The impact of bilingualism on executive functions in healthy adults

Facoltà di Medicina e Psicologia
Dipartimento di Psicologia
Dottorato in Psicologia e Scienza Cognitiva
XXXIV Ciclo

Jasmine Giovannoli
Matricola 1529261

Tutor
Prof.ssa Maria Casagrande

Co-Tutor
Prof.ssa Diana Martella

Co-Tutor
Prof.ssa Sabine Pirchio

Contents

Abstract

Approximately half of the world's population is bilingual or multilingual. The bilingual advantage theory claims that the constant need to control known languages that are always active in the brain to use the one suitable for each specific context improves cognitive functions, specifically executive functions. However, some authors do not agree on the bilingual effect, given the controversial results of studies on this topic. This doctoral thesis aims to analyze the impact of bilingualism on executive functions.

Two systematic reviews are presented in Chapters 1 and 3. The first is focused on investigating the effect of bilingualism on cognitive and motor inhibition; the second summarizes the results of studies that assessed language ability and executive functions using verbal fluency tasks.

Subsequently, three experimental chapters are presented. The first study investigated cognitive and motor inhibition in the bilingual population and the effect of the task on performance. The second study analyzed verbal fluency performance using various variables such as qualitative, quantitative, and time-course indices. The last study explored the impact of the language used to perform the verbal fluency task by comparing the performance of a bilingual group who completed the test using both the dominant language and the L2.

All chapters of the thesis are designed to investigate the effects that contribute to a bilingual advantage in executive functions. In general, the results of the two systematic reviews and the experimental data presented in this dissertation did not sustain the bilingual effect on the executive functions. The systematic review evidenced that the differences between language groups rarely emerge. The experimental studies failed to show differences in cognitive and motor inhibition ability, and no advantages in executive functioning emerge when considering the performance on the verbal fluency task. Moreover, bilinguals performed better when they completed the task in their dominant language.

Introduction

Most of the world's population today is bilingual. In general, the term bilingual describes a person who knows two languages. Although this indication appears relatively simple, defining who is bilingual is not easy. Indeed, considering different aspects that characterize the linguistic history of the participants, it is possible to identify different types of bilingualism. In the book "The Bilingualism Reader", Wei (2020) reported thirty-seven different definitions of a bilingual person taking into account the age of acquisition of the two languages (e.g., early bilingual, a person who acquired two languages early in childhood), the degree of linguistic competence (e.g., balanced bilingual, a person with equivalent knowledge of two languages), or the type of languages known (e.g., vertical bilinguals, who are bilingual in a standard language and a distinct language or dialect).

The first studies on bilingualism date back to the early 1900s. Initially, several researchers supported the hypothesis that bilingual children had lower mental abilities than monolinguals because the knowledge of several languages would generate mental confusion with deleterious consequences on every cognitive aspect (Hakuta, 1986). Peal and Lambert (1962) were the first to contradict this negative view about the bilingualism effect. Since then, the interest in the topic of bilingualism has increased exponentially. According to the Scopus database, more than 70% of the articles on this topic were published in the last ten years. The studies investigated the effect of bilingualism in various fields. Some studies examined the effectiveness of bilingual education (e.g., Costa & Guasti, 2021), other studies investigated the relationship between bilingualism and language disorders (Kohnert et al., 2020), and others the effect of bilingualism on healthy (e.g., Borsa et al., 2018) and pathological (e.g., Sala et al., 2021) aging. However, a topic that elicited considerable interest and debate is the effect of this socio-demographic condition on cognitive functioning (e.g., Kuzyk et al., 2020) and executive functions (e.g., Antón et al., 2019). According to the model of Miyake et al. (2000), executive functions refer to cognitive flexibility (e.g., the ability to switch between tasks), inhibition (e.g., the ability to suppress dominant responses), and monitoring (e.g., the ability to update information in the working memory). Because of the positive results of several studies, some authors started to talk about the bilingual advantage. The positive effect of bilingualism would depend on the constant need to control both known languages to use the one suitable for each specific context. This process would generate more significant neurological development (Bialystok, 1999, 2001). According to the Joint Activation Model of Green (1998), both languages would always be active in the brain of a bilingual person regardless of the language used at the given moment. It would be necessary to use a general suppression mechanism to inhibit the activation of the non-target language. Green and Abutalebi (2013) highlighted the importance of the context in which language exchanges take place. They proposed the Adaptive control hypothesis and identified three possible contexts of interaction: single-language, dual-language, and dense codeswitching contexts. Depending on the communicative context in which bilinguals are immersed, the languages may cooperate or compete. For this reason, each context is characterized by a different use of processes that are the basis of communication. The use of multiple languages would seem to modify both the language network and the control network (Green and Kroll, 2019). After the publication of positive evidence on the bilingual advantage, the difficulty in replicating previous results and the publication of several studies with null findings led to questioning this theory. Paap et al. (2015) stated that “bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances” and pointed out that 80% of the tests carried out after 2011 failed to obtain results in support of the bilingual effect. Moreover, Leivada et al. (2020) suggested adopting the more neutral term “bilingual effect.” Others (e.g., Grundy, 2020) suggested the need to change the approach adopted for the study of this condition, highlighting the need to replace the question "does the bilingual advantage exist" with "when does the bilingual advantage emerge". The debate on this topic is still ongoing. Numerous reviews and meta-analyses have been published over the last few years (e.g., Brini et al., 2020; Giovannoli et al., 2020; Lehtonen et al., 2018), attempting to summarize the results of published studies. However, definitive conclusions are not achieved, and sometimes conflicting results emerge from these studies.

This doctoral thesis aims to investigate the effect of bilingualism on the executive functions of bilingual adults.

In Chapter 1, a systematic review of the effect of bilingualism on cognitive and motor inhibition is reported. The review summarizes the results of studies published up to May 2021 that included participants over 18 years of age and assessed inhibition ability using at least one of the following tests: Stroop task, Flanker task, Simon task, Go/No-Go task.

Chapter 2 includes a study exploring the effect of bilingualism on cognitive and motor inhibition in different age groups (young adults, middle-aged adults, and older adults). To assess inhibition ability, participants completed six tasks (verbal Stroop task, nonverbal Stroop task, Simon task, Flanker task, Go/No-Go task, and Global/Local task). The use of this large battery of tests to assess a single component of executive function is driven to test the effect of task selection on performance.

A systematic review investigating the effect of bilingualism on verbal fluency test performance is reported in Chapter 3. The review included all studies published up to May 2021 that assessed bilingual adults' semantic or verbal fluency performance.

Chapter 4 presents a study investigating the effect of bilingualism on verbal fluency test performance by considering quantitative, qualitative, and time-based indices of language production. The study assessed the participants' language skills and executive functioning. Furthermore, different experimental conditions were compared to determine a possible effect of the stimuli used for the assessment.

Chapter 5 reported a study investigating the effect of the language used for the verbal fluency test in a group of unbalanced bilinguals. Participants were asked to complete verbal fluency tasks in their L1 (Italian, dominant language for all participants) and their L2. Results are analyzed using qualitative, quantitative, and time-based indices of language production to allow simultaneous assessment of language skills and executive functioning.

The thesis chapters were structured as independent papers, following the typical format of scientific publications. For this reason, there will be some inevitable repetitions throughout the work to make the single papers self-explanatory.

The last chapter summarizes the main results of this thesis.

Chapter 1: Executive inhibition in young, middle-aged, and elderly people: a systematic review

Introduction

According to the Inhibitory Control Model (Green, 1998), when bilinguals use only one language, both known languages are activated in the brain. The lifelong practice that bilinguals receive from managing two languages seems to lead to a cognitive benefit. The effects of bilingualism on cognitive functions would also depend on the context in which communicative exchanges occur (Adaptive Control Hypothesis, Green & Abutalebi, 2013), which determines whether the available languages may cooperate or compete. According to Bialystok (2011), during a conversation, bilinguals must be based on the context and require constant access to the information contained in the working memory. Furthermore, they need to select the appropriate language for the specific communicative situation (inhibiting the other language) and monitor what happens during the interaction (cognitive flexibility). For these reasons, bilingualism would appear to have a greater impact on executive functions. There is currently an intense scientific debate about the effect of bilingualism on executive functions. Indeed, the results of the studies are inconsistent. Some studies found a positive impact of bilingualism on executive functions, while others did not observe a difference between bilingual and monolingual populations (for a review, see Giovannoli et al., 2020; Grundy, 2020).

Several studies focused on inhibition ability, as this process is recruited for bilingual language comprehension and production processes.

According to the Dimensional Overlap Model (Kornblum, 1994), the tasks that assess the inhibition ability can generate stimulus-stimulus (S-S) or stimulus-response (S-R) incompatibility. S-S incompatibility occurs when there is interference between the relevant and irrelevant stimulus, while S-R incompatibility occurs when there is an overlap between the irrelevant stimulus and the response. The same types of interference seem to emerge in the linguistic processes of bilinguals (Bellegarda et al., 2021). An S-S incompatibility occurs in the conflict between co-activated language representations. For example, when similar-sounding words compete, inhibitory control can allow word identification correctly. Cross-linguistic conflict is also likely to arise during bilingual language production. In contrast, S-R incompatibility would manifest in the conflict between two overt responses if a cross-linguistic conflict was not resolved before response choice. Hilchey & Klein (2011) proposed two hypotheses to explain the possible bilingual advantage in executive functions. The bilingual inhibitory control advantage (BICA) hypothesis refers to a superior ability to control interference shown by bilinguals that would lead to more efficient inhibitory processes in the presence of conflict (i.e., incongruent trials) and smaller interference effects. The bilingual executive processing advantage (BEPA) hypothesis suggests a superior executive control ability responsible for faster processing and a global RTs advantage. Specifically, the results of their review, which included all studies that had adopted the Simon task and the flanker task, confirm the BEPA hypothesis.

Inhibition ability was assessed with both verbal and non-verbal tasks. The use of verbal stimuli for the assessment of the bilingual population seems to cause less efficient performance.

The most used tasks to assess inhibition ability are the Simon task (Simon & Wolf, 1963), the Flanker task (Eriksen & Eriksen, 1974), the Stroop task (Stroop, 1935), and the Go/No-go task (Newman & Kosson, 1986). The flanker task allows assessing response inhibition and the ability to suppress conflicting information. The task can be presented alone or within the Attention Network Test (Fan et al., 2002). Participants had to discriminate a central stimulus (e.g., arrows) and ignore the flankers that can be neutral (i.e., no interference with the target), congruent (identical to the target), or incongruent (different from the target). The flanker task causes an S-S interference, but in incongruent trials, both S-S and S-R are incompatible because the relevant and irrelevant stimuli share the same dimensions as the relevant stimulus and the response.

In the Simon task, participants must discriminate the target to the left or the right of a fixation cross. Typically, one of the two alternative response choices is assigned to each hand, and each hand is aligned with a target location to the left or right of fixation. The classic version of the Simon task generates an S-R interference. In contrast, its variation, the Simon arrow task (also known as Spatial Stroop task), generates S-S interference in incongruent trials because the task-relevant direction is incompatible with the task-irrelevant location.

The Stroop task is an experimental task used to assess the cognitive inhibitory system. The classic version of this task requires participants to indicate the color of the ink of a word while ignoring its meaning. The ink and the word meaning are equal in congruent trials, while the ink and word meaning differ in incongruent trials. The Stroop task causes an S-S interference. Numerous variations of the classic task have been implemented, such as the Spatial Stroop task (e.g., Blumenfeld & Marian, 2014), in which participants must discriminate the direction of an arrow, or the numerical Stroop task (e.g., Antón et al., 2019).

The Go/No-Go is used to evaluate the inhibitory motor system. Participants were required to respond to a given stimulus (Go trials) and withhold their responses on non-target trials (No-Go trials). No-Go trials that required inhibition of non-desired automatic responses were indexed as an inhibition marker.

This critical review of the literature aims to analyze the most common tasks used to assess inhibitory control (Flanker task, Stroop task, Simon task, Go/No-Go task).

The results of several studies (Kousaie & Phillips, 2017; Paap & Greenberg, 2013; Paap et al., 2015) evidenced little convergent validity among flanker task, Simon task, and Stroop task, suggesting that these tasks should not be taken as interchangeable measures of executive control. For this reason, it is necessary to summarize the results obtained from the various studies that assessed inhibition using different paradigms to establish a clear picture of the inhibition abilities of bilinguals.

Method

The review process was conducted according to the PRISMA -Statement (Liberati et al., 2009; Moher et al., 2009). The PRISMA Statement consists of a 27-item checklist and a four-phase flow diagram and helps authors improve systematic review reporting. The protocol was not registered.

Research strategies

A systematic search of the international literature was conducted in the following electronic databases by selecting articles published in peer-review journals: PsycINFO, PsycARTICLES, MEDLINE, PubMed, Web of Science, and SCOPUS. The last research was conducted on 13 May 2021. Restrictions were made limiting the research to academic publications in English, Italian, and Spanish. No restriction of age, gender, or ethnicity was made. The search strategy used Boolean combinations of the following keywords: "bilingual*", "second language", "executive function*", "inhibition", and "cognitive inhibition". Reference lists of the selected articles were screened. A total of 4753 articles were obtained from the search procedure. Mendeley reference manager software was used for removing duplicates. The first screening was made by reading the title and abstract. The full text of the selected studies was read.

Eligibility Criteria

Inclusion criteria of the studies were: the presence of at least one bilingual group and one monolingual group, at least one inhibition task, participants aged over 18 years. Studies that considered bimodal bilingual, second language learners, and trilingual or multilingual people were excluded. Studies on clinical populations were excluded.

Data Collection

According to the PICOS approach (Liberati et al., 2009), the following information has been extracted from the selected studies: author(s) and year of publication, country, characteristics of participants (age, percentage of females, years of education, socioeconomic status, spoken languages, age of acquisition of the languages), criteria used for selecting bilingual participants, the experimental paradigm used, results of the studies. Data concerning the characteristics of participants are summarized in Table 1. The results of the studies are summarized in Tables 2-6.

Quality assessment

All the selected studies were screened to assess the risk of bias using the Standard quality assessment criteria for evaluating primary research papers from various fields (Kmet et al., 2011). The studies were included if they reached a score above 70%. Of the 84 studies, 78 met the criteria for very high-quality studies (total score > 90%), five studies reached the high-quality threshold (total score > 80%) and, only one study reached a total score of 73%. The checklist items that reported the lowest scoring levels were those concerning the description of the subject/comparison group and the adequacy of sample size.

Results

Selection of studies

The flowchart (Figure 1) shows the number of studies identified from the databases and the other sources, the number of studies examined by the authors, and assessed for eligibility. The reasons for exclusion are reported.

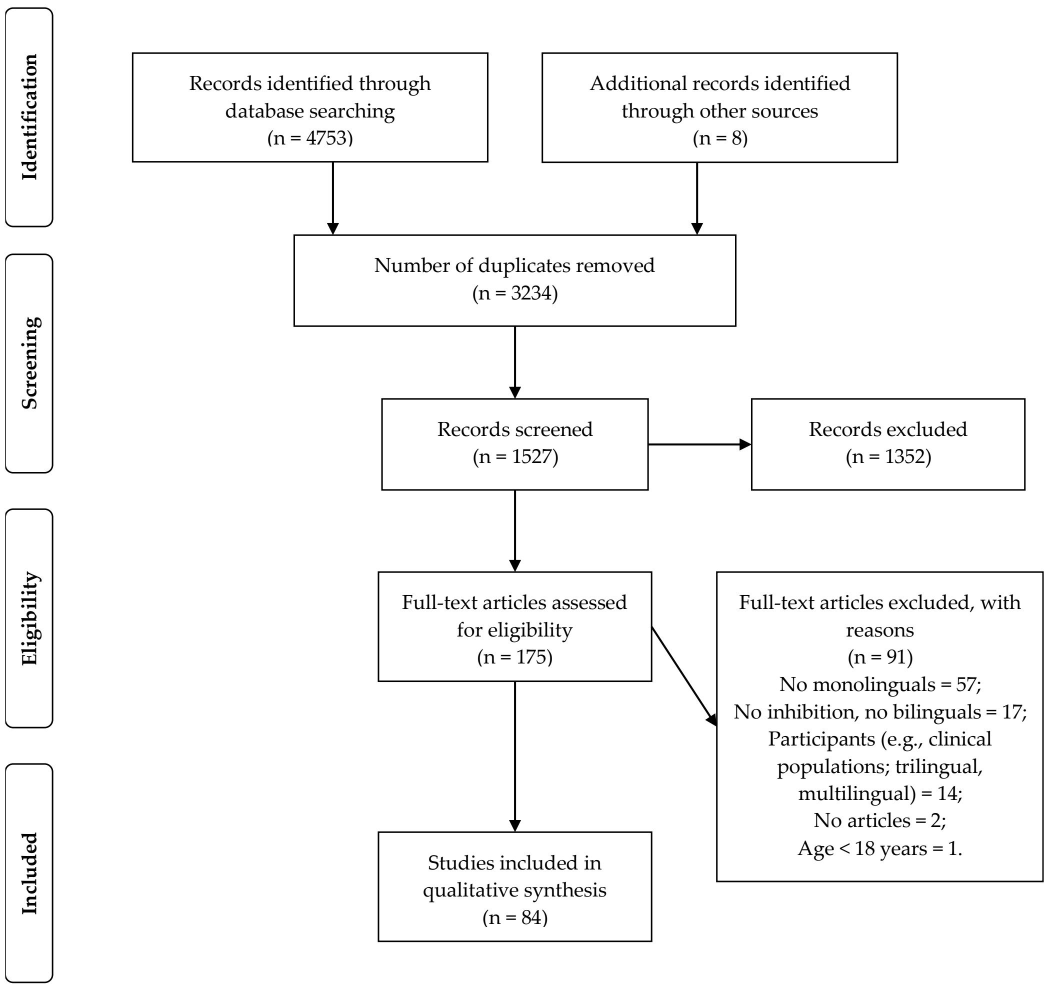


Figure 1. Studies selection flow diagram (PRISMA flow chart).

Table 1. Participants' characteristics in the included studies.

	Participants								
	Group	N	Age mean (SD)	Sex (% female)	Education	SES	Country	Language	AoA
Anderson et al., 2017	Total	35					Canada		
	M	n.r.	74.9 (4.6)		17.5 (4.0)			ENG	
	B	n.r.	74.7 (3.9)		16.7 (2.7)			ENG – ML2	
Ansaldo et al., 2015	M	10	74.5 (7.1)	60	16.1 (3.3)		Canada	FR	
	B	10	74.2 (7.4)	60	17.2 (3.1)			FR – ENG	L2: 16.7 (8.5)
Antón et al., 2019	M	90	21.8 (3.0)	74	M = B	M = B	Basque country	SPA	
	B	90	22.3 (2.9)	75				BAS - SPA	BAS: 1.0 (1.3); SPA: 1.1 (1.7)
Antón et al., 2016	Exp 1A-B						Basque country		
	M	24	68.7 (4.4)	58				SPA	
	B	24	69.4 (4.6)	58				BAS – SPA	
Barbu et al., 2020	M	28	27.9 (7.2)	82	15.4 (1.0)	M = HFLS = LFLS		FR	
	HFLS	30	25.7 (6.1)	80	15.6 (2.8)			FR – DT	
	LFLS	21	24.9 (6.6)	90	15.4 (3.0)			FR - DT	
Bellegarda & Macizo, 2021	M	24	22.0 (3.8)	50	17.0 (1.8)		Spain	SPA	
	B	24	26.6 (7.3)	87	18.1 (2.5)			ML - SPA	L2: 46% 8.4 (4.6); 54% 18.7 (3.9)
Berroir et al., 2017	M	10	74.5 (7.1)	60	16.1 (3.3)		Canada	FR	
	B	10	74.2 (7.4)	60	17.2 (3.1)			FR - ENG	L2: 16.7 (8.5)
Bialystok & DePape, 2009	M	24	23.5 (3.8)					ENG	
	B	24	22.7 (3.7)					ENG - ML	
Bialystok, Craik, et al., 2005	M	10	22-36 y	70			Canada	ENG	
	FEB	10	22-36 y	70				FR – ENG	
	CEB	9	22-36 y	70				CH – ENG	
Bialystok et al., 2004	Study 1								
	MM	10	43.0 (7.3)	50			Canada	ENG	
	MB	10	43.0 (7.3)	50			India	TAM – ENG	
	OM	10	71.6 (7.5)	50			Canada	ENG	
	OB	10	72.3 (8.7)	50			India	TAM – ENG	
	Study 2								
	MM	32	42.6 (8.8)	50			Canada	ENG	
	MB	32	42.6 (8.8)	50			India, China	ENG – TAM/CAN	

	OM	15	70.4 (5.6)	50		Canada	ENG	
	OB	15	70.2 (6.9)	50		India, Canada	ENG – FR/CAN	
	Study 3							
	M	10	38.8 (8.5)	50		Canada	ENG	
	B	10	40.6 (8.1)	50		Canada	ENG – FR	
Bialystok et al., 2009	YM	24	20.7		12.8 (1.3)	Canada	ENG	
	YB	24	19.7		12.4 (0.9)		ENG – ML	
	OM	24	67.2		14.4 (1.4)		ENG	
	OB	24	68.3		14.2 (2.4)		ENG - ML	
Bialystok et al., 2005	Study 3							
	YM	40					ENG	
	YB	56					ML1 - ENG	
	Study 4	40						
	MM		30-59 y			Canada	ENG	
	MB		30-59 y			India	ENG – ML	
	OM		60-80 y			Canada	ENG	
	OB		60-80 y			India	ENG – ML	
	Study 5	94						Study 5
	MM		30-59 y				ENG	L1, L2 < 6 y
	MB		30-59 y			China, India,	ENG – ML	
	OM		60-80 y			Canada	ENG	
	OB		60-80 y				ENG – ML	
						China, India, Can.		
Bialystok et al., 2014	Exp 1							
	YM	27	20.3 (2.5)				ENG	
	YB	44	20.4 (2.2)				ENG – ML	L2: 5.0 (n.r.)
	OM	25	71.3 (5.1)				ENG	
	OB	34	67.6 (4.5)				ENG – ML	L2: 8.0 (n.r.)
Blumenfeld & Marian, 2014	Exp 1							
	M	30	21.4 (3.9)	80		Midwestern United	ENG	
	B	30	22.0 (5.2)	70		States	ENG – SPA	SPA: 2.9 (3.8), ENG: 1.4 (0.3)
	Exp 2							
	M	60	22.2 (3.8)	87			ENG	

	B	60	21.7 (3.1)	90				SPA – ENG	SPA: 1.0 (0.2); ENG: 4.4 (2.8)
Bogulski et al., 2015	M	27	20.8 (1.9)					ENG	
	LapB	22	21.7 (4.1)					ENG – FR	
	FullB	30	21.4 (2.3)					ENG – FR	
Botezatu et al., 2021	M	24	20.8 (n.r.)	42			Pennsylvania	ENG	L2 > 6 y
	B	24	22.3 (n.r.)	12				ENG – SPA / CH	
Chabal et al., 2015	M	17	23.4 (5.1)	35				ENG	
	B	21	20.1 (2.9)	43				ENG – ML	L1 and L2 < 7 y
Chrysochoou et al., 2020	M	27	29.7 (8.0)			M = B	Greece	GRE	
	B	27	31.5 (8.6)					FR - GRE	L1, L2: 92% f.b.; 8% < 3 y
Clare et al., 2016	M	49	72.5 (8.1)	57	M = B	M = B	UK	ENG	
	B	50	74.3 (9.0)	62				ENG - WEL	
Coderre et al., 2016	M	15	25 (3)	60			Maryland	ENG	
	B	14	24 (6)	57				SPA – ENG	ENG: 6 (4)
Coderre et al., 2014	M	18	21 (2)	50			UK	ENG	first L2 contact
	GEB	19	26 (6)	58				DT – ENG	L2: 9.6 (2.3)
	PEB	22	25 (5)	59				POL – ENG	L2: 8.9 (3.1)
	AEB	17	26 (4)	53				ARA - ENG	L2: 7.9 (4.5)
Costa et al., 2008	M	100	22	85			Spain	SPA	
	B	100	22	87				CAT - SPA	CAT: 0.8 (0.6); SPA: 0.3 (0.3)
Costa et al., 2009	Exp 1						Spain		
	8% con								
	M	30	19.5	77				SPA	
	B	30	19.9	90				CAT – SPA	
	92% con								
	M	30	20.5	73				SPA	
	B	30	20.3	83				CAT – SPA	
	Exp 2								
	50% con								
	M	31	20.4	87				SPA	
	B	31	19.9	87				CAT – SPA	
	75% con								
	M	31	20.9	90				SPA	
	B	31	20.3	90				CAT - SPA	
Costumero et al., 2015	M	19	20.5 (2.9)	53	all undergraduate			SPA	
	B	20	21.1 (1.4)	60				CAT - SPA	CAT: 0.8 (0.6); SPA: 0.3 (0.3)

Cox et al., 2016	M	64	74.4 (0.3)	0		M < B	Scotland	ENG	
	B	26	74.5 (0.3)	0				ENG – ML	
Damian et al., 2018	M	25			all		UK	ENG	
	B	26			undergraduate			ENG – CAN/MAN	CH: 6.1 (3.5)
de Bruin et al., 2015	M	24	70.2 (5.7)	67	11.9 (1.7)	M = B	Scotland	ENG	
	acB	24	71.9 (7.1)	68	12.6 (2.8)			GAE– ENG	
	inacB	24	70.5 (7.7)	71	13.2 (3.6)			GAE - ENG	
De Leeuw & Bogulski, 2016	M	14	25.1 (6.1)					ENG	
	B	28	25.8 (5.5)					ENG – SPA	L2: 6.1 (2.3)
Desideri & Bonifacci, 2018	M	25	26.4 (7.1)	72			Italy	ITA	
	B	25	26.5 (7.8)	64				ITA - DT	
Desjardins & Fernandez, 2018	M	20	23.8 (3.8)		16.1 (3.2)		Texas	ENG	
	B	19	22.1 (5.9)		16.0 (2.1)			ENG – SPA	ENG: 4.1 (2.2); SPA: 1.2 (1.3)
Desjardins et al., 2020	YM	15	21.1 (1.9)		15.4 (1.5)	M = B	Texas	ENG	
	YB	16	21.4 (1.9)		14.8 (1.3)			ENG – SPA	SPA: 1.4 (1.6)
	OM	15	55.8 (5.2)		15.2 (2.3)			ENG	
	OB	15	55.9 (4.1)		15.6 (3.3)			ENG – SPA	SPA: 2.3 (2.7)
Fernandez et al., 2014	M	17	20.4 (2.4)	76	14.3 (1.2)		Florida	ENG	
	B	18	22.1 (3.8)	72	15.1 (1.2)			ENG – SPA	L2: 6.2 (4.8)
Fernandez et al., 2013	M	15	22.7 (4.7)	93	15.1 (2.0)		Florida	ENG	
	B	13	20.5 (1.5)	92	14.3 (1.2)			ENG - SPA	L2: 6.0 (2.2)
Garraffa et al., 2017	M	29	38.6 (6.6)				Italy	ITA	
	SECM	18							
	UNIM	11							
	B	34	39.7 (6.5)					ITA - SAR	
	SECB	15							
	UNIB	19							
Gathercole et al., 2014	Tot YP	85	25.5				UK		
	YM	20						ENG	
	YOEH	19						ENG – WHE	
	YWEH	23						ENG – WHE	
	YOWH	23						ENG – WHE	
	Tot OP	84	67.6						
	OM	20						ENG	
	OOEH	23						ENG – WHE	
	OWEH	17						ENG – WHE	

	OOWH	24						ENG – WHE	
Grundy et al., 2017	Study 1					M = B			
	M	28	19.1 (1.5)					ENG	
	B	31	19.3 (1.9)					ML – ML	L2: 1.7 (2.9)
	Study 2					M = B			
	M	53	21.8 (3.3)					ENG	
	B	58	22.0 (4.1)					ML – ML	
	Study 3					M = B			
	M	55	22.7 (3.1)					ENG	
	B	56	20.6 (2.7)					ML – ML	
Guðmundsdóttir & Lesk, 2019	Exp 1								
	All	132	29.9 (13.8)	72					
	M	40						ENG	
	B	58						ML – ML	
	T	34						ML – ML - ML	
Heidlmayr et al., 2015	M	22	25.5 (4.4)	59			France	FR	
	B	22	26.9 (5.5)	73				FR – DT	L2: 10.6 (0.7)
Heidlmayr et al., 2014	M	31	25.2 (4.1)	71		M = B	France	FR	
	B-L1	16	26.8 (3.7)	75			France, Germany	FR – DT	DT: 11.0 (1.2)
	B-L2	17	32.4 (5.2)	88			France, Germany	FR – DT	
Hernández et al., 2010	Exp 1				all		Spain		
	M	41	21.4 (n.r.)	90	undergraduate			SPA	
	B	41	20.9 (n.r.)	76				CAT - SPA	
Hofweber et al., 2020a	M	29	31.2 (13.3)		M = B		UK	ENG	
	B	29	34.2 (10.4)					DT - ENG	ENG: 9.8 (4.3)
Hofweber et al., 2020b	M	41	33.8 (11.8)		M = B		UK	ENG	
	B	43	32.1 (9.6)					DT - ENG	ENG: 8.8 (4.4)
Hui et al., 2020	M	12	67.8 (3.9)	42	M < B		China	CAN	ENG: 12.1 (12.4)
	B	38	67.0 (4.5)	39				CAN - ENG	ENG: 7.9 (3.5)
Kazemeini & Fadardi, 2016	M	30	25.9 (3.4)	0	M = B			PER	
	B	30	25.9 (4.4)	0				KUR - PER	PER: 4.9 (0.5)
Kirk et al., 2014	M	16	69.5 (8.6)	37		M = B	UK, Scotland	ENG	
	MD	16	69.7 (7.7)	31			Scotland	ENG – DUN	
	BD	16	72.4 (8.2)	44			Scotland	ENG - DUN	
	GAEB	16	69.8 (5.5)	37			Scotland	GAE - ENG	

	ASEB	16	72.6 (7.3)	62		UK, Scotland	ASL - ENG	
Kousaie & Phillips, 2012	YM	38	22.5 (4.5)	50	YM = YB	Canada	ENG	85% FR < 8y, 15% L1, L2 f.b.
	YB	35	23.7 (4.0)	68	OM < OB		ENG – FR	
	OM	25	68.9 (6.5)	76			ENG	
	OB	20	71.9 (5.9)	65			ENG – FR	
Kousaie & Phillips, 2017	M	21	71.7 (6.8)	86	M = B	Canada	ENG	
	B	22	68.7 (5.2)	68			FR – ENG	L2: 4.9 (5.1)
Kousaie et al., 2014	Total					Canada		
	YEM	40	21.5 (1.5)	37	15.5 (1.1)		ENG	
	YFM	30	21.8 (2.5)	33	15.1 (1.4)		FR	
	YB	51	21.5 (2.3)	35	15.5 (1.5)		FR – ENG	
	OEM	31	72.3 (6.4)	48	15.3 (2.9)		ENG	
	OFM	30	72.6 (6.6)	77	16.2 (2.6)		FR	
	OB	36	70.7 (2.6)	47	16.1 (2.8)		FR - ENG	
Kramer et al., 2015	M – control			M = B		Brazil	BPOR	
	EB	38						
	YM	10	20.6		13.0			
	MM	14	46.0		9.9			
	OM	14	72.6		5.6			
	M – control							
	LB	14			12.6		BPOR	
	EB	38						
	YB	10	22.6		13.4		BPOR – HUN	
	MB	14	43.5		12.8			
	OB	14	72.0		5.3			
	LB	14	22.5		14.4		BPOR - ENG	
Kuipers & Westphal, 2021	Exp 1					Scotland		
	M	18	18-25 y	M = B			ENG	
	B	18	18-25 y				ENG – ML	L2: 4.1 (3.1)
	Exp 2							
	M	17	20.0 (2.0)	M = B			ENG	
	B	18	22.0 (3.8)				ENG - ML	
Lee Salvatierra &	YM	66	25.9 (6.4)		16.7 (2.4)	Florida	ENG	

Rosselli, 2011	YB	67	26.7 (6.6)		16.1 (3.0)			ENG – SPA	L2: 11.0 (7.0)
	YbaB	38	26.8 (6.6)		16.3 (2.3)				L2: 10.9 (7.0)
	YunbB	29	26.5 (6.6)		15.8 (3.6)				L2: 11.6 (9.3)
	OM	42	63.4 (8.4)		14.9 (2.6)			ENG	
	OB	58	64.8 (7.3)		15.3 (2.8)			ENG - SPA	L2: 19.7 (15.7)
	ObaB	31	60.5 (5.6)		15.5 (2.7)				L2: 21.5 (16.7)
	OunbB	27	62.7 (7.3)		15.7 (2.8)				L2: 17.7 (14.4)
Luk et al., 2010	M	9	22	89				ENG	
	B	9	20	89				ENG – ML	L1, L2 < 6 y
Marzecová et al., 2013	M	17	20.0 (n.r.)	88				POL	
	B	18	23.5 (n.r.)	50				ML - ML	L2: 1.0 (1.4) age of exp.
Massa et al., 2020	YM	16	23.4 (4.5)		16.9 (2.6)	M = B		FR	
	YB	16	25.6 (3.9)		15.7 (2.3)			FR – ITA	11.1 (8.7)
	OM	16	71.1 (5.9)		16.0 (2.7)			FR	
	OB	16	72.3 (5.0)		15.1 (2.4)			FR - ITA	10.2 (9.3)
Moreno et al., 2014	nmM	15	23.6 (n.r.)	73	16.4		Canada, United States	ENG	
	mM	13	26.5 (n.r.)	69	18.0			ENG	
	B	15	23.0 (n.r.)	100	16.9			ENG – HEB/FR	L2: 7.2 (n.r.)
Morrison et al., 2019	M	23	19.7 (2.3)	74	14.6 (1.8)		Canada	ENG	FR: 6.1 (1.5)
	B	21	19.7 (1.6)	71	14.7 (1.5)			ENG – FR	FR: 4.0 (2.3)
Morrison et al., 2020	M	26	20.2 (2.2)	61	15.0 (1.8)		Canada	ENG	FR: 6.1 (1.6)
	B	28	20.5 (2.1)	71	15.2 (1.8)			ENG – FR	FR: 4.6 (2.7)
Nair et al., 2017	M	18	50.8 (4.2)		M = B	M = B	India	MAL	
	B	18	51.2 (3.9)					MAL - TUL	TUL: 5.0 (0.4)
Okada e al., 2019	All	70	18 – 22 y	61			California		
	M	31						ENG	
	B	30						ENG - ML	L2 < 5 y
Ooi et al., 2018	M	64	21.4 (3.0)	73			Edinburgh	ENG	
	ELB	63	22.1 (3.0)	75			Edinburgh	ENG – ML	L2: 12.9 (4.7)
	EEB	48	21.3 (3.1)	73			Edinburgh	ENG – ML	L2: 1.2 (1.6)
	SB	70	21.8 (2.2)	71			Singapore	ENG – ML	L2: 0.7 (1.6)
Ouzia et al., 2019	M	31	22.3 (3.7)	61				ENG	
	B	30	25.3 (4.5)	57				ENG - ML	0-6 y: 45%; 7-12 y:29%; 13-18 y: 23%
Paap et al., 2015	M	84	25					ENG	
	HPB	120	25					ENG – ML	L2: 4.9

