn, 2018). As we did not look at the complexity of the social aspect of bilingualism in this study, it should be more thoroughly investigated in future studies as there is growing evidence that the context of use of a bilingual's languages may be a main factor affecting executive functions (de Bruin, 2019; Tiv, Gulifer, Feng, & Titone, in press). It has even been suggested that the potential advantage may emerge from the specific social linguistic context that a bilingual interacts in (e.g., Fan, Liberman, Keysar, & Kinzler, 2015; Hartanto & Yang, 2016; Wu & Thierry, 2013) and investigating the social aspect of bilingualism in addition to linguistic aspects of bilingualism could shed light on the debate, and how to best operationalize bilingualism.

Third, although there is a movement toward treating bilingualism as a continuous variable that is endorsed by many (e.g., Champoux-Larsson & Dylman, 2019; DeLuca et al., 2019; Edwards, 2012; Gullifer et al., 2018; Gullifer & Titone, 2020; Incera & McLennan, 2018; Jylkkä et al., 2017; Kaushanskaya & Prior, 2015; Luk & Bialystok, 2013; Sulpizio et al., 2020; Surrain & Luk, 2019), as Kremin and Byers-Heinlein (2020) point out, there are situations where this may not be possible or preferable (e.g., with small samples). Therefore, giving up a dichotomous classification of bilingualism altogether may not be the answer when it comes to operationalizing bilingualism (Kremin & Byers-Heinlein, 2020). Using a factor mixture model or a grade-of-membership model, which allow for categorization but considers the variation within the categories, would allow analysing results based on categories, on a continuous scale, or even on both (Kremin & Byers-Heinlein, 2020). The flexibility that such models provide would allow choosing an approach that would be driven by the research question, by the latent concept that is analyzed,

and by the sample. Given the complexity of bilingualism, we believe that such a flexible approach would be appropriate to capture the different facets and nuances of the concept. Yet, using such models may not always be possible. Based on the recommendations of MacCallum et al. (2002) concerning the dichotomization of continuous variables, bilingualism should, when possible, be operationalized as a continuous variable given that its nature is continuous. Dichotomizing a continuous variable will almost inevitably lead to misleading results (MacCallum et al., 2002), and serious considerations should be taken before dichotomizing bilingualism. This choice should be supported for instance by inspecting the data in order to determine whether participants cluster into defined clusters, or by providing a clear theoretical and conceptual argument for doing so.

However, our goal here was to investigate whether operationalization can affect results, making them significant in some cases but not in others when operationalization is slightly modified. In other words, our aim was not to establish a universal or optimal way to operationalize and define bilingualism, but rather to highlight the fact that defining and measuring bilingualism in different ways can lead to different conclusions. This supports the argument that a thorough and clear description of the sample of bilinguals that is investigated is necessary if comparisons can be made across studies (Surrain & Luk, 2019). We suggest that our results illustrate the complexity and difficulty of measuring a latent construct such as bilingualism, and that future research should take the different nuances of the concept into account. Therefore, researchers should carefully support their operationalization choices based on clearly defined theoretical models and empirical results rather than on arbitrary cutoffs or measurements. We suggest that, as long as it is not established which aspects of bilingualism or of the bilingual experience affect specific cognitive mechanisms, caution should be used when choosing not to measure a specific characteristic. For the time being, it may still be more methodologically sound to use questionnaires that are detailed and extensive (such as the LSBQ), and refrain from methods that are known to be flawed (such as using a median split).

It is important to point out however that several of the experimental paradigms used to reveal an advantage in executive functions in bilinguals (such as the Simon task) are not necessarily the most reliable tools to draw conclusions on between-individual differences. While low variability between participants is necessary in order to reliably find group-level effects, which experimental paradigms are designed to do, high between-subjects variability is necessary in order to reliably detect individual differences (Hedge, Powell, & Summer 2018), which is often what we aim to do in bilingual research. This suggests that caution should be taken when interpreting results to highlight differences between monolinguals and bilinguals (or between different types of bilinguals) and that methodological changes are necessary when designing tasks and analyzing results if such paradigms are used (see Hedge et al., 2018, for recommendations).