Patra et al., 2020	M	25	30.4 (8.2)	52	M = B		UK	ENG	
	B	25	32.8 (4.8)	44				BEN - ENG	
Pelham & Abrams, 2014	M	30	19.3 (2.2)	63	14.1 (2.7)		Florida	ENG	
	LB	30	22.0 (3.4)	53	15.8 (2.6)			ENG – SPA	L2: 11.6 (4.6)
	EB	30	20.3 (1.8)	53	15.6 (1.7)			ENG - SPA	L2: 3.3 (2.2)
Rayney et al., 2021	Total	32	19.3 (1.5)						
	M	16		81				ENG	
	B	16		81				ENG - SPA	
Rodrigues & Zimmer, 2016	M	20	47.2	25	18.6		Brazil	POR	
	B	20	48.1	30	18.4			ML – ML	
Rosselli et al., 2016	All M	40		87					
	HPM	20	27.3 (7.2)		15.6 (1.2)			ENG	
	LPM	20	24.3 (5.2)		15.2 (1.0)			ENG	
	All B	74		84					
	HPB	20	26.8 (7.6)		14.9 (1.2)			SPA – ENG	L2: 6.6 (6.3)
	LPB	20	25.2 (4.9)		14.9 (1.1)			SPA – ENG	L2: 8.2 (5.0)
	UnbB	34	26.9 (6.8)		14.7 (1.1)			SPA – ENG	L2: 10.4 (8.1)
Sanchez-Azanza et al., 2020	M	49	19.9 (1.9)	88	M < B		Spain	SPA	
	B	41	22.9 (2.2)	71				CAT – SPA	SPA (L2): 2.6 (1.8)
Shulley & Shake, 2016	M	52	20.1 (2.2)		13.4 (1.6)		Kentucky	ENG	L2: 14.7 (3.0)
	B	52	21.3 (1.9)		14.2 (1.4)			ENG - ML	L2: 12.1 (6.8)
Soares et al., 2019	M	27	61.9 (6.4)	52	12.6 (n.r.)	M = B	Italy	ITA	
	B	29	63.4 (5.8)	55	13.5 (n.r.)		China		L2: 18.0 (n.r.)
Taler et al., 2013	M	32	21.6 (1.6)	47	15.6 (1.1)		Canada	ENG	
	B	38	21.5 (2.3)	39	15.5 (1.6)			ENG – FR	L1, L2 < 13 y
Tao et al., 2011	M	34	20.4 (5.5)	53	all 1st year of university	EB < LB < M	Australia	ENG	
	EB	36	20.8 (2.5)	56				CH – ENG	L2: 2.9 (1.8)
	LB	30	18.9 (1.3)	63				CH - ENG	L2: 7.8 (3.7)
Treffers-Daller et al., 2020	M	30	32.3 (10.1)		3.9 (0.6)		UK	ENG	
	TurkeybornB	29	32.5 (7.9)		2.6 (1.0)			TUR -ENG	
	CyprusbornB	28	25.5 (4.0)		3.0 (0.8)			TUR - ENG	
Vivas et al., 2017	M	45	29.7 (9.5)	64		M = B	Greece	GRE	
	B	45	27.3 (7.0)	64				GRE - ALB	L2: 9.9 (8.4)
Waldie et al., 2009	M	10	28.4 (7.4)	50			New Zealand	ENG	
	B	8	24.9 (2.7)	37				MAC - ENG	ENG (L2): 11.2 (2.5)
Warmington et al., 2019	Exp 1						UK		
	M	23	23 (4m)	65				ENG	

	B	23	23 (7m)	61				ENG – HIN	ENG: 3y (5m)
	Exp 2								
	M	20	21 (7 m)	50				ENG	
	B	20	23 (5 m)	30				ENG - HIN	ENG: 3y (7m)
Woumans et al., 2015	M	30	22.1 (1.4)	73		M = B	Belgium	FR	
	unbB	34	22.3 (2.8)	79				NL – FR	L1: f.b.; L2: 9.4 (1.3)
	baB	31	21.1 (2.1)	77				NL – FR	L1: f.b.; L2: 2.6 (3.0)
Woumans et al., 2019	M	16	18.6 (0.6)	94		EM > CM, B	Belgium	NL	NL: f.b.; L2: 12.7 (1.3)
	B	18	19.8 (4.8)	83				NL - ML	NL: f.b.; L2: 4.7 5.2)
Xie & Dong, 2017	M	33	21.5 (3.6)		11.3 (2.4)		China	CH	L2: 4.1 (1.2)
	B	33	21.7 (0.8)		15.7 (0.8)			CH - ENG	L2: 11.7 (1.7)
Xie & Zhou, 2020	EM	26	37.6 (19.2)		14.2 (3.9)	EM > CM, B	New York	ENG	
	CM	31	21.5 (3.6)		11.3 (2.5)		China	CH	
	B	34	22.3 (1.8)		16.3 (1.9)		China	CH - ENG	
Yang & Yang, 2016	M	19	19.11 (0.9)	89		M = B	New York	ENG	
	B	20	20.8 (1.2)	75				KOR - ENG	
Yudes et al., 2011	Exp 2						Spain		
	M	16	21.6 (2.9)	69				SPA	
	B	16	25.7 (3.2)	62				SPA - ENG	
Zeng et al., 2019	Exp 2						Australia		
	M	20	23.1 (3.4)	45				ENG	
	B	20	22.5 (3.4)	70				ML - ENG	ENG: 6.8 (2.4)
	Exp 3								
	M	18	69.1 (6.5)	50				ENG	
	B	17	67.7 (8.1)	76				ML - ENG	ENG: 10.5 (7.8)
Zhou & Krott, 2018	M	29	21.0 (3.0)	65	M = B		UK	ENG	
	B	29	21.6 (3.2)	41			UK, China	ENG - CH	ENG: 3.3 (1.5); CH: 1.3 (1.8)

Note. N: number of participants; SES: socioeconomic status; AoA: age of acquisition; M: monolinguals; B: bilinguals; n.r.: not reported; ENG: English;; FR: French; SPA: Spanish; BAS: Basque; HFLS: high frequency switchers; LFLS: low frequency switchers; DT: German; YM: young monolinguals; YB: young bilinguals; OM: old monolinguals; OB: old bilinguals; TAM: Tamil; CAN: Cantonese; ML: multiple languages; FEB: French-English bilinguals; CEB: Chinese-English bilinguals; CH: Chinese; LapB: lapsed bilinguals; FullB: full bilinguals; GRE: Greek; f.b.: from birth; WEL: Welsh; GEB: German-English bilinguals; PEB: Polish-English bilinguals; AEB: Arabic-English bilinguals; POL: Polish; ARA: Arabic; CAT: Catalan; MAN: Mandarin; acB: active bilinguals; inacB: inactive bilinguals; GAE: Gaelic; ITA: Italian; SECM: secondary school degree monolinguals; UNIM: university degree monolinguals; SECB: secondary school degree bilinguals; UNIB: university degree bilinguals; SAR: Sardinian; YP: young adults participants; YOEH: young bilinguals only English at home; YWEH: young bilinguals Welsh and English at home; YOWH: young bilinguals only Welsh at home; OP: older adults participants; OOEH: older bilinguals only English at home; OWEH: older bilinguals Welsh and English at home; OOWH: older bilinguals only Welsh at home; T: trilinguals; B-L1: bilinguals L1 environment; B-L2: bilinguals L2 environment; PER: Persian; KUR: Kurdish; MD: monodialectal; BD: bidialectal; GAEB: Gaelic-English bilinguals; ASEB:; Asian languages-English bilinguals; DUN: Dundonian; YEM: young English monolinguals; YFM: young French monolinguals; OEM: old English monolinguals; OFM: old French monolinguals; EB: early bi-linguals; MM: middle-aged monolinguals; LB: late bilinguals; MB: middle-aged bilinguals; YbaB: young balanced bilinguals; YunbB: young un-balanced bilinguals; ObaB: old balanced bilinguals; OunbB: old unbalanced bilinguals; nmM: non musicians; mM: musicians monolinguals; HEB: Hebrew; MAL: Malayalam; TUL: Tulu; EEB: Edinburgh early bilinguals; ELB: Edinburgh late bilinguals; SB: Singapore early bilinguals; HPB: high proficiency bilinguals; HPM: high proficiency monolinguals; LPB: low proficient bilinguals; LPB: low proficient bilingua1s; HIN: Hindi; FM: French monolinguals; CM: Chinese monolinguals; NL: Dutch; KOR: Korean.

Flanker task and Attention Network test (n = 32)

Eighteen studies used the Flanker task to assess cognitive inhibition. Seventeen studies included young adult participants (Antón et al., 2019; Bogulski et al., 2015; Botezatu et al., 2021; Coderre et al., 2016; Damian et al., 2018; De Leeuw & Bogulski, 2016; Grundy et al., 2017; Hofweber et al., 2020a; Hofweber et al., 2020b; Kuipers & Westphal, 2021; Luk et al., 2010; Paap et al., 2015; Sanchez-Azanza et al., 2020; Treffers-Daller et al., 2020; Xie & Dong, 2017; Xie & Zhou, 2020; Warmington et al., 2019; Zhou & Krott, 2018) while one study included older adult participants (Kousaie & Phillips, 2017). Results are summarized in Table 2.

Table 2. Results of studies that used the Flanker task.

	Flanker task		
	RT	**ACC**	**Flanker effect**
Antón et al., 2019	M = B	M = B	M = B
Bogulski et al., 2015	M = B		
Botezatu et al., 2021	M = B	M = B	
Coderre et al., 2016	M = B		
Damian et al., 2018	M = B		
De Leeuw & Bogulski, 2016	M = B		M = B
Hofweber et al., 2020a	M = B	M = B	M > B (92-8 condition) M = B (75-25, 50-50 conditions)
Hofweber et al., 2020b	Inc. trials M > B (92-8 condition) M = B (75-25, 50-50 conditions) Cong. trials M = B (all conditions)	M = B	
Grundy et al., 2017	Study 1, 2, 3 M = B		
Kousaie & Phillips, 2017	M = B	M < B	
Kuipers & Westphal, 2021	M = B	M < B	
Luk et al., 2010	M = B		
Paap et al., 2015[1]	M = HB		M = HB
Soares et al., 2019	M = B		
Sanchez-Azanza et al., 2020	M = B	M = B	M = B
Treffers-Daller et al., 2020[2]	ANCOVA[3] M > CB M = TB TB = CB Matched sample M = CB = TB		ANCOVA M > TB > CB Matched sample M > CB CB < TB M = TB
Warmington et al., 2019	M = B	M = B	M = B
Xie & Zhou, 2020	Cong. and neutral trials EM > CM, CEB CM = CEB Inc. trials EM = CM = CEB		
Xie & Dong, 2017	M = B		M = B
Zhou & Krott, 2018[4]	M = B	M = B	M < B

[1] Mixing cost: M = HB. [2] Proportion score (flanker effect / mean RT congruent trials): M > TB > CB. Matched sample: M > CB, M = TB, CB = TB. [3] Covariates: age, education, working memory, non-verbal reasoning. [4] τ parameter: M = B; μ parameter: M > B (incongruent trials), M = B (congruent trials).

Note. RT: reaction times; ACC: accuracy; M: monolinguals; B: bilinguals; HPB: high proficiency bilinguals; CB: Cyprus born bilinguals; TB: Turkey born bilinguals; EM: English monolinguals; CM: Chinese monolinguals.

Thirteen studies used the flanker task within the Attention Network Test (ANT). Ten studies included young adult participants (Costa et al., 2008, 2009; Desideri & Bonifacci, 2018; Marzecová et al. 2013; Ooi et al., 2018; Pelham & Abrams, 2014; Tao et al., 2011; Vivas et al., 2017; Woumans et al., 2015; Yang & Yang, 2016), two studies middle-aged participants (Nair et al., 2017; Rodrigues & Zimmer, 2016), while one study older adult participants (Soares et al., 2019). Results are summarized in Table 3.

Table 3. Results of the studies that used the Attention Network Test.

	Attention Network Test		
	RT	**ACC**	**Conflict effect**
Costa et al., 2008	M > B	M = B	M > B
Costa et al., 2009	Exp 1 M = B Exp 2 M > B	Exp 1 M = B Exp 2 M = B	
Desideri & Bonifacci, 2018	M > B (overall, inc. and neutral trials) M = B (cong.)	M = B	M > B (RT) M = B (ACC)
Marzecová et al., 2013	M = B	M < B	M > B (RT) M = B (ACC)
Nair et al., 2017[1]	M > B	M = B	
Ooi et al., 2018	M > EEB, SB ELB > SB	M = EEB = SB	M > SB ELB > SB
Pelham & Abrams, 2014	M > (LB = EB) (inc.)	M = EB = LB	
Rodrigues-Zimmer et al., 2016	M = B	M = B	M = B
Soares et al., 2019	M = B (inc., neu) M > B (cong.)		M = B (RT)
Tao et al., 2011	ANCOVA[2] M > EB M = LB LB = EB	M = EB = LB	RT M > (EB = LB) ACC M = EB M > LB EB > LB
Vivas et al., 2017	M < B	M = B	M = B
Woumans et al., 2015	M > unbB, bB	M = B	M = B
Yang & Yang 2016[3]	M > B (overall, inc., cong) M = B (neu)	M = B	M > B

[1]Inverse efficiency score (mean RT / % ACC): M > B. [2]Covariates: age, Raven, parental education. [3]Inverse efficiency score (mean RT / % ACC): M > B.

Note. RT: reaction times; ACC: accuracy; M: monolinguals; B: bilinguals; EEB: Edinburgh early bilinguals; ELB: Edinburgh late bilinguals; SB: Singapore early bilinguals; LB: late bilinguals; EB: early bilinguals; UnbB: unbalanced bilinguals; bB: balanced bilinguals.

Young adults (18-34y) (n = 27)

Considering accuracy, monolinguals were less accurate than bilingual participants in three studies (Kuipers & Westphal, 2021; Marzecová et al., 2013; Sanchez-Azanza et al., 2020). No significant differences emerged in the other studies, or no accuracy analyses were conducted.

Considering reaction times (RTs), sixteen studies showed no significant differences between monolinguals and bilinguals. In twelve studies (Costa et al., 2008; Costa et al., 2009; Desideri & Bonifacci, 2018; Hofweber et al., 2020a; Ooi et al., 2018; Pelham &Abrams, 2014; Tao et al., 2011; Treffers-Daller et al., 2020; Xie & Zhou, 2020; Yang & Yang, 2016; Vivas et al., 2017; Woumans et al., 2015), significant differences emerged between monolingual and bilingual participants. Hofweber et al. (2020a) found that monolinguals had slower RTs than bilinguals in the incongruent trials in a condition considering a high percentage of congruent trials (i.e., 92% congruent trials, 8% incongruent trials), whereas no significant differences emerged in other conditions (75% congruent trials and 25% incongruent trials; 50% congruent trials and 50% incongruent trials). Moreover, there was no significant difference in any condition for congruent trials. In Treffers-Daller et al.'s study (2020), monolinguals were slower than the two groups of bilinguals included in the study (Cyprus-born bilinguals) in the incongruent trials. In contrast, there was no significant difference in Turkish-born bilinguals and the congruent trials. The RTs of the two bilingual groups did not differ. However, no between-group differences emerged when comparing groups of participants equivalent in age, education, working memory, and non-verbal reasoning. Xie and Zhou (2020) did not evidence significant differences in the incongruent trials, while bilinguals were faster than English monolinguals but not compared to Chinese monolinguals in the congruent and neutral trials.

In Costa et al.' study (2009), bilinguals were faster than monolinguals only in Experiment 2. Desideri and Bonifacci (2018) found that bilinguals were faster in neutral and incongruent trials, while Yang and Yang (2016) observed faster RTs in congruent and incongruent trials but not in neutral trials. In Tao et al.'s study (2011), only the early bilingual group was faster than monolinguals, while no significant differences emerged between monolinguals and late bilinguals and between the two groups of bilinguals. Monolinguals were slower than Edinburgh early bilinguals and Singapore early bilinguals but not than Edinburgh late bilinguals (Ooi et al., 2018), slower only in incongruent trials (Pelham and Abrams, 2014), or overall slower than bilinguals (Woumans et al., 2015). Conversely, Vivas et al. (2017) found that monolinguals were faster than bilinguals.

Nine studies showed a greater flanker effect in the monolingual group than in the bilingual group. However, Hofweber et al. (2020a) found that this difference emerged when considering 92% of congruent trials and 8% of incongruent trials. Treffers-Daller et al. (2020) confirm a higher flanker effect in monolinguals than bilinguals but highlighted that this effect was modulated by age, education, working memory, and non-verbal reasoning. Hofweber et al. (2020b) observed a greater flanker effect in monolinguals than bilinguals only when the index was calculated considering accuracy but not RTs. Conversely, two studies (Desideri & Bonifacci, 2018; Marzecová et al., 2013) showed a greater flanker effect for RTs, but not for accuracy in monolinguals compared to bilinguals. Also, in Tao et al.'s study (2011), monolinguals presented a high flanker effect than bilinguals for RTs, while these results were confirmed only considering late bilinguals when accuracy was considered. Ooi et al. (2018) found a greater flanker effect in monolinguals than Singapore early bilinguals; no difference was found considering the other groups of bilinguals considered (Edinburgh early and late bilinguals).

Several studies have further analyzed the results by calculating other indices. An advantage in conflict management for bilinguals compared to monolinguals was found considering a proportion score (flanker effect/mean RTs congruent trials) (Traffers-Daller et al., 2020), incongruity index (neutral trials compared to incongruent trials), and congruity index (congruent trials compared to neutral trials) (Anton et al., 2019). Paap et al. (2015) did not find differences between monolinguals and bilinguals considering the mixing cost (i.e., the difference in mean RTs between blocks of single-task trials and the mean on repeated trials). No difference between groups was observed considering the t parameter (i.e., the tail of the RTs distribution). At the same time, monolinguals had a higher μ parameter (i.e., the major delay in response when a subject encountered interference) in the incongruent condition. (Zhou and Krott, 2018). Finally, Yang and Yang (2016) found that bilinguals were more efficient than monolinguals considering an efficiency score that simultaneously considered RTs and accuracy.

Middle-aged adults (35-64y) (n = 2)

Rodrigues and Zimmer (2016) did not find significant differences between monolinguals and bilinguals in any aspect considered (RTs, ACC, flanker effect). Nair et al. (2017) found that monolinguals were slower than bilinguals, while no significant differences in accuracy emerged between the groups.

Older adults (> 65y) (n = 2)

Kousaie and Philips (2017) did not observe significant differences in RTs between groups, while monolinguals were less accurate than bilinguals. Soares et al. (2019) found that monolinguals were slower than bilinguals in congruent trials, while no differences emerged in neutral and incongruent trials and the flanker effect.

General results

Considering all studies that used the Flanker task and the ANT, no significant differences emerged between bilinguals and monolinguals in fourteen studies. In seventeen studies, at least one difference emerged in favor of the bilingual group. Specifically, in seven studies, differences emerged only relative to RTs, in two studies only in the percentage of accuracy, in Hofweber et al. (2020a) only in the Flanker effect, while in seven studies relative to the conflict effect and RTs or accuracy. In two studies, at least one result emerged in favor of the monolingual group. Specifically, Vivas et al. (2017) showed lower RTs for monolinguals, while Zhou and Krott (2018) evidenced a lower Flanker effect.

Stroop task (n = 34)

Thirty-four studies used the Stroop task for the assessment of cognitive inhibition.
Twenty-six studies included young adult participants (Antón et al., 2019; Bellegarda & Macizo, 2021; Bialystok & DePape, 2009; Bialystok et al., 2008; Bialystok et al., 2014; Blumenfeld & Marian, 2014; Coderre & van Heuven, 2014; Damian et al., 2018; Heidlmayr et al., 2014; Heidlmayr et al., 2015; Hernandez et al., 2010; Kazemeini & Fadardi, 2016; Kousaie & Phillips, 2012; Kousaie et al., 2014; Massa et al., 2020; Morrison et al., 2019; Morrison et al., 2020; Okada et al., 2019; Patra et al., 2020; Rainey et al., 2021; Rosselli et al., 2016; Shulley & Shake, 2016; Taler et al., 2013; Waldie et al., 2009; Xie & Dong, 2017; Zhou & Krott, 2018), one study middle-aged participants (Garraffa et al., 2017) and ten studies older adult participants (Anderson et al. 2017; Ansaldo et al., 2015; Antón et al., 2016; Berroir et al., 2017; Clare et al., 2016; Hui et al., 2020; Kousaie & Phillips, 2017; Kousaie & Phillips, 2012; Kousaie et al., 2014; Massa et al., 2020). Results are summarized in Table 4.

Table 4. Results of the studies that used the Stroop task.

	Stroop task		
	RT	**ACC**	**Stroop effect**
Anderson et al., 2017	M < B (color, inhibition, inhibition/switching) M = B (word)	M > B (inhibition) M = B (inhibition/switching)	
Ansaldo et al., 2015		M = B (cong.) M < B (inc.)	
Antón et al., 2019	M = B	M = B	M = B
Antón et al., 2016[1]	M = B	M = B	M = B
Bellegarda & Macizo, 2021	M = B	M = B	
Berroir et al., 2017			M = B
Bialystok & DePape, 2009	M = B		M = B
Bialystok, Craik, & Luk, 2008	YM = YB OM = OB		YM = YB OM = OB
Bialystok et al., 2014	M < B (color naming)		YM > YB

	M = B (word reading) OM > OB (interference) YM = YB (interference)		OM > OB
Blumenfeld & Marian, 2014	M = B	M = B	M = B
Clare et al., 2016	M = B (colours) M > B (words)		
Coderre & van Heuven, 2014	L1 and L2 M = GEB = PEB = AEB		L1 and L2 M > AEB AEB = GEB GEB = PEB = M L1 PEB > AEB L2 AEB = GEB = PEB
Damian et al., 2018	M > B		
Garraffa, Obregon, & Sorace, 2017			SECM = SECB UNIM = UNIB
Heidlmayr et al., 2014[2]	M = B	M = B	M > B
Heidlmayr et al., 2015	M = B	M = B	
Hernandez et al., 2010[3]	M = B	M = B	M > B (SIE)
Hui et al., 2020		M = B (color, word) M < B (interference)	
Kazemeini & Fadardi, 2016	M = B (cong.) M > B (inc.)	M > B (cong., inc.)	M > B
Kousaie & Phillips, 2012	YM > YB OM = OB	YM = YB OM = OB	YM = YB OM = OB
Kousaie & Phillips, 2017	M < B (inc.) M = B (cong.)	M < B (inc.) M = B (cong.)	
Kousaie et al., 2014		M > B (overall) MA, B < MF (word, colour) MF < MA, B (inc.)	MF > MA, B (ACC, word) MF > MA > B (ACC, color)
Massa et al., 2020	M = B	M > B	M = B
Morrison et al., 2019		M = B	
Morrison et al., 2020		M = B	
Okada et al., 2019	M < B		
Patra et al., 2020	M < B (inc., neu) M = B (cong.)		M = B (SIE)
Rainey et al., 2021	M = B	M = B	M = B
Rosselli et al., 2016		M = B	HPB < LPB M < unbB, LPB unbB = HPB, LPB
Shulley & Shake, 2016	M > B	M = B	
Taler et al., 2013[4]		M = B	
Waldie et al., 2009	M = B	M = B	
Xie & Dong, 2017	M = B		M = B
Zhou & Krott, 2018[5]	M = B	M < B	M > B

[1]Congruity and incongruity effects: M = B. [2]Facilitation effect (neutral trials – congruent trials): M = B. [3]Facilitation effect: M < B. [4]Costs: YM > YB, OM > OB; Facilitation effect: OM < OB. [5]τ parameter: M > B (inc.), M = B (cong.). μ parameter: M = B.

Note. RT: reaction times; ACC: accuracy; M: monolinguals; B: bilinguals; YM: young monolinguals; YB: young bilinguals; OM: old monolinguals; OB: old bilinguals; GEB:; PEB:; AEB:; SECM: secondary school degree monolinguals; UNIM: university degree monolinguals; SECB: secondary school degree bilinguals; UNIB: university degree bilinguals; SIE: (Stroop interference effect, incongruent trials – neutral trials); EM: English monolinguals; FM: French monolinguals; HPB: high proficiency bilinguals; LPB: low proficiency bilinguals; unbB: unbalanced bilinguals.

Young adults (18-34y) (n = 26)

Considering accuracy, fourteen studies showed no significant differences between the groups. In four studies (Bialystok et al., 2009; Kazemeini & Fadardi, 2016; Kousaie et al., 2014; Massa et al., 2020), monolinguals made more errors than bilinguals, while in the study of Zhou and Krott (2018), bilinguals made more errors than monolinguals. Blumenfeld and Marian (2014) found that monolinguals made more errors than bilinguals in the incongruent condition (Experiment 1) while, in experiment two, no significant differences emerged in all conditions.

Considering RTs, fourteen studies showed no significant differences between monolinguals and bilinguals. Bialystok et al. (2014) did not observe significant differences between the groups in the word naming and interference condition, whereas, in the color naming condition, bilinguals were slower than monolinguals. In three studies, bilinguals were faster than monolinguals (Damian et al., 2018; Kousaie & Phillips, 2012; Shulley & Shake, 2016), while Kazemeini and Fadardi's study (2016), only in the incongruent condition. On the contrary, Okada et al. (2019) found that bilinguals were slower than bilinguals, while Patra et al. (2020) observed that bilinguals were slower than monolinguals in the incongruent and neutral conditions but not in the congruent condition.

Considering the Stroop effect, no significant differences emerged in eight studies. In five studies (Bialystok et al., 2014; Heidlmayr et al., 2014; Hernandez et al., 2010; Kazemeini & Fadardi, 2016; Zhou & Krott, 2018), monolinguals have a greater Stroop effect than bilinguals, while Bialystok (2009) found that bilinguals have higher costs than monolinguals. Blumenfeld and Marian (2014) observed a greater Stroop effect in bilinguals than monolinguals in the first experiment, while no significant difference emerged in the second experiment. Coderre and van Heuven (2014) observed a greater Stroop effect in monolinguals than the Arabic-English bilingual group, but not compared other groups of bilinguals (German-English and Polish-English bilingual groups). Kousaie et al. (2014) observed a smaller Stroop effect in the color condition in bilinguals than both monolingual groups considered. A significant difference emerges between bilinguals and Francophone monolinguals in the word condition, while no differences emerged between bilinguals and Anglophone monolinguals.

Four studies (Bialystok & DePape, 2009; Coderre & van Heuven, 2014; Heidlmayr et al., 2014; Hernandez et al., 2010) calculated the facilitation effect (i.e., mean RTs congruent trials minus mean RTs neutral trials). There were no significant differences in two studies (Bialystok & De Pape, 2009; Heidlmayr et al., 2014), while Hernandez et al. (2010) found that monolinguals had a lower facilitation effect than bilinguals. Coderre and van Heuven (2014) showed a higher facilitation effect for monolinguals than German-English bilinguals, while no significant difference emerged with the other groups of bilinguals (Polish-English and Arabic-English bilingual groups). Zhou & Krott (2018) evidenced that the τ parameter in the incongruent condition was higher for monolinguals compared to bilinguals, while no significant differences emerged in the congruent condition. Moreover, no significant differences emerged regarding the μ parameter.

Middle-aged adults (35-64y) (n = 1)

Garraffa et al. (2017) did not find significant differences between the groups in the Stroop effect.

Older adults (>65y) (n = 10)

Considering reaction times, three studies did not evidence significant differences between monolingual and bilingual participants. In two studies (Anderson et al., 2017, Kousaie & Phillips, 2017), bilinguals were slower than monolinguals in the incongruent condition. Five studies did not analyze RTs.

Considering accuracy, two studies did not find significant differences between the groups (Antón et al., 2016; Kousaie & Phillips, 2012). Massa et al. (2020) showed that monolinguals were more accurate than bilinguals. In two other studies (Anderson et al., 2017; Ansaldo et al., 2015), monolinguals provided more correct responses than bilinguals in the incongruent condition. Conversely, in three studies, bilinguals were more accurate in the incongruent condition (Hui et al., 2020; Kousaie & Phillips, 2017; Kousaie et al., 2014). Four studies showed no differences in the Stroop effect between the groups. Two studies (Clare et al., 2016; Kousaie et al., 2014) calculated the Stroop effect separately for the word and color conditions. In both studies, monolinguals showed a greater Stroop effect than bilinguals in the word condition. In the color condition, Clare et al. (2016) did not find significant differences between the groups, while Kousaie et al. (2014) showed a greater Stroop effect for monolinguals than bilinguals.

General results

Considering all studies that used the Stroop task, fifteen studies showed no significant differences between bilinguals and monolinguals. In fourteen studies, at least one difference emerged in favor of the bilingual group. Specifically, in seven studies, there were differences in RTs or accuracy. Four studies only related to the Stroop effect, while three on the conflict effect and RTs or accuracy. There was at least one result favoring the monolingual group in nine studies. Specifically, in seven studies, the difference between groups was relative RTs or accuracy; Rosselli et al. (2016) showed a lower Stroop effect for the monolingual group, while Anderson et al. (2017) showed slower reaction times and higher accuracy.

Simon task (n = 36)

Thirty-six studies used the Simon task to assess motor inhibition.

Twenty-four studies included young adults participants (Antón et al., 2019; Blumenfeld & Marian, 2014; Bellegarda & Macizo, 2021; Bialystok, Craik et al., 2005; Bialystok et al., 2005; Bialystok et al., 2009; Bialystok & DePape, 2009; Chabal et al., 2015; Chrysochoou et al., 2020; Coderre et al., 2014; Damian et al., 2018; Desjardins & Fernandez, 2018; Gathercole et al., 2014; Guðmundsdóttir & Lesk, 2019; Kousaie et al., 2014; Lee Salvatierra & Rosselli, 2011; Ouzia et al., 2019; Paap et al., 2015; Rosselli et al., 2016; Woumans et al., 2015; Woumans et al., 2019; Yudes et al., 2011; Zeng et al., 2019; Zhou & Krott, 2018), four studies middle-aged participants (Bialystok et al., 2004; Nair et al., 2017; Desjardins et al., 2020; Rodrigues & Zimmer, 2016), while fifteen studies older adults participants (Ansaldo et al., 2015; Berroir et al., 2017; Bialystok et al., 2004; Bialystok et al., 2005; Bialystok et al., 2009; Clare et al., 2016; Cox et al., 2016; de Bruin et al., 2015; Gathercole et al., 2014; Kirk et al., 2014; Kousaie et al., 2014; Kousaie & Phillips, 2017; Kramer et al., 2015; Lee Salvatierra & Rosselli, 2011; Zeng et al., 2019). Results are summarized in Table 5.

Table 5. Results of the studies that used the Simon task.

	Simon task		
	RT	**ACC**	**Simon effect**
Ansaldo et al., 2015	M = B	M = B	
Antón et al., 2019	M = B	M = B	M = B
Bellegarda & Macizo, 2021	M = B	M = B	
Berroir et al., 2017	M = B	M = B	M = B
Bialystok & DePape, 2009	M = B (dir., pos.) M > B (opp., confl.)	M = B	
Bialystok, Craik et al., 2005	M = FB > CB		
Bialystok et al., 2004	Study 1 YM = YB OM > OB	Study 1 YM = YB OM = OB (cong.) OM < OB (inc.)	Study 1 YM = YB OM = OB

	Study 2 M > B Study 3 n.r.	Study 2 YM = YB OM < OB Study 3 M = B	Study 2 n.r. Study 3 M = B (blocks 5-7, 10) M > B (blocks 1-4, 8, 99
Bialystok et al., 2009	YM = YB OM = OB	YM = YB OM = OB	YM = YB OM > OB
Bialystok et al., 2005	Study 3 M = B Study 4 M > B Study 5 M > B (cong., inc) M = B (control)		
Blumenfeld & Marian, 2014	M = B	M = B	
Chabal et al., 2015			M = B
Chrysochoou et al., 2020	M = B	M = B	
Clare et al., 2016			M = B
Coderre et al., 2014[1]	M = B AEB > GEB, PEB GEB = PEB		M = GEB = PEB = AEB
Cox et al., 2016		M = B	M > B
Damian et al., 2018	M > B		
de Bruin et al., 2015	M = aB = iB		M = aB = iB
Desjardins & Fernandez, 2018	M = B	M = B	
Desjardins et al., 2020	YM = YB OM = OB		YM = YB OM = OB
Gathercole et al., 2014	YM < YWEH, YOWH YM = YOEH YOEH = YWEH = YOWH OM = OB	YM = YB OM < OB	
Guðmundsdóttir & Lesk, 2019	M = B	M = B	M = B
Kirk et al., 2014	M = GEB = AEB	M = GEB = AEB	M = B
Kousaie & Phillips, 2017	M = B	M = B	
Kousaie & Phillips, 2014	MA = MF = B		(MA > MF) = B
Kramer et al., 2015	M = EB M = LB (cong.) M > LB (inc.)	M = EB = LB	M = EB = LB
Lee Salvatierra & Rosselli, 2011	M > B Simple condition	M = B Simple. Complex	M > B Simple condition

	YB = OB YM < OM YM = YB OM > OB Complex M = B YM = unbB = bB OM > unbB, bB	M = B	YB = OB YM < OM YM = YB OM > OB Complex M = B YM = unbB = bB OM > unbB, bB
Ouzia et al., 2019	M = B[2]	M = B	M = B
Nair et al., 2017	M > B	M = B	M > B
Paap, Johnson, & Sawi, 2015[3]	M = HPB		M < HPB
Rodrigues & Zimmer, 2016	M > B (center 4-color condition, lateral congruent 2-color condition) M = B (lateral incongruent 2-color and 4-color condition)	M = B M > B (lateral congruent 2-color condition)	M = B
Rosselli et al., 2016	HM = LM = HB = LB = unbB		LB > HM HM = LM = HB = unbB LM = HB = LB = unbB
Woumans et al., 2015	M = unbB = bB	M = unbB = bB	M > bB =unbB
Woumans et al., 2019	M = B	M = B	M = B
Yudes et al., 2011	M = B	M = B	M = B
Zeng et al., 2019	YM = YB OM > OB	YM = YB OM = OB	
Zhou & Krott, 2018[4]	M = B	M = B	M > B

[1]Facilitation effect: M = PEB; M > AEB, GEB; AEB < PEB, GEB; GEB = PEB. [2]ANCOVA (covariate: age): M < B. [3]Mixing cost: M = HPB. [4]μ parameter: M = B; τ parameter: M > B (inc.), M = B (cong.).

Note. RT: reaction times; ACC: accuracy; M: monolinguals; B: bilinguals; YM: young monolinguals; YB: young bilinguals; OM: old monolinguals; OB: old bilinguals; n.r.: not reported; GEB: German-English bilinguals; PEB: Polish-English bilinguals; AEB: Arabic-English bilinguals; acB: active bilinguals; inacB: inactive bilinguals; YOEH: young bilinguals only English at home; YWEH: young bilinguals Welsh and English at home; YOWH: young bilinguals only Welsh at home:; EM: English monolinguals; FM: French monolinguals; LB: late bilinguals; EB: early bilinguals; unbB: unbalanced bilinguals; bB: balanced bilinguals; HPB: high proficiency bilinguals; LPB: low proficiency bilinguals, HPM: high proficiency monolinguals; LPM: low proficiency monolinguals.

Young adults (18-34y) (n = 24)

Considering accuracy, nineteen studies did not evidence significant differences between monolinguals and bilinguals, while five studies did not analyze accuracy.

Considering RTs, fifteen studies did not find significant differences between monolinguals and bilinguals. In the third experiment, Bialystok et al. (2005) observed no significant differences between bilinguals and monolinguals. In the fourth and fifth experiments, bilinguals were faster than monolinguals in congruent and incongruent conditions. Bialystok, Craik et al. (2005) did not find significant differences between monolinguals and French-English bilinguals, while Cantonese-English bilinguals were faster than the other groups. In two studies (Damian et al., 2018; Lee Salvatierra & Rosselli, 2011), monolinguals were slower than bilinguals. Bialystok & DePape (2009) observed that bilinguals were faster in two conditions (opposite and conflictual), while there was no significant difference in the other two conditions (direction and position control). Gathercole et al. (2014) found that monolinguals were slower than two of the bilingual groups considered (Welsh and English and only Welsh at home bilinguals), while there were no significant differences for the only English at home bilingual group. Ouzia et al. (2019) did not observe significant differences between the groups. However, comparing groups of participants equivalent in age, monolinguals were faster than bilinguals. Two studies (Chabal et al., 2015; Rosselli et al., 2016) did not report RTs analyses.

Ten studies did not find significant differences, while ten did not conduct analyses regarding the Simon effect. In two studies (Woumans et al., 2015; Zhou & Krott, 2018), monolinguals showed a greater Simon effect than bilingual participants. Rosselli et al. (2016) observed a greater Simon effect for low proficiency bilinguals than high proficiency monolinguals, while no significant difference emerged between the other groups. Paap et al. (2015) showed a smaller Simon effect for monolinguals than bilinguals. Several studies calculated other indices to analyze the results further. Antón et al. (2019) computed congruity and incongruity effects, Coderre et al. (2014) calculated the facilitation effect, while Paap et al. (2015) calculated the mixing cost index, but no significant differences emerged in the studies. Blumenfeld and Marian (2014) computed the inverse efficiency score and found no significant differences between the groups.

Zhou and Krott (2018) evidenced no significant differences in the μ parameter and the τ parameter in the congruent condition, while monolinguals had a higher τ parameter in the incongruent condition.

Middle-aged adults (35-64y) (n = 4)

Two studies (Nair et al., 2017; Rodrigues & Zimmer, 2016) did not find any significant difference in accuracy, while Desjardins et al. (2020) did not analyze accuracy. Desjardins et al. (2020) did not find significant differences in RTs, while two studies (Bialystok et al., 2004; Nair et al., 2017) observed that monolinguals were slower than bilinguals. Rodrigues and Zimmer (2016) showed that bilinguals were faster than monolinguals only in the congruent condition, while no significant differences were found in the incongruent condition. In two studies (Desjardins et al., 2020; Rodrigues & Zimmer, 2016), there were no significant differences concerning the Simon effect, while two studies (Bialystok et al., 2004; Nair et al., 2017) observed a higher Simon effect for monolinguals than bilinguals.

Nair et al. (2017) showed a higher inverse efficiency score for monolinguals.

Older adults (>65y) (n = 15)

Most of the studies did not show any significant differences in accuracy (9/15), whereas four studies (Bialystok et al., 2005; Clare et al., 2016; de Bruin et al., 2015; Kousaie et al., 2014) did not conduct analyses on accuracy due to the small number of errors. Bialystok et al. (2004) observed that bilinguals made fewer errors than monolinguals in the incongruent situation, while no significant differences emerged in the congruent condition (Experiment 1). In the second experiment, bilinguals were more accurate than monolinguals, while there were no significant differences between the two groups in the third experiment. Gathercole et al. (2014) found that bilinguals were more accurate than monolinguals.

There were no significant differences in RTs (8/15) in most studies, while two studies did not analyze RTs (Clare et al., 2016; Cox et al., 2016).

Bialystok et al. (2005) did not find significant differences between bilinguals and monolinguals in the third experiment. In the fourth and fifth experiments, bilinguals were slower than monolinguals in congruent and incongruent conditions. Bialystok et al. (2004) observed that bilinguals were faster than monolinguals (Experiments 1 and 2). Kramer et al. (2015) did not find significant differences between early bilinguals and monolinguals. Conversely, late bilinguals were faster than monolinguals in the incongruent condition but not in the congruent condition. Lee Salvatierra and Rosselli (2011) observed that bilinguals were faster than monolinguals in the simple condition (classic Simon task) but not in the complex condition (i.e., four

different colored squares instead of two). Moreover, both the unbalanced bilingual and balanced bilingual groups were faster than monolinguals, whereas there were no significant differences between the two groups of bilingual participants. Zeng et al. (2019) observed that monolinguals were slower than bilinguals in incongruent but not congruent trials.
Seven studies (Bialystok 2009; Cox et al., 2016; Lee Salvatierra & Rosselli, 2011) did not find significant differences in the Simon effect (Bialystok et al., 2009; Cox et al., 2016; Lee Salvatierra & Rosselli, 2011). In three studies (Bialystok et al., 2009; Cox et al., 2016; Lee Salvatierra & Rosselli, 2011), monolinguals have a greater effect than bilinguals. Lee Salvatierra and Rosselli (2011) found that monolinguals have a greater Simon effect than bilinguals in simple but not complex conditions. Berroir et al. (2017) did not find significant differences in the inverse efficiency score between groups.

General results

Considering all the studies that used the Simon task, twenty studies did not show significant differences between bilinguals and monolinguals. In fourteen studies, at least one difference emerged in favor of the bilingual group. Specifically, in six studies, differences emerged relative to RTs or accuracy. Four studies evidenced differences only in the Simon effect, while in the other four studies in the conflict effect and RTs or accuracy. In three studies, at least one result emerged in favor of the monolingual group. In Lee Salvatierra and Rosselli (2011), young monolinguals were faster than bilinguals in the simple condition. Two studies showed a lower Simon effect for the monolinguals.

Go/No-Go task (n = 8)

Eight studies used the Go/No-Go task to assess motor inhibition ability.
Seven studies included young adult participants (Barbu et al., 2020; Costumero et al., 2015; Fernandez et al., 2013, 2014; Hofweber et al., 2020a; Moreno et al., 2014; Sanchez-Azanza et al., 2020), while one study older adult participants (Clare et al., 2016). There are methodological differences between the various studies; different versions of the task differed in the target stimuli (letters or geometric shapes of different colors) or the mode of presentation of the stimuli (visual or audio). Results are summarized in Table 6.

Table 6. Results of the studies that used the Go/No-Go task.

	Go/No-Go	
	RT	**ACC**
Barbu et al., 2020	M = HFLS = LFLS	N. of errors M = HFLS = LFLS
Clare et al., 2016		False alarms M = B
Costumero et al., 2015	M = B	N. of hits M = B
Fernandez et al., 2014	Visual and audio tasks M = B	
Fernandez et al., 2013	M = B	
Hofweber et al., 2020b		N. of errors M = B
Moreno et al., 2014	M = B	d-prime score M = B
Sanchez-Azanza et al., 2020		M = B

Note. RT: reaction times; ACC: accuracy; M: monolinguals; B: bilinguals; HFLS: high frequency switchers; LFLS: low frequency switchers.

Young adults (18-34y) (n = 7)

No studies observed significant differences between groups concerning RTs in Go trials and accuracy (assessed using different indices, see Table 6). Two studies (Fernandez et al., 2013, 2014) did not analyze accuracy due to small errors.

Older adults (>65y) (n = 1)

Clare et al. (2016) showed no significant differences in false alarms between the groups.

General results

Considering all studies that used the Go/No-Go task, no significant differences were found between bilinguals and monolinguals.

General results

Regardless of the experimental task, studies were scored for the presence (or absence) of significant differences between bilinguals and monolinguals. A score (ranging from -1 to 1) was assigned following the method used by Grundy (2020) in his quantitative analysis. Specifically, if one or more task outcomes from a study favored bilinguals over monolinguals, the study was classified as a 1 (e.g., faster RTs or greater accuracy or interference control index). If one or more outcomes favored monolinguals over bilinguals, the study was classified as a -1. If there were no group differences or some results favoring bilinguals and others favoring monolinguals, the study was 0. The data are reported in Table 7.

42.85% of studies showed a bilingual advantage in inhibition ability, while 47.95% found no differences between groups. Finally, 9.18% of studies showed better performance in the monolingual group than in the bilingual group.

Different results emerged analyzing the results of the studies according to the age range taken into consideration. Concerning young adults, 50.72% of the studies showed no significant differences, 37.68% evidenced a bilingual advantage, and 11.59% a monolingual advantage. For middle-aged adults, 28.57% of the studies showed no significant differences between the groups, while the remaining 71.42% showed a bilingual advantage. Finally, considering older adults, 45.45% of the studies showed no significant differences, 50% showed a bilingual advantage, and 4.54% showed a monolingual advantage.

Table 7. General results.

	Score (1: B > M; 0: no differences or mixed; -1: M > B)
Anderson et al., 2017	0
Ansaldo et al., 2015	1
Antón et al., 2019	0
Antón et al., 2016	0
Barbu et al., 2020	0
Bellegarda & Macizo, 2021	0
Berroir et al., 2017	0
Bialystok & DePape, 2009	1
Bialystok, Craik et al., 2005	1
Bialystok et al., 2004	1
Bialystok et al., 2009[1]	1
Bialystok et al., 2005	1
Bialystok et al., 2014	0
Blumenfeld & Marian, 2014	0
Bogulski et al., 2015	0
Botezatu et al., 2021	0
Chabal et al., 2015	0

Chrysochoou et al., 2020	0
Clare et al., 2016	1
Coderre et al., 2016	0
Coderre et al., 2014	1
Costa et al., 2008	1
Costa et al., 2009	1
Costumero et al., 2015	0
Cox et al., 2016	1
Damian et al., 2018	1
de Bruin et al., 2015	0
De Leeuw & Bogulski, 2016	0
Desideri & Bonifacci, 2018	1
Desjardins & Fernandez, 2018	0
Desjardins et al., 2020	0
Fernandez et al., 2014	0
Fernandez et al., 2013	0
Garraffa et al., 2017	0
Gathercole et al., 2014[2]	0
Grundy et al., 2017	0
Guðmundsdóttir & Lesk, 2019	0
Heidlmayr et al., 2014	1
Hernández et al., 2010	1
Heidlmayr et al., 2015	0
Hofweber et al., 2020a	1
Hofweber et al., 2020b	1
Hui et al., 2020	1
Kazemeini & Fadardi, 2016	0
Kirk et al., 2014	0
Kousaie & Phillips, 2012[3]	1
Kousaie & Phillips, 2017	0
Kousaie et al., 2014	0
Kuipers & Westphal, 2021	1
Kramer et al., 2015	1
Lee Salvatierra & Rosselli, 2011[1]	1
Luk et al., 2010	0
Massa et al., 2020	-1
Marzecová et al., 2013	1
Moreno et al., 2014	0
Morrison et al., 2019	0
Morrison et al., 2020	0
Nair et al., 2017	1
Okada e al., 2019	-1
Ooi et al., 2018	1
Ouzia et al., 2019	-1
Paap et al., 2015	-1
Patra et al., 2020	-1
Pelham & Abrams, 2014	1
Rayney et al., 2021	0
Rodrigues & Zimmer, 2016	1
Rosselli et al., 2016	-1
Sanchez-Azanza et al., 2020	0
Shulley & Shake, 2016	-1
Soares et al., 2019	0
Taler et al., 2013	0
Tao et al., 2011	1

Treffers-Daller et al., 2020	1
Vivas et al., 2017	-1
Waldie et al., 2009	0
Warmington et al., 2019	0
Woumans et al., 2015	1
Woumans et al., 2019	0
Xie & Dong, 2017	0
Xie & Zhou, 2020	0
Yang & Yang, 2016	1
Yudes et al., 2011	0
Zeng et al., 2019[1]	1
Zhou & Krott, 2018	1

[1]Young adults: 0; Older adults: 1.

[2]Young adults: -1; Older adults: 1.

[3]Young adults: 1; Older adults: 0.

Discussion

In recent years the interest in the effect of bilingualism on cognitive functioning has increased. Several studies compared monolinguals and bilinguals using the most famous and used tasks to assess cognitive and motor inhibition (Simon task, Stroop task, Flanker task, Go/No-go task). Most studies adopted these tasks interchangeably, assuming they measure the same cognitive function. Indeed, these classical cognitive tasks are not highly correlated and do not always reflect the same general ability. The results show high variability in the participants' performance, particularly evident in the studies on bilingualism. Bilingualism is a widespread phenomenon, but there is still no single definition. Although there was an initial tendency to adopt a dichotomous approach to defining bilingualism, there is now an attempt to consider this phenomenon along a continuum. There are many aspects to consider describing the bilingual experience of a person, for example, the age of acquisition, the communicative context within which exchanges take place, the percentage of daily use of each language. This systematic review summarizes the results of studies that assessed cognitive and motor inhibition in the adult bilingual population. Recently, other researchers tried to synthesize the results of studies investigating the effect of bilingualism on cognitive function and, surprisingly, did not come to the same conclusions.

Some previous meta-analyses (Donnelly et al., 2019; Lehtonen et al., 2018; Paap, 2019), investigating the bilingual advantage in executive functions, did not show evidence in favor of better performance in bilinguals. Conversely, the quantitative analysis conducted by Grundy (2020) showed evidence of a bilingual advantage. In the present review, more than half of the studies showed no difference in inhibition between monolingual and bilingual participants or a monolingual advantage. In contrast, about 40% of the studies showed a bilingual advantage. The results differ when the age of the participants is considered. Specifically, no significant differences emerged between the groups in 50.72% of the cases when studies included young adults. Differently, 50% of the studies considering older adults showed better performance of the bilingual participants. Several factors could have affected the results. Many studies adopted modified versions of the classical experimental tasks (e.g., Fernandez et al., 2013; Chrysochoou et al., 2020; Hofweber et al., 2020a). Moreover, the participants included within the studies were all classified as bilingual but different definitions were used. Some studies have included bilingual participants proficient in a third (e.g., Barbu et al., 2020; Heidlmayr et al., 2015; Marzecová et al. 2013) or fourth language (Heidlmayr et al., 2014; Marzecová et al. 2013). The classification of monolingual participants also presents critical aspects. Indeed, recruiting participants who do not have minimal exposure to a second language is very difficult (Bellegarda et al., 2021). For this reason, in several studies (e.g., Xie & Zhou, 2020), participants with limited knowledge of a second language have been classified as monolingual. However, these monolingual participants could be considered low-proficient bilinguals by changing the classification criterion (Xie & Zhou, 2020). Differently, Green and Abutalebi (2013) highlighted the role of context in determining the effects of bilingualism on cognitive functioning. Furthermore, apart from language history characteristics, additional sociodemographic characteristics can influence performance on experimental tasks. Cultural differences have been found to affect the development of executive functions (Paap, 2019). The present review included participants recruited from Europe (35 bilingual groups), America (27 bilingual groups in North America and 2 in South America), Asia (12 bilingual groups), and Australia (2 bilingual groups). Specifically, some participants were recruited from cities characterized by a multicultural context (e.g., Ooi et al., 2018) in which they were exposed to both languages from birth while others were late bilinguals (e.g., Kramer et al., 2015) or immigrants (e.g., Patra et al., 2020).

Gender is another factor that seems to affect performance in interference control ability. Previous studies showed that males have an advantage in interference control performance. Most of the studies included in this review did not balance participants by gender. Some of them presented a strong imbalance toward one of the two genders (e.g., Cox et al., 2016; Fernandez et al., 2013; Kazemeini & Fadardi, 2016). Finally, several studies have shown that the inclusion of young adult participants in studies leads to the absence of significant differences as the cognitive functions considered are at the peak of their development. The results of this review seem to partially confirm this hypothesis, although the percentage of studies in which significant differences between groups emerged is not negligible.

Moreover, Paap (2019) recently challenged this hypothesis, showing that even young adult participants show potential for improvement through executive function training. Thus, the lack of significant differences between monolinguals and bilinguals should not be attributed to age. In conclusion, the results of the present review do not allow us to draw definitive conclusions about the existence of a bilingual advantage in the resolution of tasks requiring cognitive and motor inhibition skills. Although about 50% of the studies do not show differences between monolinguals and bilinguals, better performance emerges in bilingual participants in about 40% of the studies (at least one of the outcomes considered). Several methodological and related aspects of the population of interest differ between the studies. Antón et al. (2019) stated that the results of a specific population with peculiar characteristics should not be generalized to the general population. Results should be interpreted with caution. Future studies should conduct quantitative analyses, considering aspects related to linguistic history, which were not analyzed in this review. Specifically, as suggested by Grundy (2020), future studies should seek to answer the question "when does a bilingual advantage emerge" rather than "does a bilingual advantage exist."

It is necessary to focus on the aspects that characterize the participants' linguistic and demographic history and understand which factors determine better performance in the bilingual population.

Chapter 2: Assessing executive inhibition in young, middle-aged, and older bilinguals

Introduction

According to the model of Miyake et al. (2000), executive functions refer to cognitive flexibility (e.g., the ability to switch between tasks), inhibition (e.g., the ability to suppress dominant responses), and monitoring (e.g., the ability to update information in working memory). In a subsequent review of this model, Miyake and Friedman (2012) hypothesized the existence of a general EF ability with distinct switching and updating components and the inhibitory control not separated but moderately linked to a general EF ability. The study of EFs is complex because of the task impurity problem (Miyake & Friedman, 2012). Any score derived from an EF task includes a variance attributable to non-EF processes (e.g., motor speed). Numerous studies on bilingualism focused on cognitive and executive functions development in the bilingual population. According to the Joint Activation Model of Green (1998), both languages would always be active in the brain of a bilingual person regardless of the language used at the given moment. For this reason, it would be necessary to use a general suppression mechanism to inhibit the activation of the non-target language. Moreover, the interactional context seems to play a role in the development of cognitive functions (see Green & Abutalebi, 2013) because bilinguals need to monitor what happens during a conversation (cognitive flexibility) and access the information contained in working memory. The debate on the effect of bilingualism on cognitive functions is still ongoing. Recently, some authors highlighted the need to change the approach used to study this topic, suggesting the need to overcome the dichotomous vision (there is an advantage vs. there is no advantage) and understand under what conditions such advantage is evident (e.g., Grundy, 2020). In the absence of a shared definition of bilingualism (for an overview of the use of the term bilingual, see Surrain & Luk, 2019), the populations considered within the studies could have very different language histories (e.g., age of acquisition, language pairs) making cross-study comparison difficult. In recent years, several systematic reviews and meta-analyses have been published attempting to draw conclusions about the effect of bilingualism on executive functions (e.g., Antoniou et al., 2019; Gasquoine, 2016; Giovannoli et al., 2020). Some of these studies focused on the effect of bilingualism on inhibition ability. However, the results do not always coincide. Donnelly et al. (2019) tested for the presence of a bilingual advantage in inhibition ability by considering all studies that had included at least one nonverbal interference-control task (e.g., Simon arrows, numerical Stroop). The study revealed a small but significant effect of bilingualism for interference cost. Moreover, late bilinguals showed larger advantages than early bilinguals. Paap's (2019) study also showed a small effect for bilingual advantage but concluded that there was no "compelling evidence that bilingualism enhances inhibitory control". The meta-analysis conducted by Lehtonen et al. (2018) reached a similar conclusion. Furthermore, the authors showed that the age of L2 acquisition, the immigrant status, the country in which the study was conducted, and the language pair did not seem to influence the results of the studies. Conversely, Grundy's (2020) study, which used a vote-counting procedure and considered all studies that assessed inhibition, switching ability, and working memory, found that differences between monolinguals and bilinguals are rare, but when they emerge, they favor bilinguals. A similar result emerged from the systematic review on the effect of bilingualism on inhibition ability reported in Chapter 1 of this thesis.

The aim of this study is to test the presence of a bilingual advantage in cognitive and motor inhibition in different age groups. It is hypothesized:

1) Any differences between the two language groups will favor the bilingual population group.

2) In the verbal Stroop task, no differences between the two linguistic groups will emerge due to the verbal nature of the task. In fact, it is well known that bilinguals have worse performances on verbal tasks than the

monolingual population (see Grundy, 2020), so the presence of verbal stimuli within the task should negatively affect the performance of the bilingual population.
3) According to the results of the systematic review in Chapter 1, the probability of detecting a difference is greater in the group of older bilinguals than in the other two age groups.

Method

Participants

One hundred and eighty-four people took part in the experiment. Participants were divided into two groups: the control group and bilinguals based on a preliminary assessment (see Language history). Each group was divided into three groups of different ages: 1) the control group: young adults (mean age = 25.62, SD = 3.30), middle-aged adults (mean age = 42.59, SD = 7.89), older adults (mean age 59.58, SD = 3.85); 2) Bilingual group: young adults (mean age = 26.04, SD = 3.45), middle-aged adults (mean age = 42.36, SD = 7.66), older adults (mean age = 59.00, SD = 4.04). Participants' characteristics and results will be separately reported for young, middle-age and older adults (see below).

Demographic information

Demographic information (e.g., age, gender, level of education) was collected through a structured questionnaire. A composite socioeconomic status (SES) score was calculated based on the a) educational level (0 – elementary school; 1 – secondary school certificate; 2 – high school diploma; 3 – bachelor degree; 4 – master degree; 5 – postgraduate specialization), b) type of occupation (0 – unemployed; 1 – blue-collar; 2 – white-collar), and c) position in occupation (0 – unemployed; 1 – unskilled worker; 2 – skilled/specialized professional; 3 – business owner; 4 – business owner with staff; 5 – executive member of the private or public sector).
Language Social Background Questionnaire (adapted from Anderson et al., 2018). The questionnaire contains three sections: 1) Social background, 2) Language background and 3) Community Language Use Behavior. The questions allow to assess language acquisition history, self-rated language proficiency (in writing, reading, speaking, and comprehension), and the current language usage pattern.
Language Dominance Questionnaire (Dunn & Tree, 2009). The questionnaire contains 12 questions and allow to assess the language dominance of participants.
Language Switching Questionnaire (Rodriguez-Fornells et al., 2012). The questionnaire contains 12 questions and assess language switching habits. The questionnaire provides four indices: 1) tendency to switch to L1; 2) tendency to switch to L2; 3) contextual switching (i.e., frequency of switching in particular situations); 4) unintended switches.

Language history

Participants were classified as a control group or bilingual based on the information collected through the Language and Social Background Questionnaire (Anderson et al., 2018). The questionnaire assessed language acquisition history, self-rated language proficiency (in writing, reading, speaking, and comprehension), and the current language usage pattern. Given the impossibility of recruiting participants in Italy who only knew one language (in Italy, it has been compulsory to teach at least one foreign language at school since 1962; MIUR, 1998), as in Kalamala et al. (2018), the neutral term 'control group' was chosen. Language Dominance Questionnaire (Dunn & Tree, 2009) and Language Switching Questionnaire (Rodriguez-Fornells et al., 2012) were used to measure language dominance and language switching habits.

Inhibition tasks

Flanker task. The stimuli consisted of a row of five black arrows presented on the center of the white screen. Participants had to press "A" on the keyboard when the central arrow (target) was pointing to the left and "L" if the target was pointing to the right. The experiment was introduced by a practice block of 12 trials with feedback on correctness. Afterward, a block of 48 randomly presented trials (24 congruent, and 24 incongruent) was presented. In the congruent condition, the five arrows pointed in the same direction. In the incongruent condition, the target pointed in the opposite direction of the flanker arrows. Reaction times and accuracy were recorded. Each trial started with the presentation of a fixation point (duration: 500 ms). The target stimulus remained on the screen until the participants responded. Reaction times of the correct responses and accuracy were recorded. The Flanker effect was calculated by subtracting RTs/accuracy in the congruent trials from the RTs/accuracy in the incongruent trials. Participants with an accuracy lower than 50% were removed from the analysis and response times that deviated in more than 2.5 SD from the mean of the participant were removed. Figure 1 reports an example of the procedure.

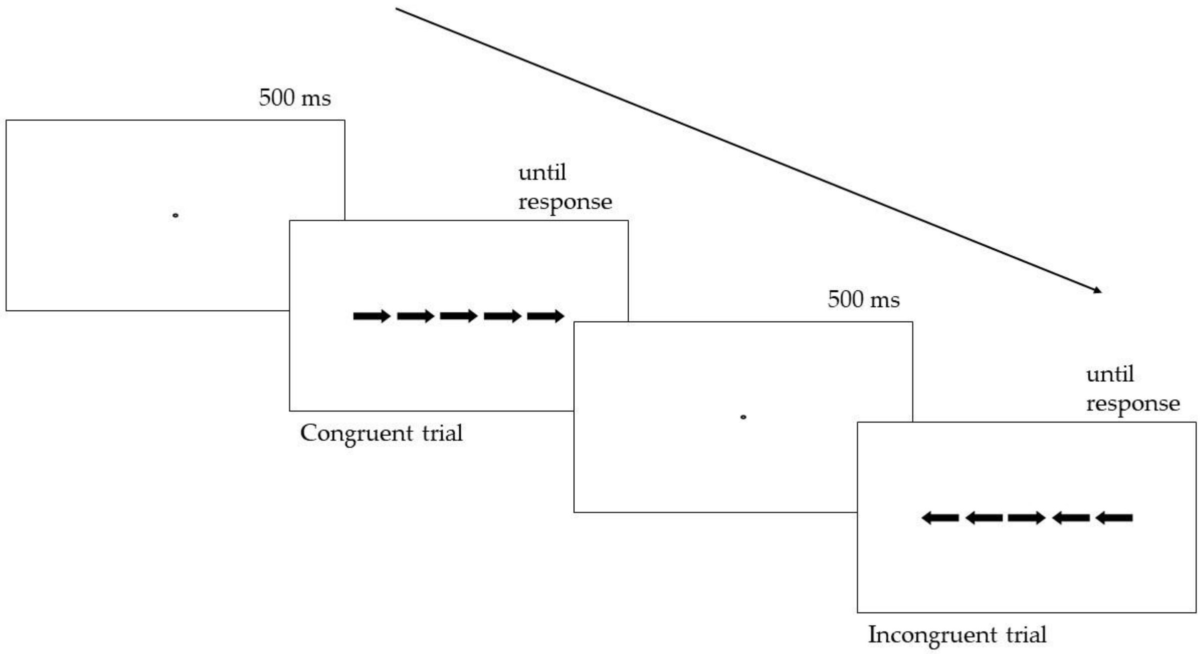


Figure 1. Example of the Flanker task procedure.

Simon task. The stimuli consisted of a black square or circle presented on the right or the left side of the white screen. Participants were instructed to press "L" on the keyboard if they saw a square and "A" if they saw a circle. The experiment was introduced by a practice block of 12 trials with feedback on correctness. Afterward, a block of 48 randomly presented trials (24 congruent, and 24 incongruent) was presented.
In the congruent condition, the stimulus was presented on the same side of the screen of the response button needed. In the incongruent condition, the stimulus was presented on the opposite side to the response button. Each trial started with the presentation of a fixation point (duration: 500 ms). The target stimulus remained on the screen until the participants responded. Reaction times of the correct responses and accuracy were recorded. The Simon effect was calculated by subtracting RTs/accuracy in the congruent trials from the RTs/accuracy in the incongruent trials. Participants with an accuracy lower than 50% were removed from the analysis and response times that deviated in more than 2.5 SD from the mean of the participant were removed. Figure 2 reports an example of the procedure.

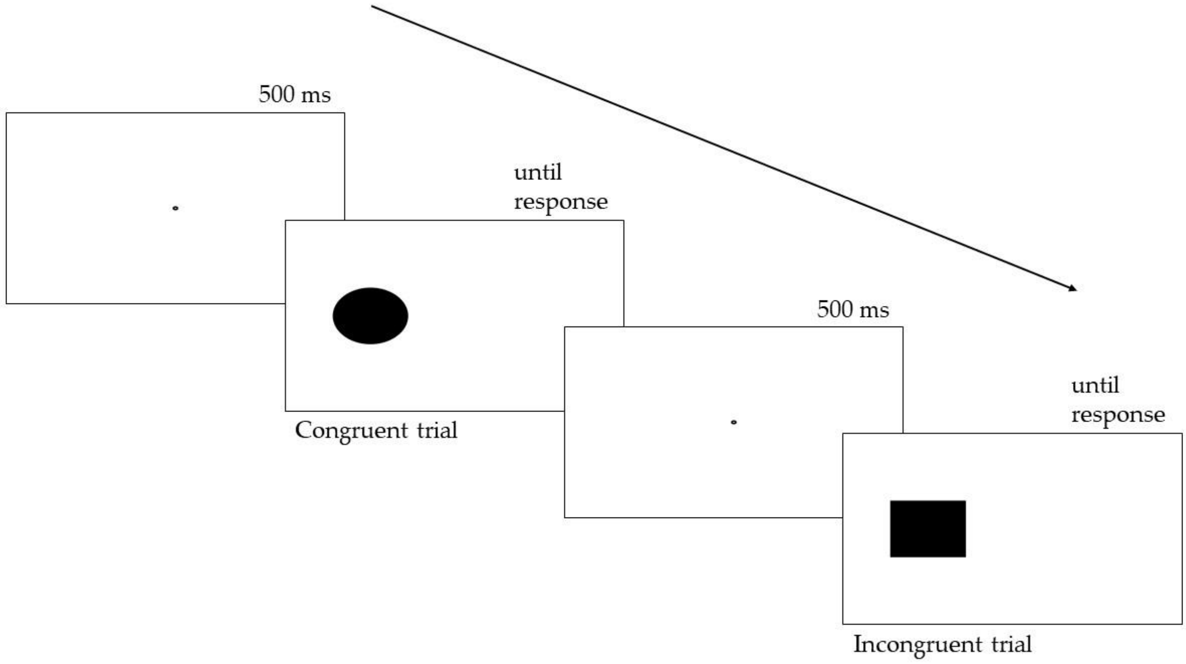


Figure 2. Example of the Simon task procedure.

Verbal Stroop task. The stimuli consisted of four words of colors (red, yellow, blue, green) presented in different colors. Participants must indicate the color of the ink by pressing a button on the keyboard (key "A" = red; key "L" = yellow; key "K" = blue; key "S" = green). To help participants remember the response keys, under the target was presented a bar with the order of colors. The experiment was introduced by a practice block of 12 trials with feedback on correctness. Afterward, a block of 48 randomly presented trials (24 congruent, and 24 incongruent) was presented. In the congruent condition, the color of the ink was the same as the word meaning. In the incongruent condition, the color of the ink was different from the word meaning. Each trial started with the presentation of a fixation point (duration: 500 ms). The target stimulus remained on the screen until the participants responded. Reaction times of the correct responses and accuracy were recorded. The Stroop effect was calculated by subtracting RTs/accuracy in the congruent trials from the RTs/accuracy in the incongruent trials. Participants with an accuracy lower than 50% were removed from the analysis and response times that deviated in more than 2.5 SD from the mean of the participant were removed. Figure 3 reports an example of the procedure.

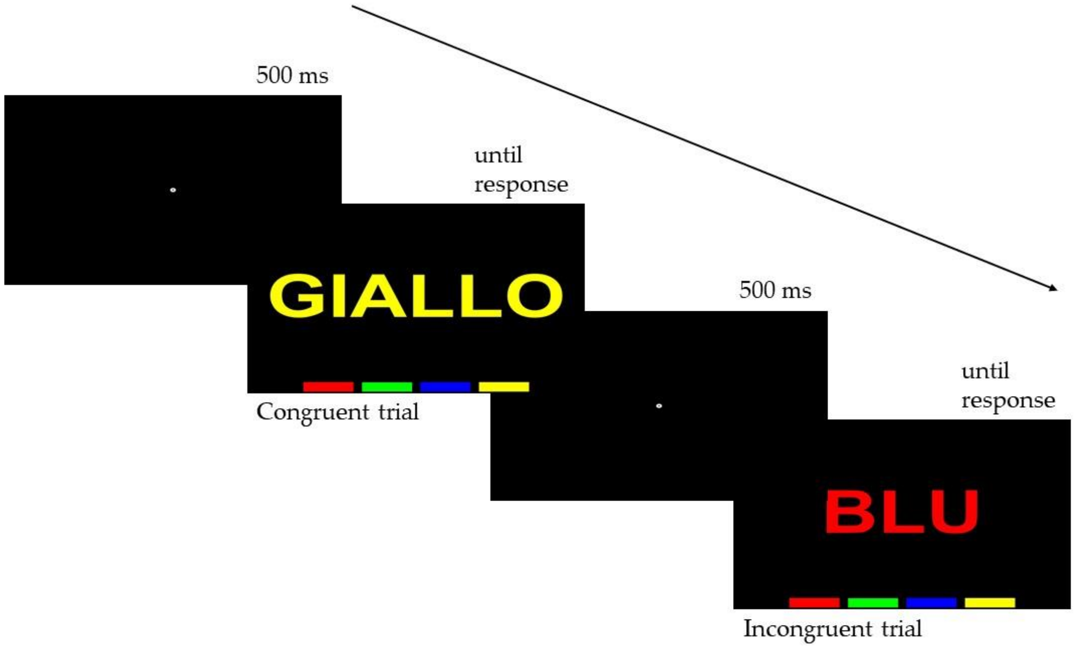


Figure 3. Example of the verbal Stroop task procedure.

Nonverbal Stroop task. The experimental task was adapted from Archibald and Kerns (1999). The stimuli consisted of colored fruits and vegetables (strawberries, bananas, peas, eggplants) presented in different colors. Participants must indicate the correct color of the fruit and vegetables by pressing a button on the keyboard (key “A” = red; key “L” = yellow; key “K” = blue; key “S” = green). To help participants remember the response keys, under the target was presented a bar with the order of colors. The experiment was introduced by a practice block of 12 trials with feedback on correctness. Afterward, a block of 48 randomly presented trials (24 congruent, and 24 incongruent) was presented. In the congruent condition, the fruits and vegetables were appropriately colored (e.g., yellow bananas, red strawberries). In the incongruent condition, the fruits and vegetables were inappropriately colored (e.g., red bananas, green strawberries). Each trial started with the presentation of a fixation point (duration: 500 ms). The target stimulus remained on the screen until the participants responded. Reaction times of the correct responses and accuracy were recorded. The Stroop effect was calculated by subtracting RTs/accuracy in the congruent trials from the RTs/accuracy in the incongruent trials. Participants with an accuracy lower than 50% were removed from the analysis and response times that deviated in more than 2.5 SD from the mean of the participant were removed. Stimuli are shown in Figure 4. Figure 5 reports an example of the procedure.

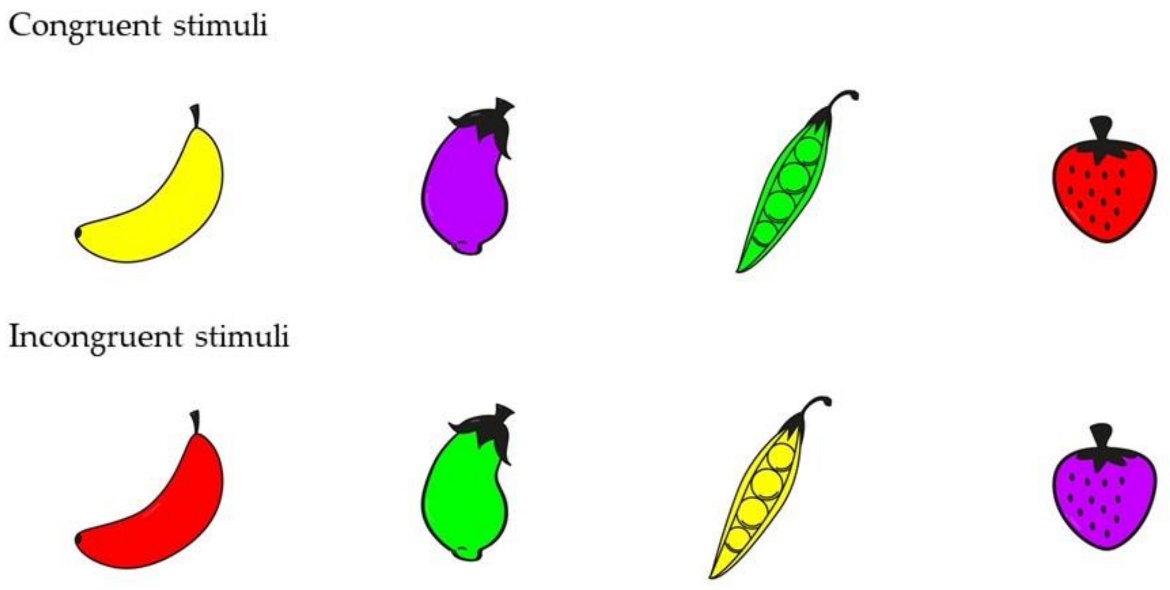


Figure 4. The target stimuli used in the nonverbal Stroop task.

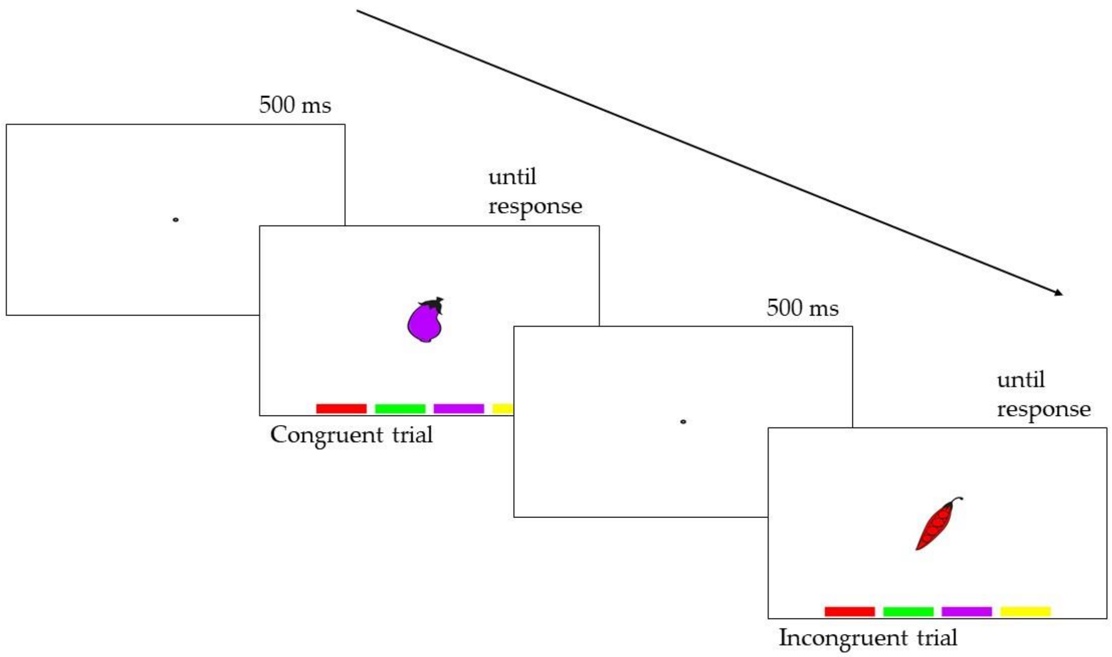


Figure 5. Example of the non-verbal Stroop task procedure.