In sum, although the current study cannot settle how bilingualism is best measured, nor which facets of it are the most relevant, it achieves its aim of showing that operationalizing bilingualism in different ways (specifically, using different facet of bilingualism, and whether bilingualism is measured as a continuous or dichotomous variable) does affect the results. Here, we showed that dividing participants into groups instead of using their scores continuously tipped the value from approaching

significance or not being significant to showing significant differences between monolinguals and bilinguals. Furthermore, our results also show that differences can (or cannot) be found depending on which aspect of the behavioral data is analyzed (i.e., accuracy vs. reaction times on congruent vs. incongruent trials, or on a score such as the Simon effect). This illustrates quite well how these results, had they come from different samples and been published as independent studies, would have led to inconsistent results where a bilingual advantage in accuracy would have been established in one study, but not in another, and where a bilingual disadvantage in speed would even have been supported by a third one. Simultaneously, there are several consistencies across the measurements and the dependant variables as well, suggesting that consistency across studies, despite using different types of operationalization, is not impossible. While this specific methodological issue is unlikely to be the sole source of all disparate results on the bilingual advantage in executive functions, it is possibly a contributing factor. Our results therefore stress the need for more research investigating operationalization and measurement of bilingualism. Future research should investigate which facets of bilingualism correlate together, which lead to similar neurological changes and effects on tasks tapping into executive functions, and which facets can and cannot be compared directly with each other. This study also highlights the need to clearly and transparently report how bilingualism is operationalized in studies investigating bilinguals in order for informed comparisons and meta-analyses to be possible. We believe that the current study contributes to the efforts toward more rigorous and standardised methodologies, and encourages increased openness, awareness, and transparency in the field of bilingual research.

References

- Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*, 689–698. doi: [10.1017/S1366728916000225](https://doi.org/10.1017/S1366728916000225)
- Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The Language and Social Background Questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*, 250–263. doi: [10.3758/s13428-017-0867-9](https://doi.org/10.3758/s13428-017-0867-9)
- Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714. doi: [10.1016/j.ecresq.2014.02.003](https://doi.org/10.1016/j.ecresq.2014.02.003)
- Berken, J. A., Chai, X., Chen, J.-K., Gracco, V. L., & Klein, D. (2016). Effects of early and late bilingualism on resting-state functional connectivity. *Journal of Neuroscience*, *36*, 1165–1172. doi: [10.1523/JNEUROSCI.1960-15.2016](https://doi.org/10.1523/JNEUROSCI.1960-15.2016)
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*, 3–11. doi: [10.1017/S1366728908003477](https://doi.org/10.1017/S1366728908003477)
- Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, *9*, 114–121. doi: [10.1111/cdep.12116](https://doi.org/10.1111/cdep.12116)
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*, 233–262. doi: [10.1037/bul0000099](https://doi.org/10.1037/bul0000099)
- 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*, 40–49. doi: [10.1016/j.neuroimage.2004.09.044](https://doi.org/10.1016/j.neuroimage.2004.09.044)
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*, 290–303. doi: [10.1037/0882-7974.19.2.290](https://doi.org/10.1037/0882-7974.19.2.290)

- Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Science*, *16*, 240–250. doi: [10.1016/j.tics.2012.03.001](https://doi.org/10.1016/j.tics.2012.03.001)
- Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. M. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging*, *29*, 696–705. doi: [10.1037/a0037254](https://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. doi: [10.1016/j.cognition.2009.06.014](https://doi.org/10.1016/j.cognition.2009.06.014)
- Bloch, C., Kaiser, A., Kuenzli, E., Zappatore, D., Haller, S., Franceschini, R., . . . Nitsch, C. (2009). The age of second language acquisition determines the variability in activation elicited by narration in three languages in Broca's and Wernicke's area. *Neuropsychologia*, *47*, 625–633. doi: [10.1016/j.neuropsychologia.2008.11.009](https://doi.org/10.1016/j.neuropsychologia.2008.11.009)
- Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology*, *8*, 1–12. doi: [10.3389/fpsyg.2017.00552](https://doi.org/10.3389/fpsyg.2017.00552)
- Borsa, V. M., Perani, D., Della Rosa, P. A., Videsott, G., Guidi, L. Weekes, B. S., . . . Abutalebi, J. (2018). Bilingualism and healthy aging: Aging effects and neural maintenance. *Neuropsychologia*, *111*, 51–61. doi: [10.1016/j.neuropsychologia.2018.01.012](https://doi.org/10.1016/j.neuropsychologia.2018.01.012)
- Champoux-Larsson, M.-F., & Dylman, A. S. (2019). A prosodic bias, not an advantage, in bilinguals' interpretation of emotional prosody. *Bilingualism: Language and Cognition*, *22*, 416–424. doi: [10.1017/S1366728918000640](https://doi.org/10.1017/S1366728918000640)
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249–253. doi: [10.1177/014662168300700301](https://doi.org/10.1177/014662168300700301)
- 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. doi: [10.1016/j.cognition.2009.08.001](https://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. doi: [10.1016/j.cognition.2006.12.013](https://doi.org/10.1016/j.cognition.2006.12.013)
- Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia*, *91*, 299–306. doi: [10.1016/j.neuropsychologia.2016.08.029](https://doi.org/10.1016/j.neuropsychologia.2016.08.029)
- Damian, M. F., Ye, W., Oh, M., & Yang, S. (2018). Bilinguals as “experts”? Comparing performance of mono- to bilingual individuals via a mousetracking paradigm. *Bilingualism: Language and Cognition*, *22*, 1176–1193. doi: [10.1017/S1366728918000901](https://doi.org/10.1017/S1366728918000901)
- de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*, 1–13. doi: [10.3390/bs9030033](https://doi.org/10.3390/bs9030033)
- DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences, USA*, *116*, 7565–7574. doi: [10.1073/pnas.1811513116](https://doi.org/10.1073/pnas.1811513116)
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*. doi: [10.1027/1618-3169/a000243](https://doi.org/10.1027/1618-3169/a000243)
- Dunn, D. M. (2018). *Peabody Picture Vocabulary Test* (5th ed., PPVT-5) [Measurement instrument]. London: Pearson.
- Edwards, J. (2012). Conceptual and methodological issues in bilingualism and multilingualism research. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 5–25). London: Wiley-Blackwell.
- Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*, 1090–1097. doi: [10.1177/0956797615574699](https://doi.org/10.1177/0956797615574699)
- Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in meta-cognitive processing. *Cognition*, *150*, 119–132. doi: [10.1016/j.cognition.2016.02.008](https://doi.org/10.1016/j.cognition.2016.02.008)
- Gathercole, V. C. M., Thomas, E. M., Jones, L., Guasch, N. V., Young, N., & Hughes, E. K. (2010). Cognitive effects of bilingualism: Digging deeper for the contributions of language dominance, linguistic knowledge, socio-economic status and cognitive abilities. *International Journal of Bilingual Education and Bilingualism*, *13*, 617–664. doi: [10.1080/13670050.2010.488289](https://doi.org/10.1080/13670050.2010.488289)

- Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, *33*, 387–396. doi: [10.1523/JNEUROSCI.3837-12.2013](https://doi.org/10.1523/JNEUROSCI.3837-12.2013)
- Grundy, J. G., Anderson, J. A. E., & Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Annals of the New York Academy of Sciences*, *1396*, 183–201. doi: [10.1111/nyas.13333](https://doi.org/10.1111/nyas.13333)
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, *117*, 123–134. doi: [10.1016/j.neuropsychologia.2018.04.037](https://doi.org/10.1016/j.neuropsychologia.2018.04.037)
- Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*, 283–294. doi: [10.1017/S1366728919000026](https://doi.org/10.1017/S1366728919000026)
- Gullifer, J. W., & Titone, D. (in press). Engaging proactive control: Influences of diverse language experiences using insights from machine learning. *Journal of Experimental Psychology: General*. doi: [10.31234/osf.io/7jq3w](https://doi.org/10.31234/osf.io/7jq3w)
- Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition*, *150*, 10–19. doi: [10.1016/j.cognition.2016.01.016](https://doi.org/10.1016/j.cognition.2016.01.016)
- Hauser, D. J., & Norbert, S. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407. doi: [10.3758/s13428-015-0578-z](https://doi.org/10.3758/s13428-015-0578-z)
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166–1186. doi: [10.3758/s13428-017-0935-1](https://doi.org/10.3758/s13428-017-0935-1)
- Incera, S., & McLennan, C. T. (2018). Bilingualism and age are continuous variables that influence executive function. *Aging, Neuropsychology, and Cognition*, *25*, 443–463. doi: [10.1080/13825585.2017.1319902](https://doi.org/10.1080/13825585.2017.1319902)
- Janus, M., & Bialystok, E. (2018). Working memory with emotional distraction in monolingual and bilingual children. *Frontiers in Psychology*, *9*, 1582. doi: [10.3389/fpsyg.2018.01582](https://doi.org/10.3389/fpsyg.2018.01582)
- Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An Internet-based study. *Journal of Cognitive Psychology*, *29*, 404–419. doi: [10.1080/20445911.2017.1282489](https://doi.org/10.1080/20445911.2017.1282489)
- 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. doi: [10.1017/S1366728914000510](https://doi.org/10.1017/S1366728914000510)
- Kovács, Á. M. (2009). Early bilingualism enhances mechanism of false-belief reasoning. *Developmental Science*, *12*, 48–54. doi: [10.1111/j.1467-7687.2008.00742.x](https://doi.org/10.1111/j.1467-7687.2008.00742.x)
- Kremin, L. V., & Byers-Heinlein, K. (2020). *Why not both? Rethinking categorical and continuous approaches to bilingualism*. PsyArXiv preprints. doi: [10.31234/osf.io/nkvap](https://doi.org/10.31234/osf.io/nkvap)
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J. de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*, 394–425. doi: [10.1037/bul0000142](https://doi.org/10.1037/bul0000142)
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. doi: [10.3758/s13428-011-0146-0](https://doi.org/10.3758/s13428-011-0146-0)
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*, 301–324. doi: [10.1016/j.cortex.2014.05.001](https://doi.org/10.1016/j.cortex.2014.05.001)
- Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 329–360). Oxford: Oxford University Press.
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*, 605–621. doi: [10.1080/20445911.2013.795574](https://doi.org/10.1080/20445911.2013.795574)
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40. doi: [10.1037/1082-989X.7.1.19](https://doi.org/10.1037/1082-989X.7.1.19)
- 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. doi: [10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067)
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115. doi: [10.1017/S1366728903001068](https://doi.org/10.1017/S1366728903001068)

- Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*, 81–93. doi: [10.1017/S1366728907003227](https://doi.org/10.1017/S1366728907003227)
- Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, *13*, 597–616. doi: [10.1080/13670050.2010.488288](https://doi.org/10.1080/13670050.2010.488288)
- Paap, K. R., Anders-Jefferson, R., Mason, L., Alvarado, K., & Zimiga, B. (2018). Bilingual advantages in inhibition or selective attention: More challenges. *Frontiers in Psychology*, *9*, 1409. doi: [10.3389/fpsyg.2018.01409](https://doi.org/10.3389/fpsyg.2018.01409)
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. doi: [10.1016/j.cogpsych.2012.12.002](https://doi.org/10.1016/j.cogpsych.2012.12.002)
- Paap, K. R., Johnson, H., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*, 615–639. doi: [10.1080/20445911.2014.944914](https://doi.org/10.1080/20445911.2014.944914)
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*, 89–112. doi: [10.1080/20445911.2016.1248436](https://doi.org/10.1080/20445911.2016.1248436)
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, 1–15. doi: [10.3389/fpsyg.2014.00962](https://doi.org/10.3389/fpsyg.2014.00962)
- Papageorgiou, A., Bright, P., Tomas, E. P., & Filippi, R. (2018). Evidence against a cognitive advantage in the older bilingual population. *Quarterly Journal of Experimental Psychology*, *72*, 1354–1363. doi: [10.1177/1747021818796475](https://doi.org/10.1177/1747021818796475)
- Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. *Bilingualism: Language and Cognition*, *23*, 459–471. doi: [10.1017/S1366728919000130](https://doi.org/10.1017/S1366728919000130)
- Poarch, G. J., & Krott, A. (2019). A bilingual advantage? An appeal for a change in perspective and recommendations for future research. *Behavioral Sciences*, *9*, 1–13. doi: [10.3390/bs9090095](https://doi.org/10.3390/bs9090095)
- Šimkovic, M., & Träuble, B. (2019). Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLoS ONE*, *14*, 1–47. doi: [10.1371/journal.pone.0220889](https://doi.org/10.1371/journal.pone.0220889)
- Stoet, G. (2010). PsyToolkit—A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*, 1096–1104. doi: [10.3758/BRM.42.4.1096](https://doi.org/10.3758/BRM.42.4.1096)
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*, 24–31. doi: [10.1177/0098628316677643](https://doi.org/10.1177/0098628316677643)
- Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage*, *205*, 116306. doi: [10.1016/j.neuroimage.2019.116306](https://doi.org/10.1016/j.neuroimage.2019.116306)
- Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, *22*, 401–415. doi: [10.1017/S1366728917000682](https://doi.org/10.1017/S1366728917000682)
- Tiv, M., Gullifer, J., Feng, R., & Titone, D. (in press). Using Network Science to map what Montréal bilinguals talk about across languages and communicative contexts. *Journal of Neurolinguistics*. doi: [10.31234/osf.io/bmry2](https://doi.org/10.31234/osf.io/bmry2)
- Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*, 86–110. doi: [10.1017/S1366728912000284](https://doi.org/10.1017/S1366728912000284)
- Wu, Y. J., & Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *Journal of Neuroscience*, *30*, 7646–7651. doi: [10.1523/JNEUROSCI.1602-10.2010](https://doi.org/10.1523/JNEUROSCI.1602-10.2010)
- Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *Journal of Neuroscience*, *33*, 13533–13537. doi: [10.1523/JNEUROSCI.4760-12.2013](https://doi.org/10.1523/JNEUROSCI.4760-12.2013)
- Yow, W. Q., Li, X., Lam, S., Gliga, T., Chong, Y. S., Kwek, K., & Broekman, B. F. P. (2017). A bilingual advantage in 54-montholds' use of referential cues in fast mapping. *Developmental Science*, *20*, 1–15. doi: [10.1111/desc.12482](https://doi.org/10.1111/desc.12482)

Cite this article: Champoux-Larsson, M.-F. and Dylman, A.S. (2021). Different measurements of bilingualism and their effect on performance on a Simon task. *Applied Psycholinguistics* *42*, 505–527. <https://doi.org/10.1017/S0142716420000661>