Global-Local task. The experiment consisted of two blocks: the global and the local blocks. The order of the blocks was counterbalanced among the participants. 12 practice trials introduced each block with feedback on correctness. In the global block were randomly presented 48 trials (16 congruent, 16 incongruent, and 16 neutrals). Participants were instructed to indicate the global dimension. The global and local shapes match in the congruent condition while they differ in the incongruent condition. In the neutral trials, the local dimension included Xs rather than shapes. In the local block were randomly presented 48 trials (16 congruent, 16 incongruent, and 16 neutrals). The participants must indicate the local dimension. Congruent and incongruent trials did not differ from the global block. In neutral trials, the global dimension included Xs rather than shapes. Each trial started with the presentation of a fixation point (duration: 500 ms). The target stimulus remained on the screen until the participants responded. Reaction times of the correct responses and accuracy were recorded. Congruency effect was computed by subtracting RTs/accuracy in congruent trials from RT/accuracy in incongruent trials; costs were calculated by subtracting RTs/accuracy in neutral trials from RTs/accuracy in incongruent trials, and benefits were computed subtracting RTs/accuracy in congruent trials from RTs/accuracy in neutral trials. Participants with an accuracy lower than 50% were removed from the analysis, and response times deviated in more than 2.5 SD from the mean of the participant were removed. Figure 6 reports an example of the procedure.

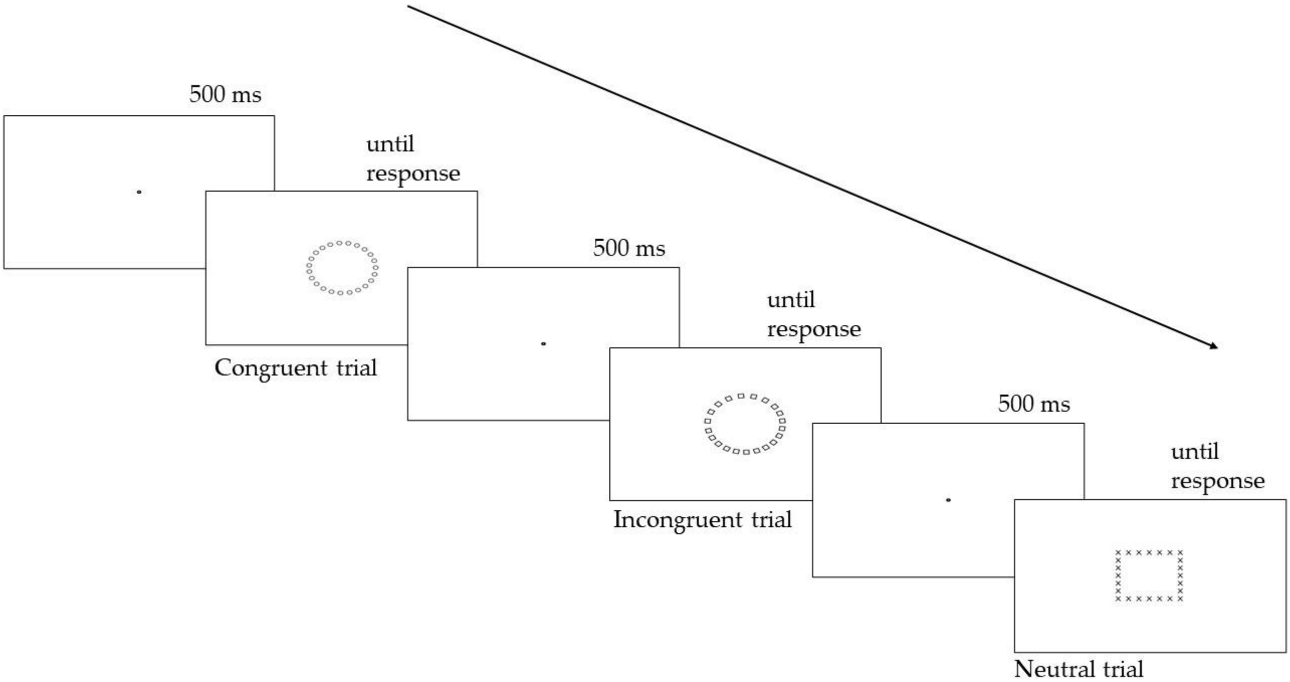


Figure 6. Example of the Global-Local task procedure.

Go/No-Go. The stimuli consisted of a green triangle, or a circle presented in the center of the screen. Participants were instructed to press the left button of the mouse if they saw a green circle (Go trials) and to withhold their response when a green triangle was presented (No-Go trials). The experiment was introduced by a practice block of 12 trials with feedback on correctness. Afterwards, a block of 72 randomly presented trials was presented. The initial screen with a fixation point (duration: 500 ms) was followed by the presentation of target stimuli (Go) or non-target stimuli (No-Go), in a randomized way considering three, four, or five Go trials for each No-Go trial. Each stimulus lasts 2000 ms or until the participant responds. Reaction times and accuracy were recorded. The incorrect responses in the No-Go trials were coded as false alarms. Participants with an accuracy lower than 50% were removed from the analysis. Figure 7 reports an example of the procedure.

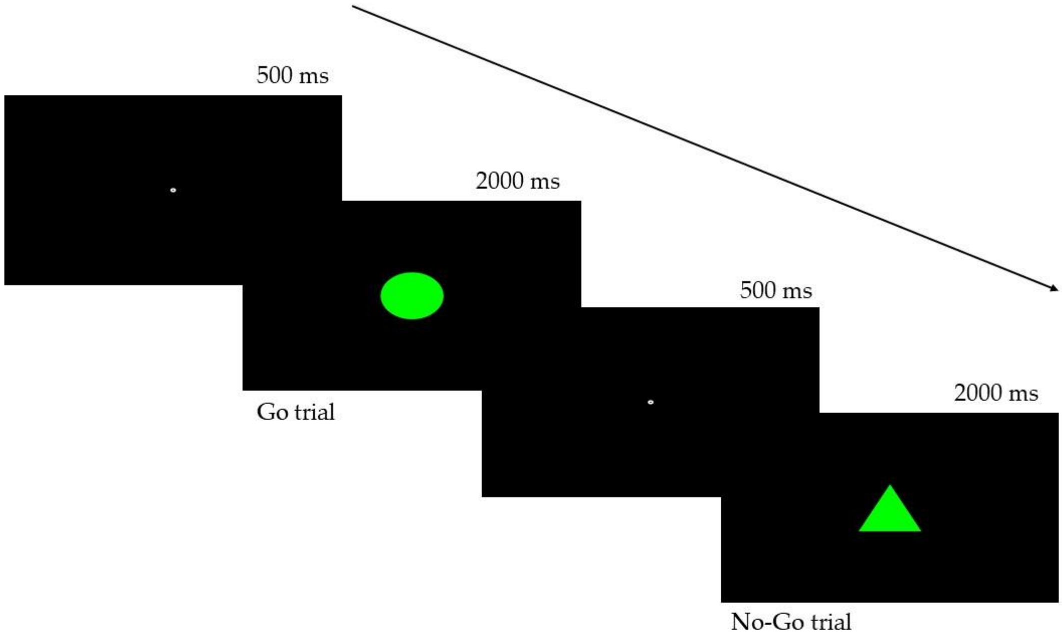


Figure 7. Example of the Go/No-Go task procedure.

General procedure

All participants completed the questionnaires and cognitive tests online. The questionnaires were administered online using the KoboToolBox platform (www.kobotoolbox.org), while the cognitive tests were programmed using OpenSesame software (Mathôt et al., 2012) and administered online using JATOS (Lange et al., 2015). All participants completed the Language and Social Background Questionnaire (adapted from Anderson et al., 2018), the Flanker task (Eriksen & Eriksen, 1974), the Simon task (Simon, 1990), the classic Stroop task (Stroop, 1935), a nonverbal version of the Stroop task (adapted from Archibald & Kerns, 1999), the Go/No-Go task (Simson et al., 1977) and the Global-Local task (Navon, 1977). The bilingual participants completed the Language Switching Questionnaire (adapted from Rodriguez-Fornells et al., 2012) and the Language Dominance Questionnaire (adapted from Dunn & Tree, 2009). The questionnaires were presented in a fixed order, while the cognitive tests followed a random order. All testing was conducted in Italian.

Data analysis

For four EF tasks (Flanker task, Simon task, verbal Stroop task, nonverbal Stroop task), after removing errors, response times that deviated in more than 2.5 SD from the mean of the participants were removed. Response accuracy and response times were analyzed using a 2 (Language Group: bilinguals, control group) x 2 (Trial type: congruent, incongruent) mixed-design ANOVAs. The conflict index (i.e., incongruent minus congruent trials) was compared across language groups using ANOVAs.

For the Global-Local task, after removing errors, response times that deviated in more than 2.5 SD from the mean of the participants were removed. Response accuracy and response times were analyzed using a 2 (Language Group: bilinguals, control group) x 2 (Condition: Global block, Local block) x 3 (Trial type: congruent, incongruent, neutral) mixed-design ANOVAs. The conflict index, the costs (i.e., incongruent – neutral trials), and benefits (congruent – neutral trials) and the total interference score [(Localcongruent – Localincongruent) + (Globalcongruent – Globalincongruent)] were compared across language groups using ANOVAs.

For the Go/No-Go task, after removing errors, response times that deviated in more than 2.5 SD from the mean of the participants were removed. False alarms (incorrect response in No-go trials), accuracy in Go trials, and RTs in Go trials were analyzed using ANOVAs.
To verify the association between inhibition tasks, linear Pearson's r correlations were used to analyze the indices of the following tasks: Flanker task (Flanker effect on RTs), Simon task (Simon effect on RTs), verbal Stroop task (Stroop effect on RTs), nonverbal Stroop task (Stroop effect on RTs), Global-Local task (total interference score on RTs), Go/No-Go task (number of False Alarms).
Planned comparisons were used to analyze the main effect of the task and the interactions.
For all the statistical analyses, the level of significance was accepted at $p< .05$. Statistical analysis was conducted using STATISTICA v10.0.

Results

Young adults

Participants

One hundred and twelve participants between 18-34 years took part in the study.
Fifty-six participants were bilingual speakers of Italian and one additional language (mean age = 26.04, SD = 3.45; 82.14% females). Bilinguals' additional language were English (35), Spanish (11), French (6), Portuguese (1), German (1), Arabic (1), Romanian (1). Most of the participants were born in Italy (N = 50) while some of them in a foreign country (Brazil, N = 2; Moldova, N = 1; Philippines, N = 1; Morocco, N = 1; Peru, N = 1). Some bilingual participants reported some experience in L3 (English, N = 18; Spanish, N = 7; Portuguese, N = 4; French, N = 4; Dutch = 4; Tagalog, N = 1; Creole, N = 1; Romanian, N = 1; Russian, N = 1) and L4 (Spanish, N = 10; French, N = 7; German, N = 2; Portuguese, N = 1; Russian, N = 1). Four participants reported some experience in Italian regional dialects.
The control group included fifty-six participants (mean age = 25.62, SD = 3.30; 69,64% females). The L1 of all participants was Italian. Some participants of the control group reported some experience in L2 (English, N = 43; Spanish, N = 2; German, N = 1), L3 (Spanish, N = 5; French, N = 4; English, N = 3) and L4 (French, N = 2; Portuguese, N = 1). Their experience in L2, L3 and L4 was constrained to foreign language courses, and they reported marginal daily use of any foreign language. Twenty-four participants of the control group reported some experience in Italian dialects. One participant was born in Belgium while all the others were born in Italy. Participants were recruited using social networks and word of mouth. See Table 1 for background information about participants and their language knowledge.

Table 1. Background information about participants and their language knowledge.

	Bilinguals	**Control group**	*p*
Young adults	N = 56	N = 56	
Age	26.04 (3.45)	25.62 (3.30)	n.s.
% females	82.14	69.64	
SES	5.18 (2.39)	4.43 (2.17)	n.s.
Language experience L1			
Age of exposure (in years)	0.59 (2.10)	0.16 (0.71)	n.s.
Percentage of daily use (%)	59.30 (18.51)	90.85 (4.95)	**
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.71 (0.80)	9.59 (0.80)	n.s.
Proficiency – Comprehension (1-10)	9.95 (0.23)	9.73 (0.70)	*
Proficiency – Writing (1-10)	9.80 (0.61)	9.53 (0.87)	n.s.
Proficiency – Reading (1-10)	9.96 (0.18)	9.73 (0.70)	*
Language experience L2			
Age of exposure (in years)	10.20 (9.24)	9.24 (5.12)	n.s.

Self-rated language proficiency			
Proficiency – Speaking (1-10)	8.36 (1.58)	5.90 (1.74)	**
Proficiency – Comprehension (1-10)	8.80 (1.08)	6.90 (1.54)	**
Proficiency – Writing (1-10)	7.86 (1.91)	5.90 (1.70)	**
Proficiency – Reading (1-10)	8.89 (1.33)	6.93 (1.36)	**
Middle-aged adults	N = 17	N = 17	
Age	42.35 (7.66)	42.59 (7.89)	n.s.
% females	82.35	64.71	
SES	6.76 (2.36)	7.00 (2.15)	n.s.
Language experience L1			
Age of exposure (in years)	0.00 (0.00)	0.41 (1.70)	n.s.
Percentage of daily use (%)	50.26 (25.08)	91.22 (7.50)	**
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.47 (1.07)	900 (2.34)	n.s.
Proficiency – Comprehension (1-10)	90.71 (0.98)	9.65 (0.86)	n.s.
Proficiency – Writing (1-10)	9.53 (1.07)	8.94 (2.41)	n.s.
Proficiency – Reading (1-10)	9.65 (1.00)	9.41 (1.73)	n.s.
Language experience L2			
Age of exposure (in years)	16.35 (12.47)	11.78 (10.92)	n.s.
Self-rated language proficiency			
Proficiency – Speaking (1-10)	8.35 (1.76)	5.57 (1.27)	**
Proficiency – Comprehension (1-10)	8.76 (1.64)	5.43 (1.27)	**
Proficiency – Writing (1-10)	7.65 (2.29)	5.43 (1.90)	*
Proficiency – Reading (1-10)	8.18 (2.40)	5.86 (1.77)	*
Older adults	N = 19	N = 19	
Age	59.00 (4.04)	59.58 (3.85)	n.s.
% females	68.42	78.95	
SES	6.47 (2.39)	5.79 (2.80)	n.s.
Language experience L1			
Age of exposure (in years)	1.05 (3.42)	1.37 (3.48)	n.s.
Percentage of daily use (%)	49.74 (16.12)	95.14 (5.76)	**
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.74 (0.56)	9.16 (0.23)	*
Proficiency – Comprehension (1-10)	9.89 (0.31)	9.31 (1.00)	*
Proficiency – Writing (1-10)	9.53 (0.90)	9.05 (1.18)	n.s.
Proficiency – Reading (1-10)	9.84 (0.37)	9.47 (0.90)	n.s.
Language experience L2			
Age of exposure (in years)	25.47 (21.78)	13.78 (9.67)	n.s.
Self-rated language proficiency			
Proficiency – Speaking (1-10)	8.21 (1.51)	5.12 (1.88)	**
Proficiency – Comprehension (1-10)	8.42 (1.57)	5.62 (1.78)	**
Proficiency – Writing (1-10)	7.68 (1.67)	4.12 (1.64)	**
Proficiency – Reading (1-10)	8.34 (1.53)	5.25 (2.05)	**

* $p < .05$; ** $p < .01$

Measures of bilingualism

There were significant differences in subjective language proficiency ratings, and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from Italian to L2. Bilinguals' language profile is summarized in Table 2.

Table 2. Bilinguals' language profile. Means (standard deviation).

	L1	**L2**	Statistical results
Young adults (N = 56)			
Age of exposure (in years)	0.59 (2.10)	10.20 (9.24)	
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.71 (0.80)	8.36 (1.58)	$t(55)= 5.90; p < .01$
Proficiency – Comprehension (1-10)	9.95 (0.23)	8.80 (1.08)	$t(55)= 8.00; p < .01$
Proficiency – Writing (1-10)	9.80 (0.61)	7.86 (1.91)	$t(55)= 6.01; p < .01$
Proficiency – Reading (1-10)	9.96 (0.18)	8.89 (1.33)	$t(55)= 7.58; p < .01$
Language dominance	19.18 (5.41)	8.11 (5.11)	$t(55)= 9.03; p < .01$
Language switching habit	8.61 (2.14)	8.84 (2.34)	$t(55)= -0.87$; p = n.s.
Middle-aged adults (N = 17)			
Age of exposure (in years)	0.00 (0.00)	16.35 (12.47)	
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.47 (1.07)	8.35 (1.76)	$t(16)= 2.72; p < .01$
Proficiency – Comprehension (1-10)	9.71 (0.98)	8.76 (1.64)	$t(16)= 3.24; p < .01$
Proficiency – Writing (1-10)	9.53 (1.07)	7.65 (2.29)	$t(16)= 3.43; p < .01$
Proficiency – Reading (1-10)	9.65 (1.00)	8.18 (2.40)	$t(16)= 2.71; p < .01$
Language dominance	17.12 (6.30)	11.00 (5.96)	$t(16)= 2.29; p = .02$
Language switching habit	7.88 (2.06)	8.06 (1.78)	$t(16)= -0.44$; p = n.s.
Older adults (N = 19)			
Age of exposure (in years)	1.05 (3.42)	25.47 (21.78)	
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.74 (0.56)	8.21 (1.51)	$t(18)= 4.79; p < .01$
Proficiency – Comprehension (1-10)	9.89 (0.31)	8.42 (1.57)	$t(18)= 4.38; p < .01$
Proficiency – Writing (1-10)	9.53 (0.90)	7.68 (1.67)	$t(18)= 4.99; p < .01$
Proficiency – Reading (1-10)	9.84 (0.37)	8.34 (1.53)	$t(18)= 2.72; p < .01$
Language dominance	14.74 (4.83)	5.53 (5.26)	$t(18)= 4.48; p < .01$
Language switching habit	7.31 (1.80)	7.47 (2.36)	$t(18)= -0.44$; p = n.s.

Experimental data results

Table 3 shows the means and standard deviation of dependent variables for each experimental task.

Flanker task

Considering RTs, the main effect of Trial type was significant ($F_{1,110} = 87.51; p < .01; \eta^2 = .44$), participants were faster in the congruent trials (428.04 ms) than the incongruent trials (446.15 ms). There main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($F < 1$) were not significant. The one-way ANOVA on the Flanker effect showed no significant differences between ($F < 1$).
Considering accuracy, the effect of Trial type was significant ($F_{1,110} = 9.1; p < .01; \eta^2 = .08$), participants were more accurate in the congruent trials (96.43%) than in the incongruent trials (94.83%). The main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($F < 1$) were not significant.
The one-way ANOVA on the Flanker effect showed no significant differences between ($F < 1$).

Simon task

Considering RTs, the main effect of Trial type was significant ($F_{1,110} = 21.56$; $p < .01$; $\eta^2 = .16$), participants were faster in the congruent trials (512.91 ms) than the incongruent trials (530.85 ms). The main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($p = .26$) were not significant. The one-way ANOVA on the Simon effect showed no significant differences between ($p = .26$).

Considering accuracy, the main effect of Trial type was significant ($F_{1,110} = 4.18$; $p < .05$; $\eta^2 = .04$), participants were more accurate in the congruent trials (94.49%) than in the incongruent trials (92.93%). The main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($F < 1$) were not significant. The one-way ANOVA on the Simon effect showed no significant differences between ($F < 1$).

Verbal Stroop task

Considering RTs, the main effect of Trial type was significant ($F_{1,109} = 72.39$; $p < .01$; $\eta^2 = .40$), participants were faster in congruent trials (756.82 ms) than in the incongruent trials (836.41 ms). The main effect of Language group was also significant ($F_{1,109} = 4.06$; $p < .05$; $\eta^2 = .03$) and bilinguals were faster than the control group (758.42 vs 834.81 ms). The Language Group x Trial Type interaction ($F < 1$) was not significant. The one-way ANOVA on the verbal Stroop effect showed no significant differences between ($F < 1$).

Considering accuracy, the main effect of Trial type was significant ($F_{1,109} = 38.10$; $p < .01$; $\eta^2 = .26$), participants were more accurate in the congruent trials (96.14%) than in the incongruent trials (91.86%). The main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($F < 1$) were not significant. The one-way ANOVA on the verbal Stroop effect showed no significant differences between ($F < 1$).

Non-verbal Stroop task

Considering RTs, the main of Trial type was significant ($F_{1,109} = 116.55$; $p < .01$; $\eta^2 = .52$). Participants were faster in the congruent trials (725.84 ms) than the incongruent trials (809.40 ms). The main effect of Language Group was marginally significant ($F_{1,109} = 3.03$; $p = .08$; $\eta^2 = .03$) and bilinguals were faster than the control group (741.53 vs 793.72 ms). The Language Group x Trial Type interaction was not significant ($F < 1$). The one-way ANOVA on the non-verbal Stroop effect showed no significant differences between groups ($F < 1$).

Considering accuracy, the main effect of Trial type was significant ($F_{1,109} = 65.98$; $p < .01$; $\eta^2 = .38$). Participants were more accurate in the congruent trials (97.07%) than in the incongruent trials (91.71%). The main effect of the Language Group ($F < 1$) and the Language Group x Trial Type interaction ($F < 1$) were not significant. The one-way ANOVA on the non-verbal Stroop effect showed no significant differences between ($F < 1$).

Global-Local task

Considering RTs, the main effect of Condition was significant ($F_{1,96} = 13.87$; $p < .01$; $\eta^2 = .13$), participants responded faster in Global block (583.66 ms) than in Local block (559.39 ms). The main effect of Trial type was significant ($F_{2,192} = 14.68$; $p < .01$; $\eta^2 = .13$), participants were faster in congruent trials than in incongruent ($F_{1,96} = 23.68$; $p < .01$; $\eta^2 = .20$) and neutral trials ($F_{1,96} = 13.87$; $p < .01$; $\eta^2 = .11$). Participants were also faster in neutral trials than incongruent trials ($F_{1,96} = 4.26$; $p = .04$; $\eta^2 = .04$). There was no main effect of Language Group ($p = .30$). All the interactions were not significant ($p > .05$). In Global block, the one-way ANOVAs on congruency effect ($F < 1$), costs ($F < 1$) and benefits ($F < 1$) were not significant different between groups. In Local trials, the one-way ANOVAs on congruency effect ($p = .17$) and benefits ($F < 1$) were not significant different while costs showed a tendency to significance ($F_{1,96} = 3.05$; $p = .08$; $\eta^2 = .03$). Bilinguals showed higher costs than control group (11.91 vs -2.70 ms). Total interference scores revealed no significant differences between groups ($F < 1$).

Considering accuracy, the main effect of Condition was not significant ($p = .27$). There was a main effect of Trial type ($F_{2,192} = 6.23$; $p < .01$; $\eta^2 = .06$). Participants were more accurate in congruent trials than incongruent ($F_{1,96} = 11.50$; $p < .01$; $\eta^2 = .11$) and neutral trials ($F_{1,96} = 7.96$; $p < .01$; $\eta^2 = .08$). No differences emerged between neutral and incongruent trials ($F < 1$). There was no main effect of Language Group ($p = .32$). All the interactions were not significant ($p > .05$). In Global and Local blocks, the one-way ANOVAs on congruency effect ($F < 1$), costs ($F < 1$) and benefits ($F < 1$) were not statistically significant. Total interference scores revealed no significant differences between groups ($F < 1$).

Go/No-Go task

Considering RTs, the main effect of Language Group was not significant ($F < 18$). The accuracy analysis of Go trials showed a main effect of Language Group ($F_{1,110} = 5.8$; $p = .02$; $\eta^2 = .05$), and bilinguals were less accurate than the control group (96.53% vs 97.49%). Considering the number of false alarms, no differences emerged between groups ($F < 1$).

Table 3. Means and standard deviations (in parentheses) of inhibition measures for young adult participants.

	Young Bilinguals	Young Control group
Flanker task		
Reaction times		
Congruent trials	423.54 (59.55)	432.55 (62.43)
Incongruent trials	441.77 (62.18)	450.52 (59.25)
Flanker effect	18.23 (18.57)	17.97 (22.23)
Accuracy		
Congruent trials	96.65 (3.84)	96.20 (3.99)
Incongruent trials	94.79 (4.07)	94.87 (4.42)
Flanker effect	-1.86 (5.38)	-1.34 (5.84)
Simon task		
Reaction times		
Congruent trials	512.36 (79.16)	513.46 (88.01)
Incongruent trials	525.90 (64.40)	535.81 (84.86)
Simon effect	13.53 (46.63)	22.36 (34.22)
Accuracy		
Congruent trials	94.42 (5.35)	94.57 (5.20)
Incongruent trials	92.93 (6.97)	92.93 (7.10)
Simon effect	-1.49 (7.26)	-1.64 (8.84)
Verbal Stroop task		
Reaction times		
Congruent trials	719.15 (168.41)	794.49 (199.61)
Incongruent trials	797.70 (202.84)	875.13 (243.72)
verbal Stroop effect	78.55 (82.62)	80.63 (111.02)
Accuracy		
Congruent trials	96.51 (4.15)	95.76 (5.30)
Incongruent trials	92.42 (6.00)	91.29 (9.31)
verbal Stroop effect	-4.09 (6.73)	-4.46 (7.82)
Non-verbal Stroop task		
Reaction times		
Congruent trials	702.02 (130.88)	749.66 (158.05)
Incongruent trials	781.03 (152.24)	837.77 (203.77)
non-verbal Stroop effect	79.00 (68.41)	86.54 (92.92)
Accuracy		
Congruent trials	96.87 (3.91)	97.27 (4.26)
Incongruent trials	91.14 (6.21)	92.27 (6.87)
non-verbal Stroop effect	-5.73 (7.22)	-4.91 (6.65)
Global-Local task		
Global block		
Reaction times		
Congruent trials	519.58 (83.62)	542.19 (95.04)
Incongruent trials	535.10 (79.98)	557.78 (106.93)
Neutral trials	530.78 (95.21)	546.52 (94.77)

Global effect	15.52 (37.31)	15.59 (55.16)
Costs	4.33 (53.82)	11.26 (56.45)
Benefits	11.20 (46.48)	4.33 (47.33)
Accuracy		
Congruent trials	95.15 (4.98)	95.54 (5.70)
Incongruent trials	94.51 (6.46)	94.51 (6.20)
Neutral trials	94.90 (5.66)	94.26 (5.96)
Global effect	-0.64 (8.95)	-1.02 (7.48)
Costs	-0.38 (8.40)	0.25 (9.28)
Benefits	-0.25 (7.54)	-1.27 (7.44)
Local block		
Reaction times		
Congruent trials	536.44 (79.66)	557.62 (116.71)
Incongruent trials	564.49 (75.87)	571.26 (118.87)
Neutral trials	552.58 (73.01)	573.96 (116.74)
Local effect	28.04 (40.63)	13.64 (61.58)
Costs	11.91 (44.79)	-2.70 (37.77)
Benefits	16.13 (48.80)	16.34 (48.38)
Accuracy		
Congruent trials	96.81 (4.25)	95.02 (10.04)
Incongruent trials	92.98 (7.17)	92.47 (8.65)
Neutral trials	94.64 (5.41)	92.60 (9.77)
Local effect	-3.83 (8.63)	-2.55 (7.96)
Costs	-1.66 (9.90)	-0.13 (9.59)
Benefits	-2.17 (6.45)	-2.42 (8.54)
Total interference score - ACC	-4.46 (12.03)	-3.57 (11.41)
Total interference score - RTs	43.57 (48.94)	29.23 (92.58)
Go/No-Go task		
RTs (Go trials)	375.22 (58.68)	381.17 (61.33)
Accuracy (Go trials)	96.53 (2.62)	97.49 (1.43)
False alarms	1.36 (1.42)	1.41 (1.69)

Middle-aged adults

Participants

34 participants between 35-54 years took part in the study.
17 participants were bilingual speakers of Italian (L1) and at least one additional language (mean age = 42.35, SD 7.66; 82.35% females). Bilinguals' L2 were Spanish (N = 6), English (N = 5), French (N = 3), German (1), Sardinian (1), Arbereshe (1). One participant was born in Ireland, while all the others were born in Italy. Some bilingual participants reported some experience in L3 (English, N = 7; Portuguese, N = 1; French, N = 1; Dutch = 1) and L4 (French, N = 3; Spanish, N = 1; German, N = 1; English, N = 1). Six participants reported some experience in Italian regional dialects.
The control group included 17 participants (mean age = 42.59, SD 7.89; 64.71% females). The L1 of all participants was Italian. Some participants of the control group reported some experience in L2 (English, N = 10; Spanish, N = 1; German, N = 1; French, N = 1), L3 (English, N = 3) and L4 (French, N = 1). Their experience in L2, L3 and L4 was constrained to foreign language courses, and they reported marginal daily use of any foreign language. Nine participants of the control group reported some experience in Italian dialects. Two participants were born in a foreign country (Germany, N = 1; Switzerland, N = 1), while all the others were born in Italy.
Participants were recruited using social networks and word of mouth. See Table 1 for background information about participants and their language knowledge.

Measures of bilingualism

There were significant differences in subjective language proficiency ratings, and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from Italian to L2. Bilinguals' language profile is summarized in Table 2.

Experimental data results

Table 4 shows the means and standard deviation of dependent variables for each experimental task.

Flanker task

Considering RTs, the main effects of Trial type ($F < 1$), Language Group ($F < 1$) and the Language Group x Trial type interaction ($F < 1$) were not significant. The one-way ANOVA on the Flanker effect showed no significant differences between groups ($F < 1$).

Considering accuracy, the main effects of Trial type ($p = .25$), Language Group ($p = .24$) and the Language Group x Trial type interaction ($F < 1$) were not significant. The one-way ANOVA on the Flanker effect showed no significant differences between groups ($p = .25$).

Simon task

Considering RTs, the main effect of Trial type was significant ($F_{1,32} = 32.91$; $p < .01$; $\eta^2 = .51$). Participants were faster in the congruent trials (529.77 ms) than the incongruent trials (563.80 ms). There was no main effect of Language Group ($F < 1$) and the Language Group x Trial type was not significant ($F < 1$). The one-way ANOVA on the Simon effect showed no significant differences between groups ($F < 1$).

Considering accuracy, the main effects of Trial type ($F < 1$), Language Group ($F < 1$) and the Language Group x Trial type interaction ($F < 1$) were not significant. The one-way ANOVA on the Simon effect showed no significant differences between groups ($F < 1$).

Verbal Stroop task

Considering RTs, the main effect of Trial type was significant ($F_{1,32} = 23.84$; $p < .01$; $\eta^2 = .43$), participants were faster in congruent trials (852.28 ms) than in the incongruent trials (915.01 ms). The main effect of Language group ($F < 1$) and the Language Group x Trial type interaction were not significant ($F < 1$). The one-way ANOVA on the verbal Stroop effect showed no significant differences between groups ($F < 1$).

Considering accuracy, the main effect of Trial type was significant ($F_{1,32} = 8.64$; $p < .01$; $\eta^2 = .21$). Participants were more accurate in the congruent trials (96.32%) than in the incongruent trials (92.52%). The main effect of Language group ($F < 1$) and the Language Group x Trial type interaction were not significant ($F < 1$). The one-way ANOVA on the verbal Stroop effect showed no significant differences between groups ($F < 1$).

Non-verbal Stroop task

Considering RTs, the main of Trial type was significant ($F_{1,30} = 29.34$; $p < .01$; $\eta^2 = .49$). The main effect of Language Group was not significant ($F < 1$). The Language Group x Trial type interaction was marginally significant ($F_{1,30} = 3.07$; $p = .09$; $\eta^2 = .09$). Both groups were faster in congruent trials than in the incongruent trials (bilinguals: $F_{1,30} = 27.41$; $p < .01$; $\eta^2 = .48$; control group: $F_{1,30} = 6.32$; $p = .02$; $\eta^2 = .17$). The one-way ANOVA on the non-verbal Stroop effect showed no significant differences between groups ($F_{1,30} = 3.07$; $p = .09$; $\eta^2 = .09$).

Considering accuracy, the main effect of Trial type was significant ($F_{1,30} = 9.95$; $p < .01$; $\eta^2 = .25$), participants were more accurate in congruent trials (95.24%) than in the incongruent trials (89.51%). The main effect of Language Group was marginally significant ($F_{1,30} = 2.93$; $p = .10$; $\eta^2 = .09$), with a higher accuracy in bilinguals than the control group (94.48% vs 90.28%). The Language Group x Trial type interaction was not significant

(p = .27). The one-way ANOVA on the non-verbal Stroop effect showed no significant differences between groups (p = .27).

Global-Local task

Considering RTs, the main effect of Condition was not significant (p = .24). The main effect of Trial type revealed a significant difference between groups ($F_{2,62}$ = 3.40; p = .04; η^2 = .10). Participants were faster in congruent trials than in incongruent trials ($F_{1,31}$ = 4.51; p = .04; η^2 = .13). The main effect of Language Group was marginally significant ($F_{1,31}$ = 3.13; p = .09; η^2 = .09), bilinguals tended to be faster than the control group (556.82 vs 665.95 ms). The Language Group x Trial type interaction was significant ($F_{2,62}$ = 4.02; p = .02; η^2 = .11). No differences emerged in neutral trials while bilinguals were marginally faster than the control group in congruent ($F_{1,30}$ = 3.95; p = .05; η^2 = .01) and incongruent trials ($F_{1,31}$ = 3.25; p = .08; η^2 = .10). All other interactions were not significant (p > .05). In Global block, the one-way ANOVAs on congruency effect (F < 1), costs (F < 1) and benefits (F < 1) were not statistically different between groups. In Local block, no differences emerged in congruency effects (p = .15) while costs ($F_{1,31}$ = 3.49; p = .07; η^2 = .10) and benefits ($F_{1,31}$ = 3.10; p = .09; η^2 = .09) were marginally significant. Bilinguals showed lower costs (-13.87 vs 104.10 ms) and higher benefits (23.87 vs -37.77 ms) than the control group. Total interference scores revealed no significant differences between groups (F < 1).

Considering accuracy, the main effect of Condition was not significant (F < 1). There was a main effect of Trial type ($F_{2,62}$ = 3.70; p = .03; η^2 = .11) and participants were more accurate in congruent trials than incongruent trials ($F_{1,31}$ = 8.81; p < .01; η^2 = .22). No other differences emerged (p > .05). There was no main effect of Language Group (F < 1) and no significant interactions (F < 1). In Global and Local blocks, the one-way ANOVAs on congruency effect, costs and benefits were not statistically significant (p > .05). Total interference scores revealed no significant differences between groups (p = .27).

Go/No-Go task

Considering accuracy and number of false alarms, no significant differences emerged between groups (F < 1). The RTs analysis showed a marginally significant difference ($F_{1,32}$ = 3.61; p = .07; η^2 = .10). Bilinguals were faster than the control group (396.91 vs 452.98 ms).

Table 4. Means and standard deviations (in parentheses) of inhibition measures for middle-aged adult participants.

	Middle-aged Bilinguals	**Middle-aged Control group**
Flanker task		
Reaction times		
Congruent trials	483.84 (75.59)	488.69 (71.86)
Incongruent trials	487.68 (69.01)	493.26 (67.92)
Flanker effect	3.84 (40.63)	4.57 (33.70)
Accuracy		
Congruent trials	93.63 (5.13)	96.09 (3.56)
Incongruent trials	92.65 (8.78)	94.53 (5.63)
Flanker effect	-0.98 (6.34)	-1.56 (6.06)
Simon task		
Reaction times		
Congruent trials	546.34 (120.39)	513.19 (63.56)
Incongruent trials	579.37 (125.96)	548.24 (65.57)
Simon effect	33.03 (36.19)	35.04 (32.91)
Accuracy		
Congruent trials	94.85 (3.76)	94.61 (5.26)
Incongruent trials	94.36 (5.30)	93.14 (6.58)
Simon effect	-0.49 (6.40)	-1.47 (6.74)
Verbal Stroop task		
Reaction times		

Congruent trials	851.63 (193.27)	852.93 (124.30)
Incongruent trials	919.91 (180.54)	910.12 (141.51)
verbal Stroop effect	68.28 (80.89)	57.19 (68.45)
Accuracy		
Congruent trials	96.32 (4.63)	96.32 (6.05)
Incongruent trials	92.40 (6.95)	92.65 (5.80)
verbal Stroop effect	-3.92 (8.26)	-3.68 (6.73)
Non-verbal Stroop task		
Reaction times		
Congruent trials	813.93 (133.25)	853.69 (222.03)
Incongruent trials	936.14 (144.01)	916.17 (170.90)
non-verbal Stroop effect	122.21 (94.71)	62.47 (97.96)
Accuracy		
Congruent trials	96.32 (4.63)	94.17 (10.54)
Incongruent trials	92.65 (6.83)	86.39 (11.41)
non-verbal Stroop effect	-3.68 (8.82)	-7.78 (11.66)
Global-Local task		
Global block		
Reaction times		
Congruent trials	530.80 (95.99)	625.31 (182.80)
Incongruent trials	632.73 (222.37)	549.92 (105.16)
Neutral trials	622.91 (165.50)	548.63 (96.63)
Global effect	19.12 (41.95)	7.42 (65.23)
Costs	1.29 (40.61)	9.82 (105.02)
Benefits	17.84 (37.05)	-2.40 (89.35)
Accuracy		
Congruent trials	95.95 (6.61)	96.48 (5.57)
Incongruent trials	93.38 (6.02)	92.19 (10.58)
Neutral trials	96.69 (5.46)	94.14 (6.64)
Global effect	-2.57 (9.39)	-4.30 (9.33)
Costs	-3.31 (7.37)	-1.95 (13.05)
Benefits	0.73 (9.08)	-2.34 (8.19)
Local block		
Reaction times		
Congruent trials	559.24 (99.04)	695.40 (376.62)
Incongruent trials	569.24 (88.16)	761.73 (502.41)
Neutral trials	583.11 (105.68)	657.63 (257.64)
Local effect	10.00 (38.54)	66.32 (152.80)
Costs	-13.87 (49.85)	104.10 (255.30)
Benefits	23.87 (39.73)	-37.77 (138.56)
Accuracy		
Congruent trials	95.59 (4.82)	96.09 (5.53)
Incongruent trials	92.97 (9.91)	93.75 (6.25)
Neutral trials	93.01 (7.62)	94.14 (6.64)
Local effect	-1.84 (6.90)	-3.12 (6.45)
Costs	0.73 (8.81)	-1.17 (9.47)
Benefits	-2.57 (7.35)	-1.95 (7.11)
Total interference score – RTs	29.12 (62.58)	73.74 (189.24)
Total interference score - ACC	-1.10 (12.35)	-5.47 (9.91)
Go/No-Go task		
RTs (Go trials)	396.91 (61.17)	452.98 (105.22)
Accuracy (Go trials)	97.06 (1.61)	96.84 (2.60)
False alarms	1.47 (1.18)	1.53 (1.66)

Older adults

Participants

38 participants aged over 55 years took part in the study.
19 participants were bilingual speakers of Italian and at least one additional language (mean age = 59.00, SD 4.04; 68.42% females). Bilinguals' L1 were Italian (N = 15), Dutch (N = 2), English (N = 1), Portuguese (N = 1). Bilinguals' L2 were English (N = 8), Italian (N = 4), Spanish (N = 2), French (N = 2), German (N = 1), Portuguese (N = 1), modern Greek (N = 1). Most of the participants were born in Italy (N = 15) while some of them in a foreign country (Brazil, N = 1; Netherlands, N = 1; Belgium, N = 1; UK, N = 1).
Some bilingual participants reported some experience in L3 (English, N = 7; Portuguese, N = 1; French, N = 1; Dutch = 1) and L4 (French, N = 3; Spanish, N = 2; German, N = 1; English, N = 1; Russian, N = 1). Three participants reported some experience in Italian regional dialects.
The control group included 19 participants (mean age = 59.58, SD 3.85; 78.95% females). The L1 of all participants was Italian. Some participants of the control group reported some experience in L2 (English, N = 6; French, N = 2; Portuguese, N = 1), L3 (English, N = 1; Spanish, N = 1; French, N = 1; German, N = 1; Russian, N = 1) and L4 (English, N = 1). Their experience in L2, L3 and L4 was constrained to foreign language courses, and they reported marginal daily use of any foreign language. Five participants of the control group reported some experience in Italian dialects. Two participants were born in a foreign country (UK, N = 1; Switzerland, N = 1) while all the others were born in Italy. Participants were recruited using social networks and word of mouth. See Table 1 for background information about participants and their language knowledge.

Measures of bilingualism

There were significant differences on subjective language proficiency ratings and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from Italian to L2. Bilinguals' language profile is summarized in Table 2.

Experimental data results

Table 5 shows the means and standard deviation of dependent variables for each experimental task.

Flanker task

Considering RTs, the main effect of Trial type (p = .19), Language Group (F < 1) and the Language Group x Trial type interaction (p =.26) were not significant. The one-way ANOVA on the Flanker effect showed no significant differences between groups (p = .26).
Considering accuracy, the main effect of Trial type (p = .15), Language Group (F < 1) and the Language Group x Trial type interaction (F < 1) were not significant. The one-way ANOVA on the Flanker effect showed no significant differences between groups (F < 1).

Simon task

Considering RTs, the main effect of Trial type was significant ($F_{1,34}$ = 12.96; p < .01; η^2 = .27). Participants were faster in the congruent trials (577.55 ms) than the incongruent trials (603.45 ms). The main effect of the Language Group (F < 1) and the Language Group x Trial Type interaction (F < 1) were not significant. The one-way ANOVA on the Simon effect showed a marginal significant difference between groups ($F_{1,34}$ = 4.05; p = .05; η^2 = .11) and bilinguals showed a lower Simon effect than the control group (11.41 vs 40.38 ms).
Considering accuracy, the main effect of Trial type (p = .11), Language Group (p = .32) and the Language Group x Trial type interaction (F < 1) were not significant. The one-way ANOVA on the Simon effect showed no significant differences between groups (F < 1).

Verbal Stroop task

Considering RTs, the main effect of Trial type was significant ($F_{1,33} = 29.46$; $p < .01$; $\eta^2 = .47$), participants were faster in congruent trials (1011.73 ms) than in the incongruent trials (1114.11 ms). The main effect of Language Group and the Language Group x Trial type interaction were not significant ($F < 1$). The one-way ANOVA on the verbal Stroop effect showed ($F < 1$).
Considering accuracy, the main effect of Trial type was significant ($F_{1,33} = 12.46$; $p < .01$; $\eta^2 = .27$). Participants were more accurate in the congruent trials (96.88%) than in the incongruent trials (92.75%). The main effect of Language Group and the Language Group x Trial type interaction were not significant ($p = .32$). The one-way ANOVA on the verbal Stroop effect showed ($p = .32$).

Non-verbal Stroop task

Considering RTs. the main of Trial type was significant ($F_{1,33} = 7.48$; $p < .01$; $\eta^2 = .18$). Participants were faster in the congruent trials (970.34 ms) than the incongruent trials (1030.73 ms). The main effect of Language Group and the Language Group x Trial type interaction were not significant ($F < 1$). The one-way ANOVA on the non-verbal Stroop effect showed ($F < 1$).
Considering accuracy, the main effect of Trial type was significant ($F_{1,33} = 12.40$; $p < .01$; $\eta^2 = .27$). Participants were more accurate in the congruent trials (97.62%) than in the incongruent trials (93.07%). The main effect of Language Group and the Language Group x Trial type interaction were not significant ($F < 1$). The one-way ANOVA on the non-verbal Stroop effect showed ($F < 1$).

Global-Local task

Considering RTs, the main effect of Condition was not significant ($F < 1$). The main effect of Trial type was significant ($F_{2,72} = 7.99$; $p < .01$; $\eta^2 = .18$), participants were faster in congruent than in incongruent trials ($F_{1,36} = 16.92$; $p < .01$; $\eta^2 = .32$). Participants were also faster in neutral trials than incongruent trials ($F_{1,36} = 4.45$; $p = .05$; $\eta^2 = .11$). No differences emerged between congruent and neutral trials ($F_{1,36} = 3.12$; $p = .08$; $\eta^2 = .08$). There was no main effect of Language Group ($F < 1$). All the interactions were not significant ($p > .05$). In Global block, the one-way ANOVAs on congruency effect ($F < 1$) and costs ($p = .25$) were not significant different between groups. Bilinguals shower marginal higher benefits than the control group ($F_{1,36} = 3.55$; $p = .07$; $\eta^2 = .09$). In Local block, the one-way ANOVAs on congruency effect, costs and benefits were not significant different ($F < 1$). Total interference scores revealed no significant differences between groups ($F < 1$).
Considering accuracy, the main effect of Condition ($p = .22$), Trial type ($p = .11$) and Language Group ($p = .22$) were not significant. All the interactions were not significant ($p > .05$). In Global and Local blocks, the one-way ANOVAs on congruency effect, costs and benefits were not statistically significant ($p > .05$). Total interference scores revealed no significant differences between groups ($F < 1$).

Go/No-Go task

Considering RTs ($F < 1$), accuracy in Go trials ($F < 1$) and false alarms ($F < 1$) no significant differences between groups emerged.

Table 5. Means and standard deviations (in parentheses) of inhibition measures for older adult participants

	Older Bilinguals	Older Control group
Flanker task		
Reaction times		
Congruent trials	532.16 (65.45)	519.15 (84.44)
Incongruent trials	533.48 (67.45)	537.52 (72.90)
Flanker effect	1.33 (42.51)	18.37 (47.97)
Accuracy		
Congruent trials	94.21 (5.37)	92.98 (9.42)
Incongruent trials	96.53 (2.95)	94.30 (8.24)

Flanker effect	2.31 (7.73)	1.32 (7.49)
Simon task		
Reaction times		
Congruent trials	592.72 (95.73)	562.38 (111.73)
Incongruent trials	604.14 (82.58)	602.76 (118.71)
Simon effect	11.41 (9.89)	40.38 (10.45)
Accuracy		
Congruent trials	96.93 (3.63)	96.08 (4.77)
Incongruent trials	95.61 (3.53)	92.89 (11.85)
Simon effect	-1.32 (6.06)	-3.19 (10.05)
Verbal Stroop task		
Reaction times		
Congruent trials	1053.78 (325.70)	969.69 (391.47)
Incongruent trials	1165.95 (440.51)	1062.28 (413.04)
verbal Stroop effect	112.17 (143.06)	92.59 (62.57)
Accuracy		
Congruent trials	97.68 (2.94)	96.08 (3.75)
Incongruent trials	92.36 (6.74)	93.14 (6.90)
verbal Stroop effect	-5.32 (7.80)	-2.94 (5.85)
Non-verbal Stroop task		
Reaction times		
Congruent trials	985.23 (205.80)	955.44 (200.77)
Incongruent trials	1030.75 (221.79)	1030.72 (232.57)
non-verbal Stroop effect	45.52 (76.11)	75.28 (170.31)
Accuracy		
Congruent trials	97.45 (4.08)	97.79 (2.99)
Incongruent trials	93.98 (5.39)	92.16 (6.40)
non-verbal Stroop effect	-3.47 (7.60)	-5.64 (7.65)
Global-Local task		
Global block		
Reaction times		
Congruent trials	621.00 (84.91)	670.60 (184.80)
Incongruent trials	649.37 (91.61)	691.84 (209.27)
Neutral trials	639.40 (99.22)	655.01 (158.10)
Global effect	28.37 (41.13)	21.24 (66.03)
Costs	9.97 (36.59)	36.83 (93.07)
Benefits	18.40 (31.58)	-15.58 (71.96)
Accuracy		
Congruent trials	97.70 (3.73)	96.38 (7.01)
Incongruent trials	94.74 (6.34)	95.39 (4.58)
Neutral trials	97.04 (4.83)	97.37 (4.33)
Global effect	-2.96 (7.62)	-0.99 8.14)
Costs	-2.30 (7.85)	-1.97 (6.27)
Benefits	-0.66 (5.85)	0.99 (5.61)
Local block		
Reaction times		
Congruent trials	636.99 (105.27)	649.18 (169.45)
Incongruent trials	663.06 (92.19)	676.06 (205.79)
Neutral trials	649.72 (102.57)	668.55 (184.919
Local effect	26.06 (51.62)	26.88 (62.67)
Costs	13.34 (46.91)	7.51 (59.90)
Benefits	12.73 (39.94)	19.37 (46.56)
Accuracy		
Congruent trials	97.70 (3.10)	92.76 (17.46)
Incongruent trials	95.07 (5.73)	91.45 (11.64)
Neutral trials	96.05 (5.19)	92.10 (11.57)

Local effect	-2.63 (6.01)	-1.31 (10.54)
Costs	-0.99 (6.67)	-0.66 (7.48)
Benefits	-1.64 (5.45)	-0.66 (10.60)
Total interference score - ACC	-5.59 (9.29)	-2.30 (17.21)
Total interference score - RTs	665.37 (114.43)	670.43 (201.29)
Go/No-Go task		
RTs (Go trials)	452.72 (67.26)	471.57 (105.28)
Accuracy (Go trials)	97.32 (1.58)	97.06 (2.18)
False alarms	0.72 (1.13)	0.88 (0.78)

Cross-task coherence

The nonverbal Stroop effect showed a marginal significant correlation with the verbal Stroop effect (r = .14; p = .07) and a significant correlation with the number of FA (r = .17; p = .03). The Flanker effect showed a marginal significant correlation with the total interference score of the Global-Local task (r = -.13; p = .09). All other pairs of effect indicated that the cross-task coherence was very low (al r between -.11 and .08).

General Discussion

This study aimed to compare the performance of the bilingual population with that of a "monolingual" control group in the main experimental tasks used to assess inhibition ability. Most of the results show no significant differences between the bilingual population and the control group. Considering young adults, they showed faster overall RTs (i.e., an indicator of monitoring ability) in the verbal Stroop and marginally in the non-verbal Stroop. Seven studies previously used a nonverbal version of the Stroop task. Four studies the numerical version (Antón et al., 2019; Antón et al., 2016; Shulley & Shake, 2016; Xie & Dong, 2017), while three studies the spatial version (Blumenfeld & Marian, 2014; Damian et al., 2018; Zhou & Krott, 2018) of the Stroop task. In half the studies, no significant differences emerged, whereas, in the remaining 50%, there was evidence in favor of a bilingual advantage. Although both are nonverbal versions of the task, it is important to consider the nature of the stimuli when interpreting the results.

The numerical version cannot be considered nonverbal but only a less linguistically charged version (Antón et al., 2016). The spatial version includes directional stimuli that create a greater conflict than non-directional stimuli (such as those included in this study). A version with nonverbal stimuli was adopted in the present study, which had never been used in bilingualism studies. There is a need to investigate the effect of stimuli on performance on the nonverbal Stroop task and see which version is more relatable to the verbal version of the task. The result on verbal Stroop is contrary to what was hypothesized. Verbal stimuli do not seem to have adversely affected the bilinguals' performance. As highlighted in the review of the previous chapter, although most of the studies using the verbal Stroop did not show any differences between the language groups, some studies have reported slower reaction times for the bilingual group (e.g., Damian et al., 2018; Shulley & Shake, 2016). It could be speculated that since the participants in this study were dominant in the Italian language, they were not affected by the presence of the verbal stimuli because they were assessed in Italian. In contrast to previous studies analyzing the Go/No-Go task (e.g., Barbu et al., 2020; Hofweber et al., 2020b), bilinguals were less accurate in the Go trials than the control group. Moreover, in the Global-Local task, there was a tendency for the bilingual population to have higher costs in the Local block than the control group. RTs showed that the control group responded with the same speed in both neutral and incongruent trials, while bilinguals were faster in neutral trials than incongruent trials.

Even considering the population of middle-aged adults, no significant differences emerged in most of the tasks. The bilingual population showed a better performance in the Global-Local task with faster general reaction times than the control group. Bilinguals responded faster in congruent trials and tended to respond faster in incongruent trials than the control group. Moreover, in the Local block, bilinguals tended to have lower costs and higher benefits than the control group. In the Go/No-Go task, bilinguals were faster than the control group. Finally, in the non-verbal Stroop, the accuracy of the bilingual population tended to be higher than that of the control group. Finally, when considering the population of older adults, there were no significant differences in any of the tasks considered. According to Grundy (2020), although differences between bilingual and monolingual groups rarely emerge, they tend to benefit the bilingual population. The results of this study are in line with this statement as the significant differences that emerged tend to favor the bilingual group. Although it has been hypothesized that results in favor of the bilingual advantage are more frequent in populations in which executive functions are in a developmental or declining phase (e.g., Antón et al., 2014), in this study, we highlight the results in the population of young adults and middle-aged adults. These findings align with Paap's (2019) findings; even in younger populations, considered at the ceiling can improve cognitive abilities through training.

Limitations and conclusion

Some limitations should be considered when interpreting the results. The present study was conducted entirely online and not in a controlled laboratory setting. This aspect may have affected the participants' performance, although the high percentage of accuracy reached in all tasks (greater than 94%) suggests a marginal influence of the adopted procedure. The high accuracy rate could also indicate that the selected version of the task was too easy for the participants included in the study, who were at the peak of their cognitive development. The participants in the control group could be considered low-proficient bilinguals if we adopted a different definition of bilingualism (e.g., incipient bilingual, a person at the early stages of bilingualism where one language is not fully developed; Wei, 2021). To date, selecting native Italian speakers who do not know a second language is almost impossible. In fact, in Italy, teaching a foreign language was compulsory in 1962, while from 1987-88, an experiment for teaching a second foreign language started in middle school. Subsequently, since 2004, the teaching of a second foreign language has become compulsory. Moreover, most degree courses at Italian universities include at least one English language course. Finally, as previously pointed out by other studies (e.g., Antón et al., 2019), even though these experimental tasks are considered measures of inhibition ability, they showed a low cross-task coherence. In conclusion, the results of this study seem to indicate a better general monitoring capacity highlighted by the faster overall RTs. Conversely, the results did not evidence a better inhibition capacity. These results are in line with the findings of Bialystok and Craik (2022): bilinguals seem to have better attentional control skills than better inhibitory skills. This aspect will be discussed in detail in the General Conclusion (pages 93-95).

Chapter 3: Executive functioning during verbal fluency tasks in bilinguals: a systematic review

Introduction

Over the last years, scientific interest in bilingualism is considerably increased. Considering the number of papers published on bilingualism in the past 25 years according to Web of Science (Thompson-Reuters, 2012), more than 70% of scientific articles have been published in the last 10 years. Being bilingual is more common than being monolingual (Bennet & Verney, 2019), and most of the world population knows two or more languages, and this trend is increasing (Marsh et al., 2019). Therefore, bilingualism is a sociodemographic factor to be considered during cognitive and clinical assessments (Bennet & Verney, 2019), especially considering the bilingualism effect that, according to some authors, would enhance cognitive functionality and counteract cognitive decline (Massa et al., 2020). Specifically, studies on bilingualism reported an effect on some cognitive functions (for a review, see Giovannoli et al., 2020; Lehtonen et al., 2018; Zhang et al., 2020) and language proficiency (for a review, see Kroll & Bialystok, 2013). It is hypothesized that bilingualism positively affects cognitive functions due to constantly monitoring the known languages to select the most suitable one for each interactional context (Joint Activation Model; Green, 1998). Furthermore, the context in which language exchanges take place would influence the development of cognitive functions (Adaptive Control Hypothesis; Green & Abutalebi, 2013). Recently, the publication of studies with null findings has led to questioning the positive effect of bilingualism (e.g., Antoniou, 2019; Paap et al., 2015). Conversely, bilingualism seems to affect language ability negatively due to less use of each known language (e.g., Bialystok, 2009). Bilinguals would seem to have a smaller vocabulary size than monolinguals; however, this difficulty would occur only when conceptual scores (i.e., vocabulary assessment considering both known languages) are not used to assess children's vocabulary size (e.g., Gross et al., 2014). A recent study (Paplikar et al., 2021) suggested that bilingual adults should also be assessed considering their known languages during neuropsychological assessments. Several authors argued that verbal stimuli in cognitive assessment tasks could negatively affect the performance of bilingual participants (e.g., Grundy et al., 2017). For these reasons, many studies have used tasks with non-verbal stimuli (e.g., Antón et al., 2014; Rosselli et al., 2016). Therefore, the authors have evaluated language production and executive functioning separately. Conversely, some authors suggest that it would be useful to use instruments, such as verbal fluencies, to assess these aspects simultaneously (e.g., Patra et al., 2020).
The verbal fluency task is a neuropsychological measure of lexical retrieval efficiency. According to a given criterion, it requires a time-restricted generation of as many words as possible. Participants must produce items belonging to a specific category in the semantic fluency condition (e.g., animals, fruits, clothing) while words starting with a given letter in the letter fluency condition. Both conditions require semantic memory and executive control functions such as working memory (e.g., participants must keep the instructions in mind), response inhibition, and conflict monitoring (e.g., participants must inhibit irrelevant responses and repetition). Higher demands are placed on executive control mechanisms in letter condition, while in semantic fluency, greater emphasis is placed on linguistic abilities (Luo et al., 2010; Patra et al., 2020). Indeed, generating words based on a phonemic cue is rarely done in everyday speech production. At the same time, concepts are clustered in semantic proprieties and resemble everyday production tasks (e.g., making a shopping list). Furthermore, letter fluency requires that participants inhibit irrelevant semantic associations. Although there are potentially more words for a letter category than a semantic category, individuals tend to generate fewer items during letter fluency than semantic fluency (Friesen et al., 2015).

To perform this task efficiently, subjects use clustering (i.e., generation of words within a subcategory) and switching (i.e., shifting from one subcategory to another) strategies. The word production over time is not

linear but is distributed in temporal clusters divided by a pause that signals the transition to a different semantic cluster (Patra et al., 2020).
Most studies adopt a quantitative approach to evaluate the performance of individuals in verbal fluency tasks. Indeed, the main indices used to assess efficiency in performing this task are the total (correct) number of words produced and the number of errors. These indices allow assessing both the linguistic proficiency and the executive functions (Patra et al., 2020).
Another quantitative index is the Fluency Difference Score (FDS), the differences in the number of correct responses between semantic and letter fluency conditions as a proportion of correct responses in the semantic fluency condition. Moreover, it is possible to analyze the performance qualitatively, considering the strategies used to complete the task and dividing the production into temporal clusters. The cluster size is considered an indicator of linguistic competence, while the number of switches indicates executive functioning. Other indices are “cluster switches", such as shifting between adjacent (e.g., canine and feline as in “dog, wolf, cat, lion”) or overlapping clusters (e.g., pet and canine as in “cat, dog, wolf”), and "hard switches", defined as shifting from a cluster to non-clustered words (e.g., pet and non-clustered as in “dog, cat, octopus”) or between two non-clustered words (e.g., “lion, butterfly”) (Abwender et al., 2001).
Lastly, analyzing the temporal distribution of the words produced can provide valuable insights into both the executive and linguistic components. Specifically, 1st RT (i.e., the time interval from the beginning of the trial to the onset of first response) and initiation parameter (i.e., starting point of the logarithmic function that is the value of y when $t = 1$ or $\ln(t) = 0$) provide information about the language component. Sub-RT (i.e., the average value of the time intervals from the onset of first response to the onset of each subsequent response) and slope (i.e., the shape of the curve that reflects how resources are monitored and used over time) on the executive components. A flatter slope indicates that participants could maintain their performance across the task despite higher interference towards the end of the trial (Friesen et al., 2015).
This systematic review aims to summarize the findings of studies investigating the performance of bilingual young and older adults in verbal fluency tasks.

Method

The review process was conducted according to the PRISMA-Statement (Liberati et al., 2009; Moher et al., 2009). The PRISMA Statement consists of a 27-item checklist and a four-phase flow diagram and helps authors improve systematic review reporting. The protocol was not registered.

Research strategies

A systematic search of the international literature was conducted in the following electronic databases by selecting articles published in peer-review journals: PsycINFO, PsycARTICLES, MEDLINE, PubMed, Web of Science, and SCOPUS. The last research was conducted on 13 May 2021. Restrictions were made limiting the research to academic publications in English, Italian, and Spanish. No restriction of age, gender, or ethnicity was made. The search strategy used Boolean combinations of the following keywords: “bilingual*”, “second language”, “executive function*”, “verbal fluency”, “semantic fluency”, “category fluency”, “phonologic* fluency” and “letter fluency”. Reference lists of the selected articles were screened. A total of 3975 articles were obtained from the search procedure. Mendeley reference manager software was used for removing duplicates. The first screening was made by reading the title and abstract. The full text of the selected studies was read.

Eligibility Criteria

The studies that respected the following characteristics were included (a) the presence of at least one bilingual group and one monolingual group, (b) at least one verbal fluency task, (c) age over 18 years. Studies on bimodal bilingual, second language learners, and trilingual or multilingual people were excluded. Studies on clinical populations were excluded.

Data Collection

According to the PICOS approach (Liberati et al., 2009), the following information has been extracted from the selected studies: author(s) and year of publication, country, characteristics of participants (age, percentage of women, spoken languages, use of languages, socioeconomic status, level of education), the experimental paradigm used, results of the studies. These data are summarized in Table 1.

Quality assessment

All the selected studies were screened to assess the risk of bias using the Standard quality assessment criteria for evaluating primary research papers from various fields (Kmet et al., 2011). The studies were included if they reached a score above 70%. Of the thirty-five studies, thirty-two met the criteria for very high-quality studies (total score > 90%) and three studies reached the high-quality threshold (total score > 80%). The checklist items that reported the lowest scoring levels were those concerning the description of the subject/comparison group and the adequacy of sample size.

Results

Selection of studies

The flowchart (Figure 1) shows the number of studies identified from the databases and the other sources, the number of studies examined by the authors, and assessed for eligibility. The reasons for exclusion are reported.

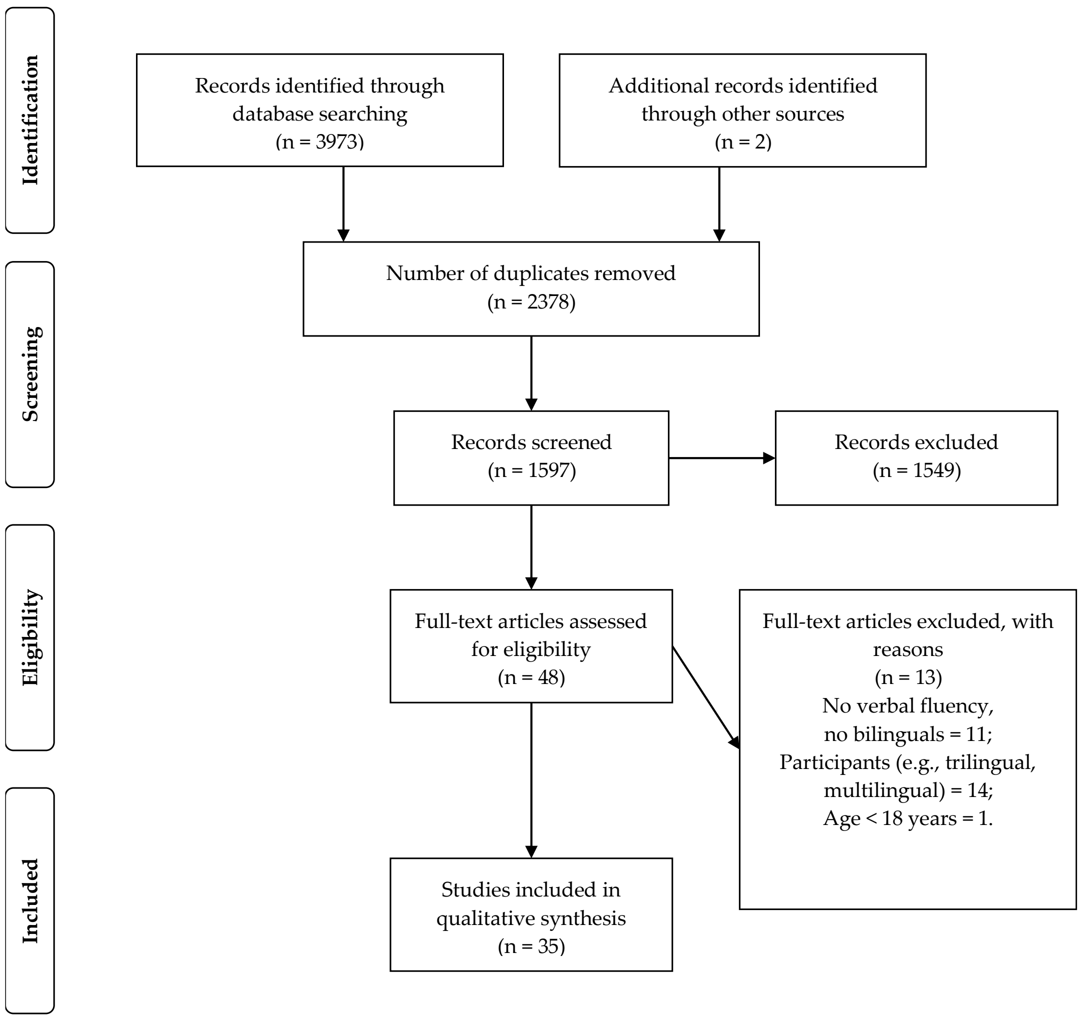


Figure 1. Studies selection flow diagram (PRISMA flowchart).

Table 1. Participants' characteristics in the included studies.

	Participants								
	Group	**N**	**Age mean (SD)**	**Sex (% female)**	**Education**	**SES**	**Country**	**Language**	**AoA**
Anderson et al., 2017	Total	35					Canada		
	M	n.r.	74.9 (4.6)		17.5 (4.0)			ENG	
	B	n.r.	74.7 (3.9)		16.7 (2.7)			ENG – ML2	
Ansaldo et al., 2015	M	10	74.5 (7.1)		16.1 (3.3)		Canada	FR	
	B	10	74.2 (7.4)		17.2 (3.1)		(Montreal)	FR – ENG	
Bennett & Verney, 2019	M	30	19.6 (1.3)	63	12.8 (0.9)	18.2 (2.9)	Mexico	ENG	SPA-L1: 75%
	B	99	20.0 (3.2)	66	12.7 (1.1)	11.0 (4.4)*		ENG - SPA	ENG-L1: 7%
	EDB	48	19.7 (2.8)	65	12.6 (1.0)	11.2 (4.3)			ENG+SPA: 18%
	BB	25	19.8 (2.1)	64	12.8 (1.2)	10.4 (4.2)			
	SDB	26	20.7 (4.5)	81	12.9 (1.1)	11.4 (5.0)			
Bialystok et al., 2009	YM	24	20.7		12.8 (1.3)		Canada	ENG	
	YB	24	19.7		12.4 (0.9)			ENG – ML2	
	OM	24	67.2		14.4 (1.4)			ENG	
	OB	24	68.3		14.2 (2.4)			ENG - ML2	
Bialystok et al., 2008	Study 1						Canada	ENG	50% < 3 y
	M	24	20.7 (n.r.)					ENG – ML2	50 % abt 6 y
	B	24	19.7 (n.r)						
	HB	12							
	LB	12							
	Study 2								
	M	16	21.3 (1.3)						
	B	50	21.4 (1.8)						
	HB	26	21.5 (1.9)						
	LB	24	21.2 (1.8)						
Blumenfeld et al., 2016	M	24	20.9 (0.7)		15.3 (0.5)			ENG	ENG: 0.4 (0.1), SPA/L2: 13.5 (0.8)
	B	25	22.0 (1.0)		15.4 (0.6)			ENG - SPA	ENG: 1.8 (0.5), SPA/L2: 2.0 (0.6)
Bogulski et al., 2015	M	27	20.8 (1.9)					ENG	
	LapB	22	21.7 (4.1)					ENG – FR	
	FullB	30	21.4 (2.3)					ENG – FR	
Friesen et al., 2015	YM	20	20.7 (1.3)					ENG	
	YB	20	21.1 (1.3)					ENG – ML2	L2: 2.9 (4.2)

	OM	20	70.9 (2.6)					ENG	
	OB	21	71.1 (3.8)					ENG – ML2	L2: 8.8 (5.9)
Gollan et al.,	M	30	19.6 (1.6)	77				ENG	ENG: 0.0 (0.0), SPA: 12.2 (3.7)
2002	B	30	20.0 (2.0)	63				SPA – ENG	ENG: 3.4 (2.7), SPA 0.1 (0.6)
Keijzer &	DM	17	78.3 (n.r.)	53	EM > DM		the Netherlands	NL	
Schmid, 2016	EM	16	76.3 (n.r.)	69	B = EM,		Australia	ENG	
	B	29	77.9 (n.r.)	45	DM		Australia	NL - ENG	
Kousaie et	Total						Canada		
al., 2014	YEM	40	21.5 (1.5)	37	15.5 (1.1)		Ottawa	ENG	
	YFM	30	21.8 (2.5)	33	15.1 (1.4)		Quebec	FR	
	YB	51	21.5 (2.3)	35	15.5 (1.5)		Ottawa	FR – ENG	
	OEM	31	72.3 (6.4)	48	15.3 (2.9)		Ottawa	ENG	
	OFM	30	72.6 (6.6)	77	16.2 (2.6)		Quebec	FR	
	OB	36	70.7 (2.6)	47	16.1 (2.8)		Ottawa	FR - ENG	
Kowoll et al.,	M	6	70.2 (8.2)				Germany	DE	
2015	B	11	68.2 (13.2)					DE – ML2	
Ljungberg et	M	26	57.5	77	11.6		Sweden	SWE	
al., 2020	SFB	26	57.5	77	11.6			SWE – FIN	
	SWEB	26	57.5	77	11.6			SWE – ENG	
Ljungberg et	M	74	54.8 (9.3)	27	10.7 (2.0)		Sweden	SWE	
al., 2013	B	104	46.5 (8.9)	53	14.3 (2.1)			SWE – ML2	
Luo et al.,	M	20	20.6 (1.3)				Canada	ENG	
2010	HVB	20	21.1 (1.4)					ENG – ML2	inf. L2: 2.9 (5.1); for. L2: 7.0 (3.6)
	LVB	20	20.3 (1.6)					ENG - ML2	inf. L2 4.3 (4.2); for. L2: 8.0 (3.4)
Marsh et al.,	M	139	55.9 (7.0)	50	B > M	B = M	Sweden	SWE	
2019	B	58	44.1 (8.3)	55				SWE - ML2	
Massa et al.,	YM	16	23.4 (4.5)		16.9 (2.6)	M = B		FR	
2020	YB	16	25.6 (3.9)		15.7 (2.3)			FR – ITA	11.1 (8.7)
	OM	16	71.1 (5.9)		16.0 (2.7)			FR	
	OB	16	72.3 (5.0)		15.1 (2.4)			FR - ITA	10.2 (9.3)
					-				
Morrison et	M	23	19.7 (2.3)	74	14.6 (1.8)		Canada	ENG	FR: 6-1 (1.5)
al., 2019	B	21	19.7 (1.6)	71	14.7 (1.5)		(Ottawa)	ENG – FR	FR: 4.0 (2.3)
Morrison et	M	26	20.2 (2.2)	61	15.0 (1.8)		Canada	ENG	FR: 6.1 (1.6)
al., 2020	B	28	20.5 (2.1)	71	15.2 (1.8)		(Ottawa)	ENG – FR	FR: 4.6 (2.7)
Paplikar et	Total						India		
al., 2021	HM	74	50.7 (5.5)	38	14.6 (2.1)		New Delhi/ Hyderabad	HIN	

	HB	67	50.8 (5.3)	34	15.1 (2.5)		New Delhi/ Hyderabad	HIN – ML2	
	BEM	32	55.7 (8.3)	56	10.8 (2.5)		Kolkata	BEN	
	BEB	32	53.7 (9.9)	37	11.9 (2.2)		Kolkata	BEN – ML2	
	TM	80	52.8 (8.1)	47	10.9 (3.3)		Hyderabad	TEL	
	TB	80	55.0 (10.5)	35	11.6 (2.3)		Hyderabad	TEL – ML2	
	KM	41	60.6 (13.8)	68	9.2 (2.0)		Bengaluru/Belgaum	KAN	
	KB	42	61.9 (13.7)	40	9.8 (1.8)		Bengaluru/Belgaum	KAN – ML2	
	MM	40	56.3 (9.2)	62	12.9 (2.1)		Trivandrum	MAL	
	MB	42	57.1 (10.3)	52	13.6 (1.8)		Trivandrum	MAM - ML2	
Patra et al., 2020	M	25	30.4 (8.2)	52	B = M		UK	ENG	
	B	25	32.8 (4.8)	44				BEN - ENG	
Portocarrero, et al., 2007	M	39	19 (n.r.)				U.S.	ENG	
	B	39	19 (n.r.)					ENG - ML2	
Prior & Gollan, 2011	M	47	20.2 (1.5)	62		Parental education	U.S. (California)	ENG	
	SEB	41	20.0 (1.6)	83				SPA – ENG	L1: 2.7 (1.9)
	MEB	43	19.4 (1.2)	74		SEB < MEB, M MEB = M		MAN – ENG	L1: 2.7 (2.2)
Rosselli et al., 2000	EM	45	63.4 (10.1)	67	16.6 (2.4)		U.S. (South Florida)	ENG	
	SM	18	61.3 (8.1)	78	13.3 (4.8)			SPA	
	SEB	19	60.6 (9.7)	53	14.5 (3.6)			SPA - ENG	
Rosselli et al., 2002[a]	EM	45	63.4 (10.1)	67	16.6 (2.4)		U.S. (South Florida)	ENG	
	SM	18	61.3 (8.1)	78	13.3 (4.8)			SPA	
	SEB	19	60.6 (9.7)	53	14.5 (3.6)			SPA - ENG	
Sadat et al., 2016	M	24	23.6 (4.4)	71			Portugal	POR	
	B	24	34.8 (5.7)	71				POR - FR	
Sandoval et al., 2010	M	30	19.7 (1.3)		14.3 (1.7)			ENG	ENG: 0.2 (0.4), L2: 10.8 (5.7)
	B	24	20.3 (2.9)		14.2 (2.6)			SPA-ENG	ENG: 2.1 (2.1), L2: 0.5 (0.9)
Seçer, 2016	M	88	age range 18-34 y	58		M = B	Cyprus	TUR	
	B	74		43				TUR - ENG	ENG: 11.83 (4.11)
Soltani et al., 2021	M	12	70.4					FA	
	B	24	72.8					ARA – FA	
	AB	12						L1: ARA	
	PB	12						L1: FA	
Taler et al., 2013	M	32	21.6 (1.6)	47	15.6 (1.1)		Canada (Ottawa, Gatineau)	ENG	
	B	38	21.5 (2.3)	39	15.5 (1.6)			ENG – FR	L1, L2 < 13 y
Tao et al., 2015	M	60	20.7 (2.5)	73		SEB < M	U.S. (California)	ENG	
	SEB	80	20.9 (2.5)	75		MEB > SEB		SPA – ENG	L1, L2 f.b.

	MEB	80	19.8 (1.1)	75	MEB = M		MAN - ENG	L1, L2 f.b.
Vega-Mendoza et al., 2015	Exp 1					Scotland (Edinburgh)		
	M	18	21.78 (2.18)	67			ENG	
	B	16	22.44 (1.97)	81			ENG – SPA	
	ML	17	20.82 (1.70)	82			ENG – SPA - ML2	
Woumans et al., 2015	M	30	22.1 (1.4)	73	M = B	Belgium	FR	
	unbB	34	22.3 (2.8)	79			NL – FR	L1: f.b.; L2: 9.4 (1.3)
	BB	31	21.1 (2.1)	77			NL – FR	L1: f.b.; L2: 2.6 (3.0)
Woumans et al., 2019	M	16	18.56 (0.63)	94	EnM > ChM, B	Belgium	NL	L1: f.b.; L2: 4.72 (5.20)
	B	18	19.82 (4.81)	83			NL - ML2	
Zeng et al., 2019	Exp 2					Australia		
	M	20	23.1 (3.4)	45			ENG	
	B	20	22.5 (3.4)	70			ENG - ML2	

Note. N: number of participants; SES: socioeconomic status; AoA: age of acquisition; M: monolinguals; B: bilinguals; n.r.: not reported; ENG: English; ML2: multiple L2s; FR: French; EDB: English dominant bilinguals; BB: balanced bilinguals; SDB: Spanish dominant bilinguals; SPA: Spanish; YM: young monolinguals; YB: young bilinguals; OM: old monolinguals; OB: old bilinguals; HB: high proficiency bilinguals; LB: low proficiency bilinguals; LapB: lapsed bilinguals; FullB: full bilinguals; DM: Dutch monolinguals; EM: English monolinguals; NL: Dutch; YEM: young English monolinguals; YFM: young French monolinguals; OEM: old English monolinguals; OFM: old French monolinguals; DE: German; SFB: Swedish-Finnish bilinguals; SWEB: Swedish-English bilinguals; SWE: Swedish; FIN: Finnish; HVB: high vocabulary bilinguals; LVB: low-vocabulary bilinguals; inf.: informal acquisition; for.: formal acquisition; ITA: Italian; HM: Hindi monolinguals; HB: Hindi bilinguals; BEM: Bengali monolinguals; BEB: Bengali bilinguals; TM: Telegu monolinguals; TB: Telegu bilinguals; KM: Kannada monolinguals; KB: Kannada bilinguals; MM: Malayalam monolinguals; MB: Malayalam bilinguals; HIN: Hindi; BEN: Bengali; TEL: Telegu; KAN: Kannada; MAM: Malayalam; SEB: Spanish-English bilinguals; MEB: Mandarin-English bilinguals; MAN: Mandarin:; POR: Portuguese; TUR: Turkish:; AB: Arabic bilinguals; PB: Persian bilinguals; FA: Persian; ARA: Arabic; ML: multilinguals; unbB: unbalanced bilinguals; f.b.: from birth.

[a]same participants of Rosselli et al., 2000

Results of the selected studies

Twenty-nine studies used phonological fluency, while thirty studies used semantic fluency. Fourteen different stimuli were used for letter fluency. The most used letters were F (20 studies), A (20 studies), and S (16 studies). Some studies (Bennet & Verney, 2019; Gollan et al., 2002) included multiple letters to test the influence of the type of stimulus used on participants' performance. Sandoval et al. (2010) required participants to produce words that began with a letter pair (24 different letter pairs, divided into high- and low-frequency word categories). A total of thirty-three categories were used in the semantic fluency task. The most used categories were animals (21 studies) and clothing items (9 studies). Two studies (Gollan et al., 2002; Sandoval et al., 2010) compared participants' performance by including multiple categories. In three studies (Gollan et al., 2002; Ljungberg et al., 2020, Ljungberg et al., 2013), participants were asked to produce words belonging to a specific category that began with a specific letter of the alphabet. Table 2 shows all the used stimuli.

Table 2. List of stimuli used in the included studies.

Letter verbal fluency task	Semantic verbal fluency task	Modified verbal fluency task
A [3-5, 8-11, 13-16, 18, 19, 21, 22, 24, 28-30]	Actions [22]	Occupations with letter B [13, 14]
B [3]	Adjectives[27]	Proper names with letter L [9]
C [3, 9, 31]	Airplane trip medications [27]	Proper names with letter M [9]
D [3, 9, 26]	Animals [4-6, 9-12, 17, 20-25, 29-35]	
E [6, 9]	Clothes [1,2 6, 9, 15, 21, 23, 26,27]	
F [3-5, 8-11, 15, 18, 19, 21, 22, 24, 25, 28-32, 35]	College majors [9]	
L [9, 31]	Colors [6, 9]	
M [3, 6, 9, 32]	Countries [6, 9]	
O [26]	Countries in Europe [9]	
P [3, 6, 9, 17, 32]	Degree courses [32]	
R [3, 9]	Food [32]	
S [1, 3-5, 8, 9, 11, 15, 18, 19, 21, 22, 24, 25, 28-30]	Frequent travel jobs [27]	
T [3]	Fruits [6, 9, 23, 26, 29]	
V [26]	Fruits and Vegetables [10, 21]	
Double-letter [27]	Function words[27]	
D-KEFS stimuli (n.s.) [1,7]	Furniture [26]	
	Girls' names [15]	
	Kitchen [22]	
	Musical instruments [9]	
	Nouns [27]	
	Occupations [9, 27, 33, 34]	
	Physical labor jobs [27]	
	Produce items [27]	
	Spices [27]	
	Sports [9]	
	Suitcase items [27]	
	Supermarket items [27]	
	Things that cost < $1 [27]	
	Things that cost between $1-$500 [27]	
	Things that cost more than $1000 [27]	
	Tools [2]	
	Vegetables [2, 6, 9, 23, 33, 34]	
	Wheels [9]	
	D-KEFS stimuli (n.s.) [1,7]	

[1]Anderson et al., 2017; [2]Ansaldo et al., 2015; [3]Bennet & Verney, 2019; [4]Bialystok et al., 2008; [5]Bialystok et al., 2009; [6] Bogulski et al., 2015; [7]Blumenfeld et al., 2016; [8]Friesen et al., 2015; [9]Gollan et al., 2002; [10]Keijzer & Schmid, 2016; [11]Kousaie et al., 2014; [12]Kowoll et al., 2015; [13]Ljungberg et al., 2020; [14]Ljungberg et al., 2013; [15]Luo et al., 2010; [16]Marsh et al., 2019; [17]Massa et al., 2020; [18]Morrison et al., 2019; [19]Morrison et al., 2020; [20]Paplikar et al., 2021; [21]Patra et al., 2020; [22]Portocarrero et al., 2007; [23]Prior & Gollan, 2011; [24]Rosselli et al., 2000; [25]Rosselli et al., 2002; [26]Sadat et al., 2016; [27]Sandoval et al., 2010; [28]Secer, 2016; [29]Soltani et al., 2021; [30]Taler et al., 2013; [31]Tao et al., 2015; [32]Vega-Mendoza et al., 2015; [33]Woumans et al., 2015; [34]Woumans et al., 2019; [35]Zeng et al., 2019. N.s.: not specified.

Letter fluency task (n = 29)

The results are summarized in Table 3.

Table 3. Results of letter fluency tasks.

	Letter fluency task							
	Quantitative analysis		*Qualitative analysis*		*Time-course analysis*			
	N° of correct words	N° of errors	Cluster size	N° of switches	1st-RT	Sub-RT	Initiation	Slope
Anderson et al., 2017	M = B	M = B (set-loss)[a] M > B (repetition)[a]						
Bennett & Verney, 2019	M = B[b] SDB < EDB = BB M = EB EB > LB[c]							
Bialystok et al., 2009	M > B							
Bialystok et al., 2008	Study 1 M > B[d] M > LB HB = LB, M Study 2 M = B HB > LB, M							
Blumenfeld et al., 2016	M = B							
Bogulski et al., 2015	ENG M = B FR FullB > LapB > M							
Friesen et al., 2015	M < B					M < B		M > B
Gollan et al., 2002	M > B[e]	M = B						
Keijzer & Schmid, 2016	DM = B EM = B							
Kousaie et al., 2014	EM > FM, B FM = B							
Ljungberg et al., 2020	M < SWEB M = SFB SWEB = SFB							

Ljungberg et al., 2013	M < B							
Luo et al., 2010	HB > M = LB				M = B	M < HB = LB		
Marsh et al., 2019	M = B		M < B	M < B				
Massa et al., 2020	M < B			M = B				
Morrison et al., 2019	M = B[f]							
Morrison et al., 2020	M = B							
Patra et al., 2020	M < B		M < B	M = B	M = B	M < B	M = B	M > B
Portocarrero et al., 2007	M = B	M = B						
Rosselli et al., 2000	SPA, ENG M = B	M = B[a]						
Rosselli et al., 2002	M = B							
Sadat et al., 2016 (Exp 1)	M = B							
Sandoval et al., 2010[g]	M > B	M = B			M = B	M < B		
Seçer, 2016	M = B							
Soltani et al., 2021	M = B							
Taler et al., 2013	M = B							
Tao et al., 2015	MEB > M > SEB							
Vega-Mendoza et al., 2015	M = B							
Zeng et al., 2019	M = B							

1st RT: First response time; Sub-RT: Subsequent response time; M: monolinguals; B: bilinguals; set-loss: words that violate any of the criterion rules of the condition; EDB: English dominant bilinguals; BB: balanced bilinguals; SDB: Spanish dominant bilinguals; EB: early bilinguals; LB: late bilinguals; HB: high proficiency bilinguals; LB: low proficiency bilinguals; ENG: English; FR: French; LapB: lapsed bilinguals; FullB: full bilinguals; DM: Dutch monolinguals; EM: English monolinguals; FM: French monolinguals; SFB: Swedish-Finnish bilinguals; SWEB: Swedish-English bilinguals; SWE: Swedish; SPA: Spanish; SEB: Spanish-English bilinguals; MEB: Mandarin-English bilinguals.

[a]The score was calculated considering errors in both letter fluency and semantic fluency. [b]All condition (F, A, S, B, C, P, M, R, D, FAS, BCT, PMR, CDM) except one (T; M > B). [c]All condition (F, S, B, C, T, P, M, R, D, FAS, BCT, PMR, CDM) except one (A; M = B). [d]ANCOVA (PPVT) M < B. [e]M > B (5 conditions out of 10: A, E, L, P, S); M = B (5 conditions out of 10: M, D, F, R, C). [f]F, A: M = B; S: M > B. Bilingual: English > French. [g]Double-letter condition.

Quantitative analysis

Twenty-nine studies conducted a quantitative analysis of performance.

In sixteen studies (Anderson et al., 2017; Bennett & Verney, 2019; Blumenfeld et al., 2016; Keijzer & Schmid, 2016; Marsh et al., 2019; Morrison et al., 2019; Morrison et al., 2020; Portocarrero et al., 2007; Rosselli et al., 2000; Rosselli et al., 2002; Sadat et al., 2016; Seçer, 2016; Soltani et al., 2021; Taler et al., 2013; Vega-Mendoza et al., 2015; Zeng et al., 2019) no significant differences emerged between monolingual and bilingual participants.

In Bennett and Verney (2019), participants were asked to produce words using ten different letters of the alphabet, and bilinguals generated fewer items in only one condition (words starting with letter "T") while no significant differences emerged in the other conditions. Furthermore, subdividing the bilingual participants by language dominance revealed differences between the groups. Specifically, the dominant Spanish bilingual group produced fewer words than dominant English bilingual and balanced bilingual groups.

In Bogulski et al.'s (2015) study, bilingual participants completed English and French tasks. No significant differences emerged between bilinguals and English monolinguals in the English version, while in the French version, bilinguals produced more words than French monolinguals.
In three studies (Bialystok et al., 2009; Gollan et al., 2002; Sandoval et al., 2010), bilingual participants produced fewer words than monolingual.
Bialystok et al. (2008) included in the study a group of monolinguals and two groups of bilinguals with a different language proficiency (LB, low proficient bilinguals; HB, high proficient bilinguals). In the first experiment, LB produced fewer items than monolinguals, while no differences emerged compared to HB. In addition, no significant differences emerged between the HB and monolingual groups. In the second experiment, HB produced more words than low proficient bilingual and monolingual.
In Gollan et al.'s study (2002), participants were asked to complete the task in ten different conditions. Monolinguals produced more words than bilinguals in five conditions, while there were no significant differences between the groups in the remaining conditions.
In Kousaie et al.'s study (2014), English - French bilinguals produced fewer words than English monolinguals but not compared to French monolinguals.
In four studies (Friesen et al., 2015; Ljungberg et al., 2013; Massa et al., 2020; Patra et al., 2020), bilinguals produced more words than monolinguals.
Luo et al. (2010) found that HB produced more items than monolinguals, while no significant difference emerged between monolinguals and LB. Ljungberg et al. (2020) observed that only one bilingual group (Swedish-Finnish bilinguals) produced more items than the Swedish monolinguals. No differences emerged between the two bilingual groups (Swedish-Finnish vs Swedish-English bilinguals).
In Tao et al.'s study (2015), monolinguals produced more words than the SEB group, while MEB produced more items than monolinguals and SEB.
Four studies (Gollan et al., 2002; Portocarrero et al., 2007; Rosselli et al., 2000; Sandoval et al., 2010) compared the number of errors and found no differences between the groups.

Qualitative analysis

Three studies conducted a qualitative analysis.
Two studies (Marsh et al., 2019; Patra et al., 2020) divided into clusters the words produced by the participants, following the method used by Troyer et al. (1997), and found that bilinguals had a larger cluster size than monolinguals. Three studies assessed the number of switches between clusters. Two of these studies (Massa et al., 2020; Patra et al., 2020) did not observe differences between monolinguals and bilinguals, while Marsh et al. (2019) found that bilinguals made more switches than monolinguals.
Three studies (Blumenfeld et al., 2016; Sadat et al., 2016; Sandoval et al., 2010) found that bilinguals generated more cognates words than monolinguals.

Time-course analysis

Four studies conducted a time-course analysis. Three studies (Luo et al., 2010; Patra et al., 2020; Sandoval et al., 2010) considered first reaction time (RT), and no differences emerged between-groups. Furthermore, Patra et al. (2020) evaluated the initiation parameter, and no significant differences emerged. All these studies (Friesen et al., 2015; Luo et al., 2010; Patra et al., 2020; Sandoval et al., 2010) evaluated subsequent RTs and evidenced a faster subsequent RT for monolinguals. Two studies (Friesen et al., 2015; Patra et al., 2020) assessed slope and showed a higher slope for the monolingual group.

Category fluency task (n = 30)

The results are summarized in Table 4.

Table 4. Results of semantic fluency tasks.

	Semantic fluency task							
	Quantitative analysis		*Qualitative analysis*		*Time-course analysis*			
	N° of correct words	N° of errors	Cluster size	N° of switches	1st-RT	Sub-RT	Initiation	Slope
Anderson et al., 2017	M > B							
Ansaldo et al., 2015	M = B							
Bialystok et al., 2009	M > B							
Bialystok et al., 2008	Study 1 M = B[a] M = HB = LB Study 2 M = B M = HB > LB							
Blumenfeld et al., 2016	M = B							
Bogulski et al., 2015	ENG M = B FR FullB > LapB > M							
Friesen et al., 2015	M = B					M < B		M = B
Gollan et al., 2002	M > B[b]	M = B						
Keijzer & Schmid, 2016	DM > B EM = B							
Kousaie et al., 2014	EM, FM = B EM > FM							
Kowoll et al., 2015	M = B							
Ljungberg et al., 2020	M = B							
Ljungberg et al., 2013	M = B							
Luo et al., 2010	M = B				M = B	M = B		
Massa et al., 2020	M = B			YM > YB OM = OB				
Paplikar et al., 2021	M = B							
Patra et al., 2020	M = B		M = B	M = B	M = B	M < B	M = B	M = B
Portocarrero et al., 2007	M > B	M = B						
Prior & Gollan, 2011	MEB = SEB							
Rosselli et al., 2000	SPA[c], ENG M > B	M = B[d]						
Rosselli et al., 2002	EM > B SM = B		M = B[e]					
Sadat et al., 2016 (Exp 1)	M = B							
Sandoval et al., 2010	M > B	M = B			M < B	M < B		
Soltani et al., 2021	M = B[f]							
Taler et al., 2013	M = B							
Tao et al., 2015	M = MEB > SEB							
Vega-Mendoza et al., 2015	M = B							
Woumans et al., 2015	L1 M = B L2, dual-language unbB < bB							

Woumans et al., 2019	M = B
Zeng et al., 2019	M = B

1st RT: First response time; Sub-RT: Subsequent response time; M: monolinguals; B: bilinguals; EDB: English dominant bilinguals; BB: balanced bilinguals; SDB: Spanish dominant bilinguals; EB: early bilinguals; LB: late bilinguals; HB: high proficiency bilinguals; LB: low proficiency bilinguals; ENG: English; FR: French; LapB: lapsed bilinguals; FullB: full bilinguals; DM: Dutch monolinguals; EM: English monolinguals; FM: French monolinguals; SFB: Swedish-Finnish bilinguals; SWEB: Swedish-English bilinguals; SWE: Swedish; SPA: Spanish; SEB: Spanish-English bilinguals; MEB: Mandarin-English bilinguals.
[a]ANCOVA (PPVT) M = B. [b]M > B (9 conditions out of 12: Countries, Countries in Europe, wheels, musical instruments, vegetables, sports, fruits, colors, animals); M = B (3 conditions out of 12: occupation, college majors, clothing). [c]M > B (fruits); M = B (animals). [d]The score was calculated considering errors in both letter fluency and semantic fluency. [e]Semantic associations (n° di cluster). [f]M = B (fruits); M > B (animals)

Quantitative analysis

Thirty studies conducted a quantitative analysis of performance.
In twenty-one studies (Ansaldo et al., 2015; Bialystok et al., 2008; Blumenfeld et al., 2016; Bogulski et al., 2015; Friesen et al., 2015; Kousaie et al., 2014; Kowoll et al., 2015; Ljungberg et al., 2013; Ljungberg et al., 2020; Luo et al., 2010; Massa et al., 2020; Paplikar et al., 2021; Patra et al., 2020; Prior & Gollan, 2011; Sadat et al. 2016; Soltani et al., 2021; Taler et al., 2013; Vega-Mendoza et al., 2015; Woumans et al., 2015; Woumans et al., 2019; Zeng et al., 2019) no significant difference emerged between monolingual and bilingual participants. Soltani et al. (2021) found that monolinguals named more words in the animal category, while no differences emerged in the fruit category or overall score.
Six studies (Anderson et al., 2017; Bialystok et al., 2008a; Gollan et al., 2002; Portocarrero et al., 2007; Rosselli et al., 2000; Sandoval et al., 2010) observed that bilingual participants produced fewer words than monolinguals.
Gollan et al. (2002) required participants to produce words belonging to twelve categories. Bilinguals generated fewer words in nine categories, while there were no differences between the two groups in the remaining categories. Keijzer and Schmid (2016) found that Dutch monolinguals named more words than bilinguals, while no significant differences emerged between the English monolinguals and bilinguals. Rosselli et al. (2000) did not observe differences between monolinguals and bilinguals in the animal category using Spanish; however, monolinguals generally produced more words than bilinguals. In another study (Rosselli et al., 2002), the authors found that English monolinguals produced more words than bilinguals while no significant difference emerged on the Spanish test. Tao et al. (2015) did not observe differences between monolinguals and Mandarin–English bilinguals, while Spanish–English bilinguals produced fewer words than monolinguals and Mandarin–English bilinguals. Four studies (Gollan et al., 2002; Portocarrero et al., 2007; Rosselli et al., 2000; Sandoval et al., 2010), considering the number of errors, did not find differences between monolinguals and bilinguals.

Qualitative analysis

Three studies conducted a qualitative analysis.
Patra et al. (2020) evidenced no difference in cluster size and number of switches between monolinguals and bilinguals. Rosselli et al. (2002) evaluated the number of semantic associations (i.e., number of clusters consisting of two or more consecutive words) and did not observe between-group differences. By contrast, Massa et al. (2020) found that young monolinguals produced more switches than young bilinguals, while there were no significant differences in the old participant's groups.
Three studies (Blumenfeld et al., 2016; Sadat et al., 2016; Sandoval et al., 2010) observed that bilinguals generated more cognates words than monolinguals.

Time course analysis
Four studies conducted a time-course analysis. Three of them (Friesen et al., 2015; Patra et al., 2020; Sandoval et al., 2010) reported that monolinguals showed a lower subsequent RT than bilinguals, while in the Luo et al.' study (2010), no significant differences emerged.
In two studies (Luo et al., 2010; Patra et al., 2020), there were no significant differences in the first RT, while in the Sandoval et al.'s study (2010), monolinguals showed lower first RT than bilinguals. Two studies (Friesen et al., 2015, Patra et al., 2020) did not observe differences in the slope.

Discussion

The current systematic review summarizes the results of 35 studies published between 2000 and 2021 that investigated the effect of bilingualism on verbal fluency tasks.
In more than half of the studies, the quantitative analysis of performance did not show significant differences between monolinguals and bilinguals, both for letter fluency and semantic fluency tasks.
No study showed evidence of a bilingual advantage, while a few studies showed some difficulties of bilinguals in semantic verbal fluency. In most studies, participants were assessed in the "animal" category, while three studies (Gollan et al., 2002; Keijzer & Schmid, 2016; Sandoval et al., 2010) assessed performance in multiple categories. The selection of categories seems to influence the performance of individuals, and not all categories would seem to be cross-linguistic (i.e., with minimal variation among people living in different countries or speaking different languages; Ardila, 2020). The category of animals seems to assume the same meaning (with minimal variation) across different languages and cultures (Ardila, 2020).
A possible explanation for the higher difficulty experienced by bilinguals in performing this task would be the linguistic interference caused by knowing more than one language. Bilinguals recall the words belonging to the category in the two known languages but must inhibit the words of the untested language. This interference effect seems to be greater when concrete words are recalled (as in semantic fluency) rather than abstract words. Rosselli et al. (2002) hypothesized that concrete words share more elements in their representations across languages, and this feature could generate greater cross-language interference. Linguistic interference is not only attention-demanding but also time-consuming. The results of time-course analyses seem to confirm this hypothesis. In fact, in three studies (Friesen et al., 2015; Patra et al., 2020, Sandoval et al., 2010), bilinguals produced the same number of words but had greater sub-RTs (i.e., the average value of the time intervals from the onset of first response to the onset of each subsequent response) than monolinguals. Longer sub-RTs in conjunction with equivalent or fewer items generated are consistent with the notion of word retrieval difficulties (Friesen et al., 2015).
Further factors must be considered, such as the age of the participants and the level of education. Several studies have shown that advancing age is associated with reduced verbal fluency (Soltani et al., 2021). However, a higher level of education seems to improve performance in this task. Considering the studies that reported a negative effect of bilingualism on semantic fluency performance, five studies (Anderson et al., 2017, Keijzer & Schmid, 2016, Rosselli et al., 2000; Rosselli et al., 2002, Soltani et al., 2021) included older adult participants (age > 60 years) with a high level of education. Standard neuropsychological tools (e.g., Mini-Mental State Examination, Folstein et al., 1975) would not appear to be sensitive enough to identify small changes in cognitive status in adult individuals with high cognitive reserve (Anderson et al., 2017). Bilingualism would slow the onset of dementia symptoms (for a review, see Brini et al., 2020; Van den Noort et al., 2019). This positive effect could delay diagnosing neurodegenerative diseases during neuropsychological evaluations since bilinguals with the same monolinguals' degree of brain atrophy could show better cognitive performance (Tao et al., 2021). It is possible to speculate that bilingual adults may be at a more advanced stage of cognitive decline than monolinguals and that this aspect affects participants' performance included in these studies. Mixed results emerged from the studies that assessed performance in letter verbal fluency. Quantitative analysis showed improved performance for bilinguals in eight studies,

while six studies showed a disadvantage for bilinguals. The "FAS" triplet was used for test performance in most studies. The choice of stimulus seemed to affect participants' performance on this test.

Bennet and Verney (2019) highlighted, for example, that Spanish bilinguals could have greater difficulty in producing words beginning with the letter "A" or "S" due to the existence of overlapping sounds in the Spanish language. The study results by Gollan et al. (2002), which evaluated participants' performance in ten conditions showing a significant difference in only half of the conditions, seem consistent with that hypothesis. Furthermore, it has been hypothesized that the presence of a greater number of cognates (words that have a similar sound in two languages, e.g., flower and floor) could improve individuals' performance in performing letter fluency (Michael & Gollan, 2005). Specifically, the probability of producing cognates appeared higher in letter fluency than in semantic fluency (Blumenfeld et al., 2016; Sadat et al., 2016). Cognates seem to facilitate lexical access (Sandoval et al., 2010) and to be more frequent when known languages are more phonological similar (Sadat et al., 2016).

Linguistic history is a further aspect to consider. Bennet and Verney (2019) found no differences between the bilingual and monolingual groups when they did not consider the linguistic history of the participants. Instead, when establishing the dominant language of the participants, the study showed that Spanish dominant bilinguals performed worse than English dominant bilinguals, balanced bilinguals, and monolinguals. In Kousaie et al. (2014), bilinguals underperformed English dominant bilinguals but not French monolinguals, and in Bialystok et al.'s study (2008), monolinguals produced more words only compared to low proficient bilinguals.

Using a modified version of the task, more complex than the classic task, that required participants to produce words from a pair of letters (double-letter condition), bilinguals showed greater sub-RT in letter fluency tasks (Friesen et al., 2015; Luo et al., 2010; Patra et al., 2020; Sandoval et al., 2010). However, this outcome was associated with the production of a greater number of words than the monolingual group. This result seems to indicate superior control and continued generating words over time. Specifically, in studies performing time-course analysis, bilinguals showed a flatter slope (Friesen et al., 2015; Patra et al., 2020). This would seem to indicate the ability to maintain performance over time despite increased interference toward the end of the task (an indicator of better executive functioning).

Studies that conducted a qualitative analysis of responses produced in letter fluency showed a greater cluster size in bilinguals than monolinguals (Marsh et al., 2019; Patra et al., 2020).

Conclusions

No definitive conclusions can be drawn about a bilingual advantage or disadvantage on the verbal fluency tasks. Future studies should conduct quantitative analyses to interpret better the results showed by this review. Comparing the effect size of the various studies could help to understand the results better. The results of this systematic review indicate the need to take some precautions in using this type of task, especially in the context of clinical practice. Standardized instruments for a monolingual population and associated normative data seem not to be appropriate to test a bilingual population. To date, research is attempting to establish new guidelines for neuropsychological assessments of the bilingual population. For example, Paplikar et al. (2021) provided recommendations suggesting that assessments of bilinguals should be conducted in their dominant language and that responses in their second known language should also be considered in the scoring. Study results show a change in performance depending on the stimuli used and the linguistic history of the participants. Both factors should be considered when conducting a neuropsychological assessment. Given the significant increase in the bilingual population, some questions related to language history should be included by default in every neuropsychological evaluation. This assessment can help to interpret the results more accurately. However, creating normative data for the bilingual population is arduous. Bilingual people can have very different language histories, and the number of possible language pairs is very numerous. Furthermore, trilingualism and multilingualism are becoming more prevalent, which could also affect individuals' performance.

Chapter 4: Assessing executive functions using verbal fluency in bilinguals

Introduction

Bilingualism is a socio-demographic factor that seems to determine an advantage in some cognitive functions (e.g., executive functions, for a review see Grundy, 2020) and a disadvantage in linguistic abilities (e.g., picture naming, Gollan et al., 2005; verbal fluency, Anderson et al., 2017; Gollan et al., 2002). Different explanations have been proposed to explain the poor quality of linguistic performance in bilingual participants. According to the Inhibitory control model (Green, 1998), both known languages are always active in the brain, and bilinguals must inhibit the non-target language for successful communication. This process is time-consuming and attentional demanding and could explain why bilinguals produce fewer words (Ardila, 2020). Conversely, according to the "weaker link" model (Michael & Gollan, 2005), reduced language production results from bilinguals making less use of known languages. Moreover, bilinguals are likely to develop context-specific vocabulary depending on their use of each language (Rhys & Thomas, 2013). Verbal fluency, a neuropsychological tool widely used in clinical practice, is a useful instrument for simultaneously assessing language skills and executive functions. As highlighted in the systematic review in Chapter 3, most studies that used verbal fluency to evaluate performance in the bilingual population considered only quantitative indices (e.g., number of correct responses). Conversely, measuring performance with qualitative indices (e.g., the size of the clusters produced) or indices of language production over time (e.g., subsequent mean RT) allows analyzing not only language skills but also executive functioning (see Patra et al., 2020). In verbal fluency, words are produced in temporal clusters. To maximize language production, participants produce words belonging to a certain sub-category (i.e., cluster) and, once this is exhausted, move on to the next sub-category (i.e., switching). Successful performance requires access to the information contained in the memory and strong executive abilities responsible for search strategies, response initiation, shifting, and flexibility.

Previous studies (Luo et al., 2010; Sandoval et al., 2010) compared different experimental conditions and showed that the letter condition demands executive control mechanisms more than semantic condition. In semantic fluency, greater emphasis is placed on linguistic abilities (Luo et al., 2010; Patra et al., 2020). Furthermore, the number of possible items is greater in the letter condition, but participants generally produce more correct words in the semantic condition.

The participants' performance seems to depend on different factors. For example, the categories and letters used for the assessment would appear to play a role in determining the participants' performance. According to Ardila's study (2020), language and word length do not seem to influence performance while age and level of education do.

This study aims to analyze a group of bilinguals' verbal fluency task performance using quantitative, qualitative, and language production indices over time.

It is assumed that the bilinguals will perform poorly on the verbal fluency task considering the indicators of the linguistic component (the onset of first response, the initiation parameter, and the cluster size). Conversely, the bilingual group will show a better performance regarding the indicators related to the executive control component (the Fluency Difference Score, the subsequent-RT, the slope, and the number of switches). No difference is expected between the two groups for the number of correct responses since it is considered an indicator of the linguistic and executive control components. Considering age groups, the probability of detecting a difference is greater in the group of older bilinguals. Furthermore, the performance of older bilinguals will be poorer than that of younger bilinguals due to physiological cognitive decline.

Method

Participants

Seventy-two people took part in the experiment. Participants were divided into two groups: the control group and bilinguals based on a preliminary assessment (see Language history). Each group was divided into two groups of different ages: 1) the control group: young adults (mean age = 24.67, SD = 2.90), older adults (mean age = 58.33, SD = 3.52); 2) Bilingual group: young adults (mean age = 26.83, SD = 3.29), older adults (mean age = 60.17, SD = 4.57). Participants' characteristics and results will be separately reported for young and older adults (see below).

Demographic information

The same as experiment 1 (see Chapter 2).

Language history

The same as experiment 1 (see Chapter 2).

Cognitive tests

Raven's standard progressive matrices (Raven et al., 1998). Participants were asked to indicate the missing piece of a figure. The test consists of sixty items, and it assesses general human intelligence and abstract reasoning. All the items are non-verbal, and for this reason, it is considered a "culture fair" test.

Verbal fluency – Letter condition (Caltagirone et al. 1995). Participants were asked to produce as many words as possible that start with letters F, A, S in sixty seconds. The restrictions for the letter conditions were to produce unique Italian words that are not proper names (no name of people and no name of places, e.g., Florence). Each participant was tested individually. After providing the instruction, the participant started a trial only when the tester said "start" to ensure a definitive starting point for each trial. Responses were recorded with a digital voice recorder.

Verbal fluency – Semantic condition (Novelli et al., 1986; Zarino 2014). Participants were asked to produce as many words as possible in two categories (animals and fruits) in sixty seconds. Each participant was tested individually. After providing the instruction, the participant started a trial only when the tester said "start" to ensure a definitive starting point for each trial. Responses were recorded with a digital voice recorder.

Procedure

All participants completed the questionnaires and cognitive tests online. The questionnaires were administered online using the KoboToolBox platform (www.kobotoolbox.org), while the cognitive tests were administered during an online interview. All participants completed the Language and Social Background Questionnaire (adapted from Anderson et al., 2018), Raven's standard progressive matrices, the letter and semantic fluency tasks. The bilingual participants completed the Language Switching Questionnaire (LSQ, adapted from Rodriguez-Fornells et al., 2012) and the Language Dominance Questionnaire (LDQ, adapted from Dunn & Tree, 2009). The questionnaires and the cognitive tests were presented in a fixed order. All testing was conducted in Italian.

Statistical analysis

For the verbal fluency tasks, each correct response was time-stamped using PRAAT (Boersma & Weenink, 2015) to index the onset of a response from the beginning of the trial (i.e., "start") and to calculate the variables in the time-course analysis.

The following variables were measured for each trial:

- *Number of correct responses (CR)*: the number of responses produced in sixty seconds, excluding errors and repetitions.

In the semantic condition, the CR was the total number of responses excluding errors (i.e., words that did not belong to the designated categories and cross-linguistic intrusions) and repetitions. Subcategories (e.g., fish) were not counted if a specific item of the subcategories was also provided (e.g., tuna).
In letter condition, the CR was the total number of responses excluding errors (i.e., words that did not start with the specified letter, proper names, and cross-linguistic intrusions) and repetitions.
- *Fluency Difference Score (FDS)*: (CR semantic fluency – CR letter fluency) /CR semantic fluency
- *Time-course analysis*: First-RT (1st-RT), Subsequent-RT (Sub-RT), initiation parameter, and slope were calculated based on the timing of the responses. CR was grouped into 5 seconds bins over each 60 seconds trial, resulting in 12 bins. The group means of CR in each of the twelve bins were calculated for each letter and semantic fluency trial. The means of CR for each trial were plotted using a line graph (x variable, bin number; y-variable, mean CR). This graph was fitted with a logarithmic function and two measures derived from this plot: initiation parameter and slope.
1st-RT is the time interval from the beginning of the trial to the onset of the first response. Sub-RT is the mean value of the time intervals from the onset of the first response to the onset of each subsequent response. The initiation parameter is the starting point of the logarithmic function that is the value of y when $t = 1$ or $In(t) = 0$. Slope refers to the rate of the retrieval output as a function of the change in time over 60 seconds.
- *Clustering and switching analyses*. Repetitions were included for the clustering and the switching analyses. In letter condition, the method developed by Troyers et al.'s (1997) was used to cluster the responses. Words were included in the same cluster if they met one of the following criteria: words that begin with the same first two letters (family and fall); words that differ only by a vowel sound regardless of the actual spelling (fun and fan); words that rhyme (floor and for); or words that are homonyms (foot: anatomical part of the body, and foot: unit of measure).
In semantic conditions, a new method was developed. For the "animal" condition (see Table 1), Troyers et al.'s method was modified, adding new categories according to previous studies (e.g., Kosmidis et al., 2004; Jaimes-Bautista et al., 2020). For the "fruit" condition (see Table 2), a new method based on the strategies used in previous studies (e.g., Jaimes-Bautista et al., 2020; Kavé et al., 2008; Kosmidis et al., 2004; Patra et al., 2020; Soltani et al., 2021) was developed.
Two variables were generated after clustering the responses. Cluster size was calculated beginning with the second word in each cluster. A single word was given a cluster size of zero (e.g., dog), two words cluster obtained a cluster size of one (e.g., dog, cat belong to Pets animal cluster and cluster size of one), and so on. The mean cluster size for a trial was calculated by summing the size of each cluster and dividing the score by the number of clusters. The number of switches was the number of transitions between clusters. For example, dog, cat; penguin, parrot; camel, horse, llama contain two switches – before penguin and before camel. Similarly, in letter fluency – fortune, forgive; fall, fast; flower, fly, flute include two switches – before fall and before flower.
Analyses of variance (ANOVAs) were used to compare the performance of participants with Language Group as the independent variable (Bilinguals, Control Group). ANOVAs were conducted to compare test performance in different conditions (semantic vs letter; animal vs fruit; F vs A vs S).

Table 1. Cluster and exemplars for animals.

Animals	
Living Environment	
Africa	aardvark, antelope, buffalo, camel, chameleon, cheetah, chimpanzee, cobra, dromedary, eland, elephant, gazelle, giraffe, gnu, gorilla, hippopotamus, hyena, impala, jackal, lemur, leopard, lion, manatee, meerkat, mongoose, monkey, ostrich, panther, rhinoceros, suricate, tiger,

	wildebeest, warthog, zebra
Artic/Far North	auk, caribou, musk ox, penguin, polar bear, reindeer, seal
Australia	emu, kangaroo, kiwi, opossum, platypus, Tasmanian devil, wallaby, wombat
Farm	chicken, cow, donkey, duck, ferret, goat, goose, horse, lamb, mule, pig, sheep, turkey
North America	badger, bear, beaver, bobcat, caribou, chipmunk, cougar, deer, elk, fox, moose, mountain lion, puma, rabbit, raccoon, skunk, squirrel, wolf
Water	alligator, auk, beaver, crocodile, clam, dolphin, fish, frog, lobster, manatee, muskrat, mussel, newt, octopus, otter, oyster, penguin, platypus, salamander, sea lion, seal, shark, ray, toad, turtle, whale
Human use	
Beasts of burden	camel, donkey, dromedary, horse, llama, ox
Fur	beaver, chinchilla, fox, mink, rabbit
Pets	budgie, canary, cat, dog, gerbil, golden retriever, guinea pig, hamster, parrot, rabbit
Zoological Categories	
Arachnids	scorpion, spider, tick
Bird	albatross, budgie, condor, duck, eagle, finch, goose, kiwi, macaw, parrot, parakeet, pelican, penguin, robin, toucan, woodpecker
Bovine	bison, buffalo, cow, musk ox, yak
Canine	coyote, dog, fox, hyena, jackal, wolf
Deer	antelope, caribou, eland, elk, gazelle, gnu, impala, moose, reindeer, wildebeest
Feline	bobcat, cat, cheetah, cougar, jaguar, leopard, lion, lynx, mountain lion, ocelot, panther, puma, tiger
Fish	bass, guppy, ray, salmon, trout
Insect	ant, bee, beetle, butterfly, cockroach, dragonfly, flea, fly, ladybug, mosquito, praying mantis, wasp
Insectivores	aardvark, anteater, bat, hedgehog, mole, shrew
Molluscs	snail, clam, mussel, oyster
Primate	ape, baboon, chimpanzee, gibbon, gorilla, human, lemur, marmoset, monkey, orangutan, shrew
Rabbit	coney, hare, pika, rabbit
Reptile/Amphibian	alligator, anaconda, chameleon, crocodile, frog, gecko, iguana, lizard, newt, salamander, snake, toad, tortoise, turtle
Rodent	beaver, chinchilla, chipmunk, gerbil, gopher, groundhog, guinea pig, hamster, hedgehog, marmot, mole, mouse, muskrat, porcupine, rat, squirrel, woodchuck
Weasel	badger, ferret, marten, meerkat, mink, mongoose, otter, polecat, skunk, stoat, suricate

Table 2. Clusters and exemplars for fruit.

Fruits	
Seasons	
Autumn fruits	apple, chestnut, citron, clementine, grapefruit, grapes, kiwi, lemon, orange, persimmon, pomegranate, prickly pear, quince, raspberry, tangerine
Spring fruits	apple, cherry, grapefruit, kiwi, lemon, medlar, melon,

	mulberry, orange, pear, raspberry, strawberry, tomato
Summer fruits	apple, apricot, blackberry, blueberry, cherry, fig, grape, melon, mulberry, peach, pear, plum, prickly pear, prune, raspberry, sour cherry, strawberry, tomato, watermelon
Winter fruits	apple, citron, clementine, grapefruit, kiwi, lemon, orange, tangerine, pear
Type of fruits	
Berries	blackberry, blueberry, currant, mulberry, raspberry, strawberry
Citrus	citron, chinotto, clementine, grapefruit, lemon, lime, orange, mandarin, tangerine
Nuts and seeds	almond, cashew, chestnut, coconut, hazelnut, peanut, pine nut, pistachio, pumpkin seeds, sunflower seeds, walnut
Tropical fruits	alchechengi, avocado, banana, coconut, date, dragon fruit, durian, goji, guava, jackfruit, lime, lucuma, lychee, mango, maracuja, papaya, passion fruit, pineapple, tamarind
Eat as vegetables	cucumber, olive, tomato

Results

Young adults

Participants

48 participants between 18-34 years took part in the study.
24 participants were bilingual speakers of Italian and one additional language (mean age = 26.83, SD = 3.29; 70.83% females). Bilinguals' additional languages were English (15), Spanish (5), French (3), German (1). Most of the participants were born in Italy (N = 22), while two of them in a foreign country (Brazil, N = 1; Moldova, N = 1).
Some bilingual participants reported some experience in L3 (English, N = 7; Dutch = 3; Spanish, N = 2; Portuguese, N = 2; French, N = 1; German, N = 1; Romanian, N = 1; Russian, N = 1) and L4 (Spanish, N = 4; English, N = 1). Two participants reported some experience in Italian regional dialects.
The control group included 24 participants (mean age = 24.67, SD = 2.90; 70.83% females). The L1 of all participants was Italian. Some participants of the control group reported some experience in L2 (English, N = 18, L3 (Spanish, N = 2). Their experience in L2 and L3 was constrained to foreign language courses, and they reported marginal daily use of any foreign language. Twelve participants of the control group reported some experience in Italian dialects. All participants were born in Italy. Participants were recruited using social networks and word of mouth.
There were no differences between the language groups in SES and total Raven's score. Bilinguals were older than the control group. See Table 3 for background information about participants and their language knowledge.

Table 3. Background information about participants and their language knowledge.

	Bilinguals	**Control group**	*p*
Young adults	N = 24	N = 24	
Age	26.83 (3.29)	24.67 (2.90)	*
% females	70.83	70.83	
SES	5.58 (2.24)	4.50 (2.36)	n.s.
Total score Raven (60)	50.54 (4.89)	48.10 (5.12)	n.s.
Language experience L1			
Age of exposure (in years)	0.50 (1.72)	0.12 (0.61)	n.s.
Percentage of daily use (%)	58.15 (21.87)	90.18 (5.40)	**
Self-rated language proficiency			

Proficiency – Speaking (1-10)	9.83 (0.56)	9.62 (0.71)	n.s.
Proficiency – Comprehension (1-10)	9.92 (0.28)	9.83 (0.48)	n.s.
Proficiency – Writing (1-10)	9.83 (0.38)	9.54 (0.83)	n.s.
Proficiency – Reading (1-10)	10.00 (0.00)	9.83 (0.48)	n.s.
Language experience L2			
Age of exposure (in years)	10.65 (7.06)	8.83 (5.67)	n.s.
Self-rated language proficiency			
Proficiency – Speaking (1-10)	8.67 (1.37)	5.54 (1.69)	**
Proficiency – Comprehension (1-10)	9.12 (0.99)	6.27 (1.68)	**
Proficiency – Writing (1-10)	8.42 (1.61)	5.27 (1.74)	**
Proficiency – Reading (1-10)	9.33 (0.76)	6.27 (1.35)	**
Older adults	N = 12	N = 12	
Age	58.33 (3.52)	60.17 (4.57)	n.s.
% females	83.33	75.00	
SES	6.42 (2.06)	6.00 (2.66)	n.s.
Total score Raven (60)	46.70 (7.99)	41.75 (6.20)	n.s.
Language experience L1			
Age of exposure (in years)	1.67 (4.25)	0.58 (2.02)	
Percentage of daily use (%)	47,00 (14.19)	96.50 (3.82)	**
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.75 (0.45)	9.33 (0.89)	n.s.
Proficiency – Comprehension (1-10)	10.00 (0.00)	9.58 (0.79)	n.s.
Proficiency – Writing (1-10)	9.75 (0.62)	9.17 (1.19)	n.s.
Proficiency – Reading (1-10)	10.00 (0.00)	9.83 (0.39)	n.s.
Language experience L2			
Age of exposure (in years)	24.25 (15.15)	11.00 (8.05)	n.s.
Self-rated language proficiency			
Proficiency – Speaking (1-10)	8.17 (1.40)	4.00 (2.16)	**
Proficiency – Comprehension (1-10)	8.58 (1.56)	4.25 (0.50)	**
Proficiency – Writing (1-10)	7.75 (1.60)	3.50 (1.73)	**
Proficiency – Reading (1-10)	8.50 (1.51)	4.25 (1.71)	**

* $p < .05$; ** $p < .01$

Measures of bilingualism

There were significant differences on subjective language proficiency ratings and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from Italian to L2 and vice versa. Bilinguals' language profile is summarized in Table 4.

Table 4. Bilinguals' language profile. Means (standard deviation).

	L1	**L2**	Statistical results
Young adults (N = 24)			
Age of exposure (in years)	0.50 (1.72)	10.65 (7.06)	$t(23)$=6.42; $p < .01$
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.83 (0.56)	8.67 (1.37)	$t(23)$= 4.07; $p < .01$
Proficiency – Comprehension (1-10)	9.92 (0.28)	9.12 (0.99)	$t(23)$= 4.16; $p < .01$
Proficiency – Writing (1-10)	9.83 (0.38)	8.42 (1.61)	$t(23)$= 4.17; $p < .01$
Proficiency – Reading (1-10)	10.00 (0.00)	9.33 (0.76)	$t(23)$= 4.29; $p < .01$
Language dominance	19.33 (5.66)	7.12 (4.77)	$t(23)$= 6.38; $p < .01$
Language switching habit	8.42 (2.08)	8.50 (2.23)	$t(23)$= -0.17; p = n.s.
Older adults (N = 12)			
Age of exposure (in years)	1.67 (4.25)	24.25 (15.15)	$t(11)$= -5.08; $p < .01$

Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.75 (0.45)	8.17 (1.40)	$t(11)= 3.64$; $p < .01$
Proficiency – Comprehension (1-10)	10.00 (0.00)	8.58 (1.56)	$t(11)= 3.14$; $p < .01$
Proficiency – Writing (1-10)	9.75 (0.62)	7.75 (1.60)	$t(11)= 3.63$; $p < .01$
Proficiency – Reading (1-10)	10.00 (0.00)	8.50 (1.51)	$t(11)= 3.45$; $p < .01$
Language dominance	14.67 (5.04)	7.17 (5.56)	$t(11)= 2.50$; $p = .01$
Language switching habit	6.75 (1.96)	6.58 (2.19)	$t(11)= 0.38$; p = n.s.

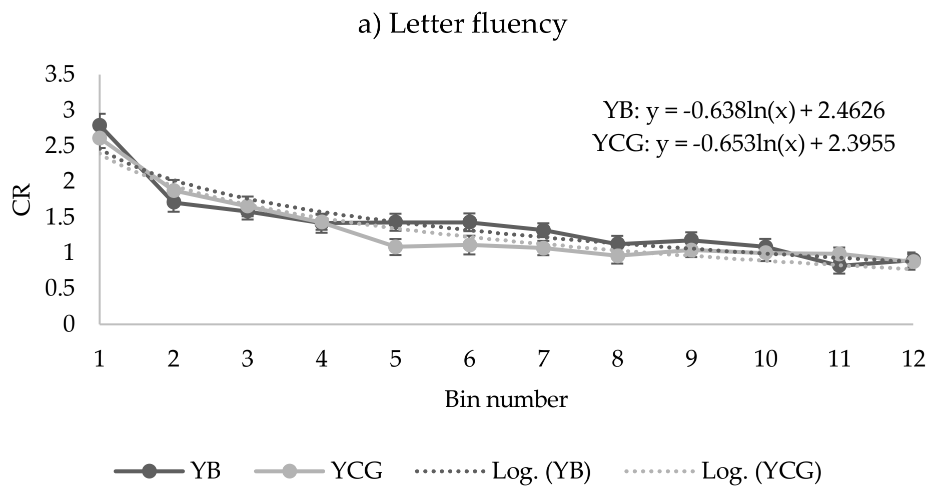
Experimental data results

Table 5 shows the means and standard deviation of dependent variables for each experimental task.

Verbal fluency - Semantic vs Letter condition

The CR showed a main effect of Condition ($F_{1,46} = 23.80$; $p < .01$; $\eta^2 = .34$) and participants produced more CR in semantic (mean = 19.18) than in letter condition (mean = 16.24). The sub-RT revealed a main effect of Condition ($F_{1,46} = 39.82$; $p < .01$; $\eta^2 = .46$) and participants had lower Sub-RT in semantic (mean = 21.57 s) than in letter fluency (mean = 25.17 s). Cluster size revealed a main effect of Condition ($F_{1,46} = 49.52$; $p < .01$; $\eta^2 =$ .52) and it was larger in semantic (mean = 1.05) than in letter condition (mean = 0.59). 1stRT and number of switching did not show any significant differences.

The FDS score did not show significant differences between group ($F < 1$). In both letter and semantic fluency initiation parameter and the slope was not significantly different ($F < 1$). Figure 1 represent the time-course of the CR.



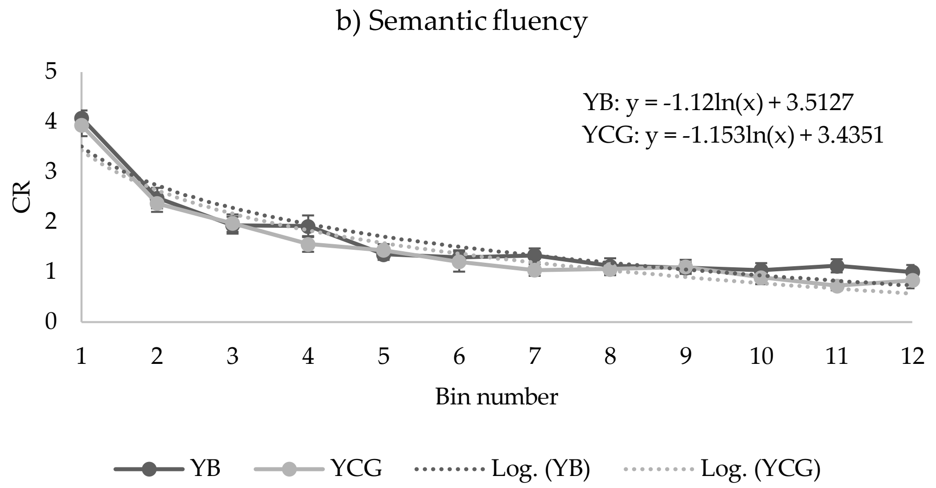


Figure 1. Group comparison (young adults) of CR produced as a function of 5-sec time interval in a) letter and b) semantic fluency. Error bars represent standard errors of the mean.

Semantic verbal fluency - Animal vs Fruit conditions

The CR showed a main effect of Condition ($F_{1,46} = 71.20$; $p < .01$; $\eta^2 = .61$) and participants produced more CR in animal (mean = 22.54) than in fruit condition (mean = 15.81). The Group x Condition interaction was marginally significant ($F_{1,46} = 3.25$; $p = .08$; $\eta^2 = .07$). Both groups produced more CR in animal than in fruit condition (bilinguals: $F_{1,46} = 52.43$; $p < .01$; $\eta^2 = 53$; control group: $F_{1,46} = 22.01$; $p < .01$; $\eta^2 = .32$). Moreover, in animal condition the difference between groups was marginally significant ($F_{1,46} = 2.92$; $p = .09$; $\eta^2 = .06$), bilinguals tended to produce more CR than the control group. The sub-RT revealed a main effect of Condition ($F_{1,46} = 65.95$; $p < .01$; $\eta^2 = .59$) and participants had lower Sub-RT in fruit (mean = 19.07 s) than in animal condition (mean = 24.07 s). The Group x Condition interaction was marginally significant ($F_{1,46} = 2.73$; $p = .10$; $\eta^2 = .06$), bilinguals had longer Sub-RT than the control group in animal condition ($F_{1,46} = 4.13$; $p = .05$; $\eta^2 = .08$). Cluster size revealed a main effect of Condition ($F_{1,46} = 25.04$; $p < .01$; $\eta^2 = .35$) and cluster size was larger in fruit (mean = 1.27) than in animal condition (mean = 0.83). Number of switches revealed a main effect of Condition ($F_{1,46} = 132.76$; $p < .01$; $\eta^2 = .74$). Participants made more switches in animal (mean = 12.23) than in fruit condition (mean = 6.52). 1stRT did not show significant differences between groups.

Letter verbal fluency – F vs A vs S conditions

The CR did not show a main effect of Group ($F < 1$), Condition ($F < 1$) nor an interaction ($p > .05$). The sub-RT revealed a main effect of Condition ($F_{2,92} = 5.35$; $p < .01$; $\eta^2 = .10$) and participants had lower Sub-RT in F condition than A ($F_{1,46} = 5.66$; $p = .02$; $\eta^2 = .11$) and S conditions ($F_{1,46} = 10.60$; $p < .01$; $\eta^2 = .19$). Cluster size revealed a main effect of Condition ($F_{2,92} = 5.85$; $p < .01$; $\eta^2 = .11$) and cluster size was lower in F condition than both A ($F_{1,46} = 9.20$; $p < .01$; $\eta^2 = .17$) and S conditions ($F_{1,46} = 13.86$; $p < .01$; $\eta^2 = .23$). Number of switches revealed a main effect of Condition ($F_{2,92} = 9.39$; $p < .01$; $\eta^2 = .17$). Participants made more switches in F condition than both the A ($F_{1,46} = 12.47$; $p < .01$; $\eta^2 = .21$) and the S conditions ($F_{1,46} = 14.49$; $p < .01$; $\eta^2 = .24$). 1stRT did not show significant differences between groups.

Table 5. Means and standard deviations (in parentheses) of verbal fluency tasks for young adult participants.

	Young Bilinguals	**Young Control Group**
CR		
Semantic	19.75 (3.54)	18.60 (3.75)
Animal	23.83 (5.59)	21.25 (4.86)
Fruit	15.67 (3.82)	15.96 (3.78)
Letter	16.79 (4.34)	15.69 (4.42)

F	16.58 (4.92)	16.71 (5.14)
A	16.71 (4.96)	14.92 (5.18)
S	17.08 (4.97)	15.46 (4.94)
FDS	0.14 (0.19)	0.15 (0.22)
1st-RT		
Semantic	0.94 (0.42)	1.00 (0.62)
Animal	0.92 (0.68)	1.11 (0.92)
Fruit	0.96 (0.76)	0.89 (0.60)
Letter	0.81 (0.53)	1.01 (0.61)
F	0.87 (0.77)	0.79 (0.62)
A	0.78 (0.76)	1.19 (0.98)
S	0.78 (0.65)	1.05 (0.90)
Sub-RT		
Semantic	22.13 (2.68)	21.02 (3.44)
Animal	25.14 (3.24)	23.01 (3.97)
Fruit	19.11 (3.21)	19.03 (4.44)
Letter	25.43 (2.15)	24.91 (2.43)
F	24.91 (3.35)	22.90 (3.96)
A	25.48 (4.00)	25.36 (3.87)
S	25.90 (3.36)	26.46 (3.26)
Initiation - Semantic	3.51 (0.75)	3..43 (0.77)
Slope - Semantic	1.12 (0.35)	1.15 (0.43)
Initiation - Letter	2.46 (0.67)	2.39 (0.57)
Slope - Letter	0.64 (0.25)	1.03 (1.83)
Cluster size		
Semantic	1.07 (0.30)	1.02 (0.32)
Animal	0.90 (0.38)	0.76 (0.32)
Fruit	1.25 (0.50)	1.29 (0.51)
Letter	0.63 (0.42)	0.55 (0.32)
F	0.44 (0.37)	0.42 (0.27)
A	0.69 (0.54)	0.66 (0.61)
S	0.75 (0.60)	0.56 (0.49)
Number of switches		
Semantic	9.56 (2.08)	9.19 (1.77)
Animal	12.67 (3.73)	11.79 (2.62)
Fruit	6.46 (1.82)	6.58 (1.64)
Letter	9.93 (2.20)	9.80 (2.75)
F	11.04 (3.56)	11.29 (3.57)
A	9.37 (2.84)	8.67 (3.62)
S	9.37 (2.52)	9.46 (3.31)

Note. CR: Correct responses; FDS: Fluency Difference Score, Sub-RT: Subsequent reaction times.

Older adults

Participants

24 participants aged over 55 years took part in the study.

12 participants were bilingual speakers of Italian and at least one additional language (mean age = 60.17, SD = 4.57; 83.33% females). Bilinguals' L1 were Italian (N = 11), English (N = 1). Bilinguals' L2 were English (N = 5), Spanish (N = 2), French (N = 2), Italian (N = 1), German (N = 1), modern Greek (N = 1). All the participants except one (UK) were born in Italy (N = 10).

Some bilingual participants reported some experience in L3 (English, N = 6; Spanish, N = 3; French, N = 2) and L4 (French, N = 2; Spanish, N = 1; N = 1). Two participants reported some experience in Italian regional dialects.

The control group included 12 participants (mean age = 58.33, SD = 3.52; 75% females). The L1 of all participants was Italian. Some participants of the control group reported some experience in L2 (English, N =

5; French, N = 1) and L3 (German, N = 1; French, N = 2; Russian, N = 1). Their experience in L2 and L3 was constrained to foreign language courses, and they reported marginal daily use of any foreign language. Four participants of the control group reported some experience in Italian dialects. One participant was born in a foreign country (UK, N = 1) while all the others were born in Italy. Participants were recruited using social networks and word of mouth. There were no differences between the language groups in SES and total Raven's score. See Table 3 for background information about participants and their language knowledge

Measures of bilingualism

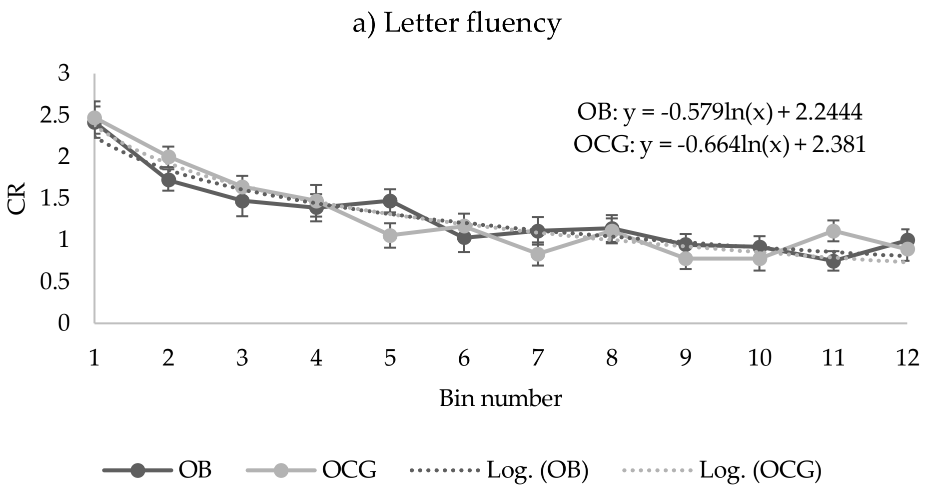
There were significant differences on subjective language proficiency ratings and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from L1 to L2 and vice versa. Bilinguals' language profile is summarized in Table 4.

Experimental data results

Table 6 shows the means and standard deviation of dependent variables for each experimental task.

Verbal fluency - Semantic vs Letter condition

The CR showed a main effect of Condition ($F_{1,22} = 37.60$; $p < .01$; $\eta^2 = 0.63$) and participants produced more CR in semantic (mean = 19.52) than in letter condition (mean = 15.33). The sub-RT revealed a main effect of Condition ($F_{1,22} = 6.98$; $p = .01$; $\eta^2 = .24$) and participants had lower Sub-RT in semantic (mean = 21.78 s) than in letter fluency (mean = 24.36 s). Cluster size revealed a main effect of Condition ($F_{1,22} = 87.98$; $p < .01$; $\eta^2 = .80$) and it was larger in semantic (mean = 1.16) than in letter condition (mean = 0.53). The Group x Condition interaction was marginally significant ($F_{1,22} = 3.11$; $p = .09$; $\eta^2 = .12$). Both groups showed a higher cluster size in semantic fluency than in letter fluency (bilinguals: $F_{1,22} = 62.10$; $p < .01$; $\eta^2 = .74$; control group: $F_{1,22} = 28.99$; $p < .01$; $\eta^2 = .57$). Furthermore, bilinguals showed a higher cluster size in semantic fluency than the control group ($F_{1,22} = 4.36$; $p = .05$; $\eta^2 = .16$). 1st-RT and number of switching did not show any significant differences. The FDS score did not show significant differences between group ($p = .27$). In both letter and semantic fluency initiation parameter and the slope was not significantly different ($F < 1$). Figure 2 represents the time-course of the CR.



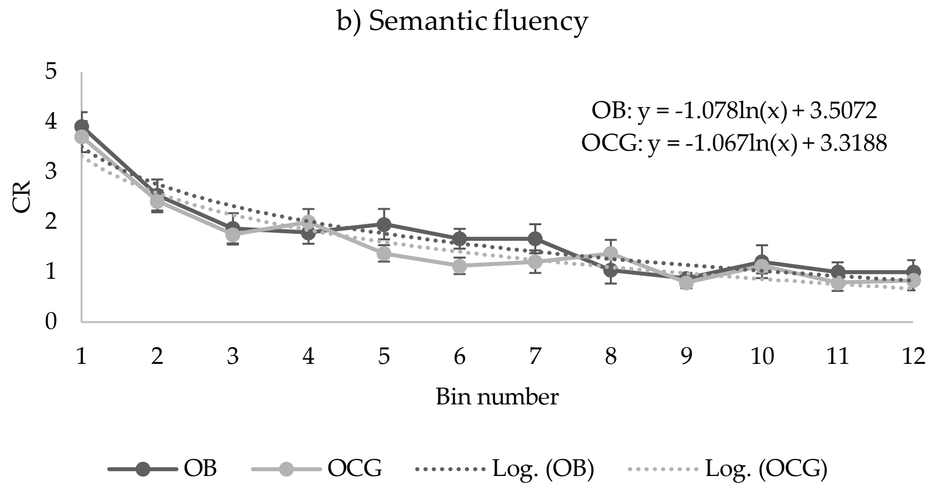


Figure 4. Group comparison (older adults) of CR produced as a function of 5-sec time interval in a) letter and b) semantic fluency. Error bars represent standard errors of the mean.

Semantic verbal fluency - Animal vs Fruit conditions

The CR showed a main effect of Condition ($F_{1,22}$ = 24.06; p < .01; η^2 = .52) and participants produced more CR in animal (mean = 21.96) than in fruit condition (mean = 17.08). 1st-RT revealed a main effect of Group ($F_{1,22}$ = 6.89; p = .01; η^2 = .24) and bilinguals had longer 1st-RT than the control group (1.33 vs 0.91 s). The sub-RT revealed a main effect of Condition ($F_{1,22}$ = 10.97; p < .01; η^2 = .33) and participants had lower Sub-RT in fruit (mean = 19.76 s) than in animal condition (mean = 23.80 s). Cluster size revealed a significant effect of Group ($F_{1,22}$ = 4.36; p = .05; η^2 = .16) and bilingual had larger cluster size than the control group (1.29 vs 1.03). Number of switches revealed a main effect of Condition ($F_{1,22}$ = 23.05; p < .01; η^2 = .51). Participants made more switches in animal (mean = 10.67) than in fruit condition (mean = 7.21).

Letter verbal fluency – F vs A vs S conditions

The CR and 1st-RT analysis did not show a main effect of Group (F < 1), Condition (F < 1) nor an interaction (p > .05). Sub-RT showed a marginally significant effect of condition ($F_{2,44}$ = 2.57; p = .09; η^2 = .10) and of the interaction Group x Condition ($F_{2,44}$ = 2.81; p = .07; η^2 = .11). No significant difference emerges between the two language groups. For bilinguals, the sub-RT in the "F" condition were significantly lower than in the "A" ($F_{1,22}$ = 11.35; p < .01; η^2 = .34) and "S" ($F_{1,22}$ = 5.61; p = .03; η^2 = .20) conditions. In the control group, sub-RT in the "F" condition are significantly lower than in the "A" condition ($F_{1,22}$ = 4.85; p = .04; η^2 = .18). Cluster size revealed a main effect of Condition ($F_{2,44}$ = 3.59; p = .04; η^2 = .14) and cluster size was lower in F condition than A condition ($F_{1,22}$ = 7.68; p = .01; η^2 = .26). Number of switches revealed a main effect of Condition ($F_{2,44}$ = 5.10; p = .01; η^2 = .19). Participants made more switches in F condition than the A condition ($F_{1,22}$ = 8.57; p < .01; η^2 = .28). The difference between F and S condition was marginally significant ($F_{1,22}$ = 3.79; p = .06; η^2 = .15).

Table 6. Means and standard deviations (in parentheses) of verbal fluency tasks for older adult participants.

	Older Bilinguals	**Older Control Group**
CR		
Semantic	20.54 (4.91)	18.50 (3.04)
Animal	23.25 (6.25)	20.67 (4.75)
Fruit	17.83 (4.43)	16.33 (3.02)
Letter	15.36 (4.18)	15.31 (3.58)
F	14.50 (4.62)	16.42 (3.60)
A	15.42 (3.78)	14.58 (4.42)
S	16.17 (6.31)	14.92 (4.38)

FDS	0.25 (0.18)	0.17 (0.13)
1st-RT		
Semantic	1.33 (0.49)	0.91 (0.26)
Animal	1.33 (0.72)	0.74 (0.37)
Fruit	1.33 (0.67)	1.08 (0.45)
Letter	1.69 (1.43)	1.60 (2.15)
F	1.36 (1.98)	0.60 (0.54)
A	2.37 (3.96)	1.66 (1.88)
S	1.34 (1.02)	2.54 (6.23)
Sub-RT		
Semantic	21.63 (3.30)	21.93 (2.94)
Animal	23.27 (3.28)	24.32 (4.45)
Fruit	19.99 (5.53)	19.53 (3.69)
Letter	24.55 (2.44)	24.17 (3.54)
F	21.70 (3.52)	23.74 (4.17)
A	25.31 (3.10)	26.10 (3.96)
S	26.65 (3.98)	22.68 (8.39)
Initiation - Semantic	3.51 (0.48)	3.32 (0.62)
Slope - Semantic	1.08 (0.28)	1.06 (0.35)
Initiation - Letter	2.24 (0.43)	2.38 (0.60)
Slope - Letter	0.58 (0.15)	0.66 (0.27)
Cluster size		
Semantic	1.29 (0.33)	1.03 (0.26)
Animal	1.19 (0.48)	0.88 (0.54)
Fruit	1.39 (0.56)	1.19 (0.59)
Letter	0.54 (0.28)	0.53 (0.30)
F	0.36 (0.28)	0.43 (0.42)
A	0.78 (0.50)	0.55 (0.40)
S	0.31 (0.09)	0.47 (0.13)
Number of switches		
Semantic	8.87 (2.37)	9.00 (2.41)
Animal	10.75 (3.52)	10.58 (2.84)
Fruit	7.00 (2.59)	7.42 (2.84)
Letter	9.97 (3.23)	9.83 (2.82)
F	10.67 (3.34)	11.50 (3.80)
A	9.00 (3.88)	8.83 (3.19)
S	10.25 (3.91)	9.17 (3.41)

Note. CR: Correct responses; FDS: Fluency Difference Score, Sub-RT: Subsequent reaction times.

General discussion

The verbal fluency task is a neuropsychological test commonly used to assess language skills. Indeed, most studies adopt this task to evaluate performance only quantitatively. However, qualitative and time-course analysis can be used to assess executive functioning. In line with previous studies (e.g., Patra et al., 2020), participants show better performance in semantic fluency than in letter fluency, producing higher CR and showing longer sub-RT and larger cluster sizes. These results highlight that participants could continue generating responses longer in the semantic condition.

Furthermore, older bilinguals showed a larger cluster size than the control group in the semantic condition, indicating a better ability to use cluster strategies. When comparing participants' performance in the two semantic categories, participants showed better performance in the "animal" condition with more CR and larger cluster size and maintained a better performance over time (higher sub-RT). It should be noted that the number of possible items for the animal category is higher than for the fruit category. Moreover, a difference between young adult bilinguals and the control group emerged. Bilinguals evidenced longer sub-RT than the control group, but, according to Luo et al. (2010), the interpretation of this variable depends on the number of CR. Bilinguals produced more CR than the control group, but this difference was only marginally significant, indicating a poorer performance of bilinguals. It may be hypothesized that cross-linguistic interference influenced the language production of bilingual participants. Bilinguals could have greater difficulty in retrieving terms despite knowing them. This interference increased with increasing linguistic competence in the L2 (Paap et al., 2019). Although the participants in this study were unbalanced bilinguals, their self-report assessment of L2 language proficiency was high. Older adult bilinguals tended to have a larger cluster size than the control group and a longer 1st-RT; thus, bilinguals used the cluster strategy to recall words better, but the time required for task preparation was longer. Both young and older adults found the "F" the most demanding condition considering letter fluency. Participants showed shorter sub-RT and smaller cluster size, indicating a faster-declining rate of retrieval because most of the CR were produced early during the trial.

Moreover, a smaller cluster size and a higher number of switches could indicate a failure to cluster (i.e., difficulties in retrieving new words within a cluster; Kavé et al., 2008). In the group of young adults, the bilinguals showed worse performance than the control group in the letter F. Bilinguals produced the same number of CR but had longer Sub-RT. This result indicates that more time was required to produce the same number of words. In contrast, no differences emerged in the group of older adults indicating that both groups experienced the same difficulty. As evidenced in previous studies (Bennet & Verney, 2019; Gollan et al., 2002) and reported in the systematic review included in Chapter 3, this study confirms that the letter used for the assessment affects participants' performance.

Limitations and conclusion

In conclusion, contrary to what was expected, the results do not show any advantages regarding the indices relating to the executive component, and no disadvantages emerge regarding the indices relating to the linguistic component. Conversely, the differences that emerged would seem to confirm a disadvantage for the group of bilinguals.

Some characteristics of the participants could have influenced the results. The bilingual participants were all unbalanced late bilinguals. Future studies should include different types of bilinguals to verify the effect of these characteristics on performance on the verbal fluency task. Furthermore, the bilingual participants knew different L2s, resulting in a different degree of cross-linguistic interference.

Considering the control group, according to a different definition of bilingualism, they could be considered low-proficient bilinguals. In Italy, teaching a foreign language was compulsory from 1962, while from 1987-88, an experiment for teaching a second foreign language started in middle school. Subsequently, since 2004, the teaching of a second foreign language has become compulsory. Moreover, most degree courses at Italian universities include at least one English language course. For this reason, selecting native Italian speakers who do not know a second language is almost impossible.

Considering the qualitative analysis of semantic fluency, the method used to classify words into clusters may have affected the results of this study. Specifically, the animal classification included twenty-four possible categories, while the fruit classification included only nine. This aspect could have influenced the number of switches and the width of the clusters. Therefore, we highlight the need to create new standardized procedures for administering and evaluating of semantic fluencies that include quantitative, qualitative, and time-course analysis guidelines (e.g., Lehtinen et al., 2021).

Chapter 5: Performance difference in verbal fluency in bilinguals: the role of language proficiency

Introduction

The verbal fluency test is a neuropsychological test widely used in clinical and experimental settings. It is a brief test that does not require special equipment and is easy to use. Although verbal fluency is mainly used to assess language skills, this test can also provide information about executive functioning (see Patra et al., 2020). Several studies on bilingualism used this test to compare the performance of monolinguals and bilinguals. Some studies did not show significant differences between the groups' performance (e.g., Morrison et al., 2020; Zeng et al., 2019), while others showed better performance by bilingual (e.g., Massa et al., 2020; Patra et al., 2020) or monolingual (e.g., Anderson et al., 2017; Bialystok et al., 2009) participants. The recent meta-analysis by Lehtonen et al. (2018), which considered all studies published from 1999 to 2018, revealed the presence of a disadvantage for bilinguals in performing this task. Two main hypotheses have been proposed to explain the reduced linguistic performance of bilinguals. According to the weaker link hypothesis (Michael & Gollan, 2005), the reduced linguistic performance of bilinguals depends on the reduced use of each known language. For instance, bilinguals would show greater difficulty recalling low-frequency words than monolinguals (Sandoval et al., 2010).

In contrast, according to the Inhibitory Control model (Green, 1998), language difficulties depend on the activation of both languages in the brain, even when only one language is used. This activation would lead to cross-linguistic interference, which would affect bilinguals' performance. Specifically, bilinguals would have greater difficulty recalling high-frequency words, which would generate greater cross-linguistic interference since they are available in both known languages (Sandoval et al., 2010).

Although the verbal fluency test is widely used, a limited number of studies (e.g., Rosselli et al., 2002; Sandoval et al., 2010) investigated bilinguals' performance on verbal fluency tests by assessing performance in both known languages. These studies highlighted some factors that appear to influence bilinguals' performance. Specifically, the language used in the assessment seems to influence the type of words produced. For example, in Rosselli et al. (2002), English-Spanish bilinguals generated more English grammatical words than Spanish. Bilinguals seem to be more likely to produce cognates (i.e., similar translations between languages). This result supports the hypothesis that dual-language activation affects verbal fluency performance. Furthermore, cultural aspects seem to influence language production as different living environments could produce cross-linguistic differences (Pekkala et al., 2009).

This study aims to assess the linguistic and executive performance of a group of bilingual young adults. As previously pointed out by Sandoval et al. (2010), it is assumed that bilinguals will show better performance when assessed in their dominant language producing a higher number of correct responses and longer subsequent response time in the dominant language. To our knowledge, this is the first study that includes unbalanced bilinguals dominant in the Italian language, and that assesses their performance in verbal fluency using qualitative, quantitative indices and time-course analysis.

Method

Participants

23 bilinguals between 18-34 years took part in the study (73,91% females, mean age: 26.70, SD 3.29). All participants were bilingual speakers of Italian and one additional language (English, n =15; Spanish, n = 4; French, n = 3; German, n = 1)

Most of the participants were born in Italy (N = 21) while two of them in a foreign country (Brazil, N = 1; Moldova, N = 1).

Some bilingual participants reported some experience in L3 (English, N = 7; Dutch = 3; Spanish, N = 2; Portuguese, N = 2; French, N = 1; German, N = 1; Romanian, N = 1; Russian, N = 1) and L4 (Spanish, N = 4; English, N = 1). Two participants reported some experience in Italian regional dialects. Participants were recruited using social networks and word of mouth.

Measures of bilingualism

There were significant differences in subjective language proficiency ratings and bilinguals judged their knowledge of Italian superior to their knowledge of L2 (speaking, comprehension, reading and writing abilities). According to the results of LDQ, Italian was the dominant language. No significant differences emerged in the tendency to switch from Italian to L2 and vice versa. Bilinguals' language profile is summarized in Table 1.

Table 1. Bilinguals' language profile. Means (standard deviation).

	L1	**L2**	Statistical results
Young adults (N = 23)			
Age of exposure (in years)	0.52 (1.75)	10.65 (7.06)	$t(22) = -6.42$; $p < .01$
Self-rated language proficiency			
Proficiency – Speaking (1-10)	9.83 (0.58)	8.61 (1.37)	$t(22) = 4.13$; $p < .01$
Proficiency – Comprehension (1-10)	9.91 (0.29)	9.09 (1.00)	$t(22) = 4.23$; $p < .01$
Proficiency – Writing (1-10)	9.83 (0.39)	8.35 (1.61)	$t(22) = 4.23$; $p < .01$
Proficiency – Reading (1-10)	10.00 (0.00)	9.30 (0.76)	$t(22) = 4.36$; $p < .01$
Language dominance	19.13 (5.70)	7.48 (4.54)	$t(22) = 6.10$; $p < .01$
Language switching habit	8.48 (2.11)	8.56 (2.25)	$t(22) = -0.17$; p = n.s.

Demographic information

The same variables measured in the previous study (see Chapter 4) were used in this study.

Language history

The same questionnaire used in the previous study (see Chapter 4) were used in this study.

Verbal fluency tests

The same tests used in the previous study (see Chapter 4). The tests were performed in Italian and in the participant's L2.

Procedure

All participants completed the questionnaires and cognitive tests online. The questionnaires were administered online using the KoboToolBox platform (www.kobotoolbox.org), while the cognitive tests were and administered during an online interview. All participants completed the Language and Social Background Questionnaire (adapted from Anderson et al., 2018), the letter and semantic fluency tasks. The bilingual participants completed the Language Switching Questionnaire (LSQ, adapted from Rodriguez-Fornells et al., 2012) and the Language Dominance Questionnaire (LDQ, adapted from Dunn & Tree, 2009). The questionnaires were presented in a fixed order while verbal fluency tests were presented in a counterbalanced order among the participants. All testing was conducted in Italian.

Statistical analysis

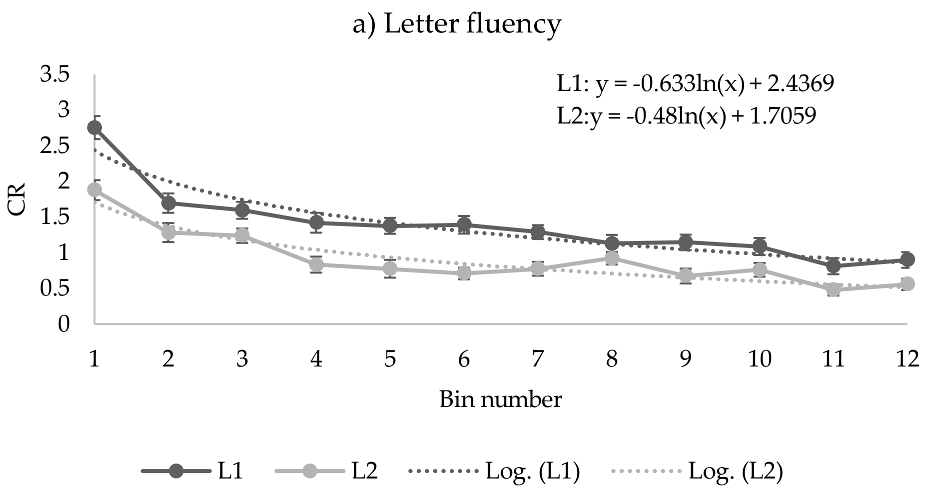
The same variables measured in the previous study (see Chapter 4) were used.
Analyses of variance (ANOVAs) were used to compare bilingual's performance in the two languages, with language as the independent variable (L1, L2). ANOVAs were conducted to compare test performance in different conditions (semantic vs letter; animal vs fruit; F vs A vs S).

Results

Table 2 shows the means and standard deviation of dependent variables for each experimental task.

Verbal fluency - Semantic vs Letter condition

The CR showed a main effect of Condition ($F_{1,44}$ = 23.33; p < .01; η^2 = .35) and participants produced more CR in semantic (mean = 16.58) than in letter condition (mean = 13.74). The main effect of Language was significant ($F_{1,44}$ = 35.79; p < .01; η^2 = .45) and participants produced more words in L1 (mean = 18.09) than L2 (mean = 12.22).
1st-RT evidenced a main effect of Condition ($F_{1,44}$ = 4.31; p = .04; η^2 = .09) and participants showed longer 1st-RT in letter (mean = 1.49) than in semantic condition (mean = 0.96). The main effect of Language was significant ($F_{1,44}$ = 5.68; p = .02; η^2 = .11) and participants showed longer 1st-RT in L2 (mean = 0.89) than L1 (mean = 1.57). The interaction was significant ($F_{1,44}$ = 6.15; p = .02; η^2 = .12). In semantic condition there was no significant difference between L1 and L2 (F < 1). In letter condition, participants showed longer 1st-RT in L2 than L1 ($F_{1,44}$ = 6.51; p = .01; η^2 = .13).
The sub-RT revealed a main effect of Condition ($F_{1,44}$ = 26.19; p < .01; η^2 = .37) and participants had lower Sub-RT in semantic (mean = 21.24 s) than in letter fluency (mean = 24.46 s). The main effect of Language was significant ($F_{1,44}$ = 8.29; p < .01; η^2 = .16) and participants showed longer sub-RT in L1 (mean = 23.80 s) than in L2 (mean = 21.90 s).
Cluster size revealed a main effect of Condition ($F_{1,44}$ = 34.43; p < .01; η^2 = .44) and it was larger in semantic (mean = 0.96) than in letter condition (mean = 0.55). The main effect of Language was significant ($F_{1,44}$ = 4.21; p = .05; η^2 = .09) and participants evidenced a larger cluster size in L1 (mean = 0.84) than L2 (mean = 0.67).
Number of switching showed a main effect of Language ($F_{1,44}$ = 27.57; p < .01; η^2 = .38) and participants made more switches using L1 (mean = 9.71) than L2 (6.75).
The FDS score did not show significant differences between group (F < 1). In both condition, participants evidenced a lower int in L2 than L1 (semantic: $F_{1,44}$ = 14.10; p < .01; η^2 = .24; letter: $F_{1,44}$ = 17.32; p < .01; η^2 = .28).
Slope was significant different only in letter condition ($F_{1,44}$ = 4.34; p = .04; η^2 = .09) and participants evidenced a flatter slope in L2 than L1. Figure 1 represents the time-course of the CR.



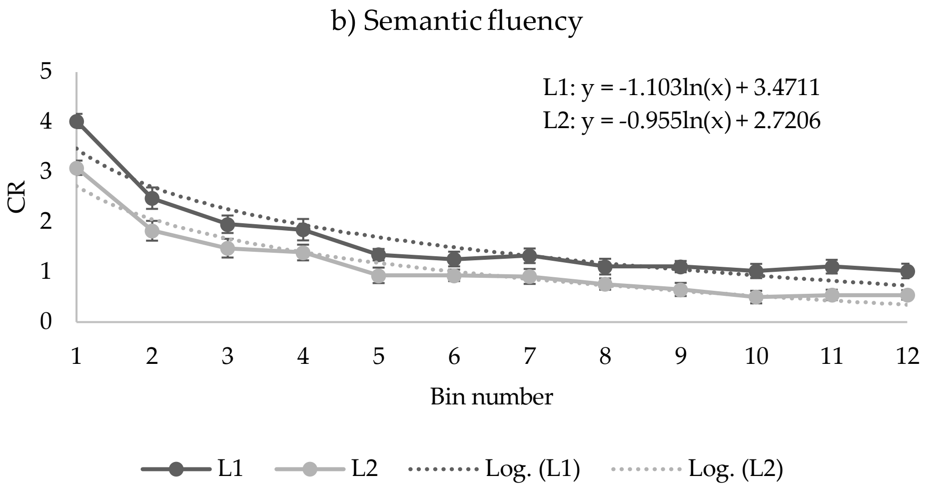


Figure 1. Language comparison of CR produced as a function of 5-sec time interval in a) letter and b) semantic fluency. Error bars represent standard errors of the mean.

Semantic verbal fluency - Animal vs Fruit conditions

The CR showed a main effect of Condition ($F_{1,44} = 50.74$; $p < .01$; $\eta^2 = .53$) and participants produced more CR in animal (mean = 19.43) than in fruit condition (mean = 13.72). The main effect of Language was significant ($F_{1,44} = 26.12$; $p < .01$; $\eta^2 = .37$) and participants produced more words in L1 (mean = 19.59) than L2 (mean = 13.56). The Group x Condition interaction was significant ($F_{1,44} = 9.70$; $p < .01$; $\eta^2 = .18$). Participants produced more CR in animal than in fruit condition using both languages (L1: $F_{1,44} = 52.41$; $p < .01$; $\eta^2 = 54$; L2: $F_{1,44} = 8.03$; $p < .01$; $\eta^2 = .15$). 1stRT did not show significant differences between groups.
The sub-RT revealed a main effect of Condition ($F_{1,44} = 37.34$; $p < .01$; $\eta^2 = .46$) and participants had shorter Sub-RT in fruit (mean = 18.91 s) than in animal condition (mean = 23.57 s). The main effect of Language was significant ($F_{1,44} = 5.15$; $p = .03$; $\eta^2 = .10$) and bilinguals had longer sub-RT using L1 (mean = 22.18 s) than L2 (20.30 s). The interaction was marginally significant ($F_{1,44} = 3.49$; $p = .07$; $\eta^2 = .07$). In animal condition, participants had longer sub-RT using L1 than L2 ($F_{1,44} = 9.99$; $p < .01$; $\eta^2 = .19$). No differences emerged in fruit condition.
Cluster size revealed a main effect of Condition ($F_{1,44} = 18.89$; $p < .01$; $\eta^2 = .30$) and cluster size was larger in fruit (mean = 1.14) than in animal condition (mean = 0.78). The main effect of Language was significant ($F_{1,44} = 4.32$; $p = .05$; $\eta^2 = .09$) and bilinguals had larger cluster size using L1 (mean = 1.06) than L2 (mean = 0.85).
Number of switches revealed a main effect of Condition ($F_{1,44} = 76.19$; $p < .01$; $\eta^2 = .63$). Participants made more switches in animal (mean = 10.39) than in fruit condition (mean = 5.78). The main effect of Language was significant ($F_{1,44} = 22.49$; $p < .01$; $\eta^2 = .34$) and bilinguals made more switches using L1 (mean = 9.56) than L2 (mean = 6.61). The interaction was significant ($F_{1,44} = 8.79$; $p < .01$; $\eta^2 = .17$). In both conditions, participants made more switches using L1 than L2 (animal: $F_{1,44} = 21.74$; $p < .01$; $\eta^2 = .33$; fruit: $F_{1,44} = 4.91$; $p = .03$; $\eta^2 = .10$).

Letter verbal fluency – F vs A vs S conditions

The CR evidenced a main effect of Condition ($F_{2,84} = 4.99$; $p < .01$; $\eta^2 = .11$) and bilinguals produced more word in S condition than A ($F_{1,42} = 14.18$; $p < .01$; $\eta^2 = .25$). The main effect of Language was significant ($F_{1,42} = 25.36$; $p < .01$; $\eta^2 = .38$) and participants generated more words using L1 (mean = 16.59) than L2 (mean = 10.87). The interaction between the main effects was significant ($F_{2,84} = 3.70$; $p = .03$; $\eta^2 = .08$). Bilinguals produced more word in all conditions using L1 (F: $F_{1,42} = 20.70$; $p < .01$; $\eta^2 = .33$; A: $F_{1,42} = 30.42$; $p < .01$; $\eta^2 = .42$; S: $F_{1,42} = 8.19$; $p < .01$; $\eta^2 = .16$). 1st-RT evidenced a main effect of Language was significant ($F_{1,42} = 6.05$; $p = .02$; $\eta^2 = .13$) and participants showed longer 1st-RT in L2 (mean = 2.15) than L1 (mean = 0.84). The interaction was marginally significant ($F_{2,84}$ = 2.50; p = .09; η^2 = .06). In A condition, participants generated evidenced longer 1st-RT in L2 than L1 ($F_{1,42}$ = 4.05; p = .05; η^2 = .09). Sub-RT did not show significant differences.

Cluster size revealed a main effect of Condition ($F_{2,84}$ = 4.79; p = .01; η^2 = .10) and cluster size was lower in F condition than S condition ($F_{1,42}$ = 13.66; p < .01; η^2 = .25). The main effect of Language was significant ($F_{1,42}$ = 4.04; p = .05; η^2 = .09) and bilinguals showed larger cluster size using L1 (mean = 0.62) than L2 (mean = 0.41).

Number of switches revealed a main effect of Condition ($F_{2,84}$ = 3.21; p = .05; η^2 = .07). Participants made more switches in F condition than A condition ($F_{1,42}$ = 4.61; p = .04; η^2 = .10). The main effect of Language was significant ($F_{1,42}$ = 18.34; p < .01; η^2 = .30) and bilinguals made more switches using L1 (mean = 9.85) than L2 (mean = 6.95). The interaction was significant ($F_{2,84}$ = 3.63; p = .03; η^2 = .08). In F and A conditions, participants made more switches using L1 than L2 (F: $F_{1,42}$ = 17.56; p < .01; η^2 = .29; A: $F_{1,42}$ = 13.48; p < .01; η^2 = .24).

Table 2. Means and standard deviations (in parentheses) of verbal fluency tasks.

	L1	**L2**
CR		
Semantic	19.59 (3.53)	13.56 (4.41)
Animal	23.70 (5.68)	15.17 (5.54)
Fruit	15.48 (3.79)	11.96 (4.03)
Letter	16.59 (4.32)	10.88 (3.08)
F	16.56 (5.03)	10.52 (3.57)
A	16.48 (4.94)	9.29 (3.51)
S	16.74 (4.78)	12.81 (4.28)
FDS	0.14 (0.20)	0.12 (0.41)
1st-RT		
Semantic	0.94 (0.42)	0.99 (0.68)
Animal	0.88 (0.66)	1.03 (0.72)
Fruit	1.00 (0.75)	0.94 (0.88)
Letter	0.84 (0.53)	2.14 (2.40)
F	0.91 (0.77)	1.35 (1.45)
A	0.79 (0.77)	3.47 (6.34)
S	0.81 (0.65)	1.62 (1.45)
Sub-RT		
Semantic	22.18 (2.73)	20.30 (2.88)
Animal	25.22 (3.29)	21.92 (3.78)
Fruit	19.13 (3.28)	18.68 (4.73)
Letter	25.42 (2.20)	23.50 (4.19)
F	25.02 (3.38)	22.80 (5.84)
A	25.38 (4.06)	23.79 (6.34)
S	25.85 (3.43)	25.97 (4.92)
Initiation - Semantic	3.47 (0.73)	2.72 (0.62)
Slope - Semantic	1.10 (0.35)	0.95 (0.23)
Initiation - Letter	2.43 (0.67)	1.70 (0.51)
Slope - Letter	0.63 (0.26)	0.49 (0.22)
Cluster size		
Semantic	1.06 (0.30)	0.85 (0.37)
Animal	0.90 (0.39)	0.65 (0.35)
Fruit	1.22 (0.49)	1.06 (0.51)
Letter	0.62 (0.43)	0.48 (0.38)
F	0.44 (0.37)	0.33 (0.26)
A	0.67 (0.55)	0.40 (0.50)
S	0.75 (0.61)	0.50 (0.32)
Number of switches		
Semantic	9.56 (2.13)	6.61 (2.10)
Animal	12.65 (3.81)	8.13 (2.67)
Fruit	6.48 (1.85)	5.09 (2.37)

Letter	9.85 (2.21)	6.89 (2.39)
F	11.00 (3.63)	7.00 (2.55)
A	9.39 (2.90)	6.05 (3.14)
S	9.17 (2.37)	7.81 (3.06)

Note. CR: Correct responses; FDS: Fluency Difference Score, Sub-RT: Subsequent reaction times.

Discussion

The verbal fluency test is useful for assessing language skills and executive functioning, widely used in clinical and experimental settings. This study assessed whether the performance of unbalanced bilinguals varies according to the language used for the assessment. It is important to clarify this aspect since the percentage of bilinguals is constantly increasing and suitable instruments for assessing this population are not always available. Neuropsychological instruments are culturally biased, and normative data for a certain population may not be suitable for another population (Pekkala et al., 2009). Moreover, there is still a need to clarify whether only responses given in the assessment language should be considered in the rating of bilinguals or responses provided in the second known language should also be accepted to avoid penalizing bilingual participants (e.g., Lehtinen et al., 2021).

In line with Sandoval et al. (2010) and our hypothesis, bilinguals perform better in both semantic and letter fluency when tested in their dominant language.

Participants produced more CR, made better use of the cluster strategy, and maintained better performance over time when using L1 (i.e., longer sub-RT). According to previous studies (e.g., Luo et al., 2010), letter fluency was the most demanding condition.

Participants produced fewer CR in a longer time. Moreover, when performing the test using L2, they needed more time before generating words. Considering the different stimuli of letter fluency, bilinguals produce fewer CR when using L2 in all conditions.

Considering L1, no differences emerged between the stimuli. Conversely, considering L2, participants produced fewer CR in the letters A and F than in the letter S. Because participants with different language pairs were included, it is impossible to determine which language these conditions were more difficult.

As highlighted by Rosselli et al. (2002), who assessed bilingual participants using the same triplet, no differences emerged between the different conditions for the Spanish language. In contrast, when the test was conducted in English, the S was the easiest condition and the A the most difficult. Future studies should compare different groups of bilinguals with different language pairs to determine how language performance varies across languages.

Considering semantic fluency, the animal condition is easier than the fruit condition in L1 and L2. The choice of semantic category influences participants' performance. In particular, the animal category would seem to be the most suitable for assessing bilinguals as it tends to be stable across cultures (Ardila, 2020).

Finally, considering the linguistic production over time, both in the semantic and letter conditions, the bilingual participants showed a higher initiation parameter when performing the test in L1 than in L2. The initiation parameter reflects the initial linguistic resources to perform the task, and the result confirms a better linguistic mastery in L1. Conversely, in the letter fluency, a flatter slope emerged in L2. According to previous studies (e.g., Patra et al., 2020), a flatter slope indicates that participants could maintain their performance across time. However, in this study, this index would not seem to reflect a better performance but should be interpreted considering other aspects (e.g., the average number of CR per bin). In this study, the absence of variability in language production over time depends on the total number of CR, which was very small. The reduced language production in L2 can be hypothesized to be due to cross-linguistic interference generated by L1, i.e., the dominant language for all participants.

Limitations and conclusions

In conclusion, the results of this study show that the language used to assess bilingual participants' performance in the verbal fluency test plays a role in determining the results. Bilinguals achieve better performance when they are evaluated in their dominant language. This aspect is particularly relevant in both experimental and clinical settings. Future studies should determine the dominant language of bilingual participants and consider this aspect when interpreting the results of their studies, especially when instruments with verbal stimuli are used. Finally, it would be useful to conduct further analyses of the following participants' data. Analyzing the type of words produced (e.g., concrete vs. abstract, cognates vs. non-cognates) would provide more information on the strategies used by the participants in performing this experimental task.

Finally, all instructions were given in Italian, even those related to verbal fluency in L2. This aspect may have influenced participants' performance. Future studies should include experienced evaluators in the different languages known to the participants.

General conclusions

This doctoral thesis investigated the effect of bilingualism on executive functions. Bilingualism is a sociodemographic condition that characterizes more than half of the world's population; therefore, this topic has aroused wide scientific interest. Several authors focused on the effect of bilingualism on executive functions. According to the Joint Activation Model (Green, 1998), both languages are always active in a bilingual's brain regardless of the language used at the given moment. The constant need to control both known languages to use the one suitable for each moment would positively affect cognitive functions.

As evidenced by the systematic review investigating the relationship between bilingualism and cognitive and motor inhibition included in Chapter 1, more than half of the studies that examined this relationship did not show a significant difference between the bilingual and monolingual populations. Conversely, most of the remaining studies showed evidence favoring a bilingual advantage. Paap & Greenberg (2013) suggested that the positive findings in favor of a bilingual advantage could be Type 1 errors. If this assumption is true, the same number of studies should exist in favor of a monolingual population advantage (Bialystok & Craik, 2022). The review in Chapter 1 challenges this assumption; in fact, only about 9% of the studies show results in favor of a monolingual advantage. This result is in line with the statement by Grundy (2020), "when group differences appear on EFs tasks, even if they are rare, the difference almost always favors bilinguals outperforming monolinguals." Different authors suggested understanding the circumstances under which this advantage emerges. Indeed, both methodological factors (e.g., type and complexity of the tasks) and factors relating to the participants' linguistic history (e.g., age of acquisition, use of languages) would affect the results.

The study reported in Chapter 2 aimed to investigate the effect of the experimental task used for the assessment. The importance of this methodological aspect was previously highlighted by Paap et al. (2015), who suggested including at least two tasks to assess the same function to exclude that the effect was task-specific. Moreover, as other studies highlighted (e.g., Paap & Sawi, 2016), the experimental tasks used to evaluate executive functions showed low convergent validity. Therefore, the study included all main tasks to assess cognitive and motor inhibition. Moreover, since verbal stimuli seemed to affect the performance of bilingual participants, the nonverbal version of all the tasks was included. Only the Stroop task was administered in both verbal and non-verbal versions to test the effect of verbal stimuli. Furthermore, participants from different age groups (young adults, middle-aged adults, and older adults) were included because, as noted in the review in Chapter 1, the percentage of differences between groups varies according to the age of the participants. The study results did not show a better inhibition ability for the bilingual population. The absence of significant differences could depend on several factors.

Considering the experimental tasks, the classic version of the tasks may be too simple for the population included in this study. All participants completed the tasks correctly, and the accuracy rate was greater than 94% in all tasks. As evidenced by Bialystok and Craik (2022), if the tasks did not exceed the resources of each group, it is difficult to observe significant differences in the groups' performance, especially when including participants at the peak of their cognitive development (such as those included in this study). Future studies should include different versions of the same task while manipulating the complexity of the test to confirm this hypothesis.

Despite no differences emerging at the behavioral level, a positive effect of the condition of bilingualism should not be excluded. Wong et al. (2016) evidenced that bilingualism leads to differences in brain structures, such as regional activation in areas associated with cognitive control and executive functioning and increased gray and white matter volume. Future studies should simultaneously conduct behavioral assessments as well as neuroimaging studies.

Conversely, the significant differences between the groups that emerged in Study 1 (i.e., faster overall RTs) could be interpreted as a better general monitoring ability for the bilingual group. Recently, Bialystok and Craik (2022) suggested that immersion in a bilingual environment positively affects attentional control. They defined attentional control as a mechanism that maintains current goals in an active state, facilitates cognitive operations that accomplish these goals, suppress interference, and switches processing resources to a different set of operations when cognitively beneficial. Specifically, they hypothesize that bilingualism does not increase attentional resources but makes the use of attentional resources more efficient. Consistent with this assumption, better reaction times would be attributable to fewer atypical longer RTs indicating fewer lapses in attentional control. Future studies should use different statistical approaches from the traditional ones to analyze the results to investigate this hypothesis (Bialystok & Craik, 2022). To examine whether the bilingual population has a better attentional control would be useful to use ex-Gaussian analyses (). This analysis uses the entire distribution of scores and extracts different measures associated with the mean tendency (μ, automatic aspects of processing) and exponential (ţ, monitoring, and attentional control) components of the overall RT.

Moreover, data-trimming procedures to exclude extreme RTs do not allow the analysis of attentional lapses. This study excluded reaction times that exceeded the mean by 2.5 SD. It would be useful to repeat the analyses without data trimming to test this hypothesis. Lastly, in this study, the presence of verbal stimuli did not appear to have negatively affected the performance of the bilingual participants. It is possible to hypothesize that this result depends on the characteristics of the participants (i.e., bilinguals dominant in the Italian language) and the language used in the assessment. Other features of the participants may have influenced the results of this study and the other two experimental studies included in this thesis. A dichotomous approach was used to classify the participants into bilingual and control groups. The choice to use the term control group was because the participants included in this group had partial knowledge of an L2 (L3 and L4 in some cases) due to the teachings given during compulsory schooling. By choosing a different definition of bilingualism, these participants would have been considered low-proficient bilinguals. Similarly, some bilinguals had marginal knowledge of an L3 and an L4 and could be defined as trilingual or multilingual. This aspect could have influenced the results of this study. Although there are not many studies on trilingualism and multilingualism, the results indicate that the consequences of trilingualism are not simply an extension of the consequences of bilingualism (Schroeder & Marian, 2017). No differences seem to emerge between trilinguals and bilinguals, considering inhibition ability. Instead, trilinguals seem to have better interference suppression but not better response inhibition (Jiang et al., 2022).

Some authors (e.g., Bialystok & Craik, 2022) pointed out that using a dichotomous approach to classifying participants (monolingual vs. bilingual) is not suitable to capture the differences and the peculiar characteristics of the groups. For this reason, it would be more appropriate to consider bilingualism as a process along a continuum. It would be useful to reanalyze the studies' results in this thesis using a different classification of participants that considers the peculiar aspects of their linguistic history. Indeed, as Green & Abutalebi (2013) suggested, the context in which bilinguals are immersed determines which and how much certain cognitive skills will be developed. The concept of "language entropy" (i.e., a measure of the social diversity in which each language is used) could be used for a classification of participants that takes these aspects into account (Gullifer & Titone, 2020). Language entropy was positively associated with the brain scores and is "a comprehensive and sensitive measure to model language balance and diversity within groups bilinguals and multilinguals" (Li et al., 2021). Moreover, a greater language entropy was related to smaller RT indices of conflict monitoring and goal maintenance but larger RT indices of inhibition of response conflict (Li et al., 2021).

In Chapter 3, a systematic review investigated executive functioning in bilinguals by analyzing the results obtained in the verbal fluency task. Patra et al. (2020) highlighted that previous studies assessed language skills and executive functioning separately. In contrast, the verbal fluency task allows assessing both components simultaneously. Most of the studies included in the systematic review did not show significant differences in quantitative indices. Conversely, the results of the qualitative and time-course analysis were mixed. The results suggested the importance of interpreting the results of verbal fluency using all the

indexes available to provide a more in-depth assessment of both linguistic competence and executive ability. Furthermore, the linguistic history of the participants seems to affect, positively or negatively, the performance of the bilingual population. This aspect is particularly relevant in clinical settings where standardized instruments for the bilingual population are not available.
According to the systematic review findings, the studies reported in Chapters 4 and 5 were conducted to investigate performance in verbal fluency using the different indices available. Moreover, only one other study included participants who were fluent in Italian (Massa et al., 2020).
In Chapter 4, it is reported that a study compared the performance in the verbal fluency task of a group of unbalanced bilinguals with that of a control group. This study showed that the unbalanced bilinguals had greater difficulty in verbal fluency.
Previous studies (e.g., Paap et al., 2019) evidence that cross-language interference affects bilingual performance. The amount of interference (from an L2 on L1) increase as L2 proficiency increase because L2 practice comes at the expense of more L1 use. Although the participants in this study were unbalanced bilinguals, their self-assessment of their L2 proficiency showed very high scores reflecting very good L2 mastery.
Finally, Chapter 5 reported a study that evaluated the performance in the verbal fluency task of a group of unbalanced bilinguals who performed the task in both their dominant language (L1) and their L2.
The results of this study are relevant in terms of both clinical and experimental aspects. Indeed, considering all the indices, bilinguals performed better when they performed the task in their dominant language. This result suggests the importance of assessing the linguistic dominance of bilinguals and considering this aspect when interpreting the results, especially when using instruments with verbal stimuli.
The language pairs known by participants could have influenced participants' performance (both in this study and in the previous study). Although the FAS triplet is the most used in phonological fluency, different language pairs might have produced a different degree of cross-linguistic interference. Previous studies (Bennet & Verney, 2019; Gollan et al., 2002) showed that triplet choice affects the performance of bilingual participants. Future studies should investigate this aspect by including participants with restricted knowledge of defined language pairs.
The results of the studies that have included verbal fluencies showed that only qualitative analysis of performance is insufficient to have an adequate understanding of the abilities of the assessed persons. However, it is necessary to develop standardized guidelines for administering and assessing verbal fluencies (Lehtinen et al., 2021). As the systematic review included in Chapter 3 revealed, studies use different categories for assessment, different ways of calculating the scores obtained in the case of quantitative analyses, and several classifications for qualitative analyses. For a correct comparison between the studies, the need to standardize the procedures used seems evident.
Moreover, Paplikar et al. (2021) suggest the need to conduct the assessment in the participant's dominant language, that the administered be adequately trained in the language of the evaluation, and the need to consider linguistic intrusions for scoring purposes.
Finally, previous studies (for a review, see Tao et al., 2021) indicate a protective effect of bilingualism on the cognitive and neural decline. However, this positive effect could delay diagnosing neurodegenerative diseases during neuropsychological evaluations since bilinguals with the same brain atrophy as monolinguals could show better cognitive performance. Further studies are necessary to investigate the best approach for the psychological assessment of bilinguals.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323-338.

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*(1), 250-263.

Anderson, J. A., Saleemi, S., & Bialystok, E. (2017). Neuropsychological assessments of cognitive aging in monolingual and bilingual older adults. *Journal of neurolinguistics*, *43*, 17-27.

Ansaldo, A. I., Ghazi-Saidi, L., & Adrover-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of clinical and experimental neuropsychology*, *37*(5), 455-470.

Antón, E., Carreiras, M., & Duñabeitia, J. A. (2019). The impact of bilingualism on executive functions and working memory in young adults. *PloS one*, *14*(2), e0206770.

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., ... & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in psychology*, *5*, 398.

Antón, E., García, Y. F., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly?. *Journal of Memory and Language*, *90*, 147-160.

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics*, *5*, 395-415.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*(2), 115-129.

Ardila, A. (2020). A cross-linguistic comparison of category verbal fluency test (ANIMALS): a systematic review. *Archives of Clinical Neuropsychology*, *35*(2), 213-225.

Barbu, C. A., Gillet, S., & Poncelet, M. (2020). Investigating the effects of language-switching frequency on attentional and executive functioning in proficient bilinguals. *Frontiers in Psychology*, *11*, 1078.

Bellegarda, M. A., & Macizo, P. (2021). Cognitive control and bilingualism: The bilingual advantage through the lens of dimensional overlap. *Frontiers in Psychology*, *12*, 173.

Bennett, J., & Verney, S. P. (2019). Linguistic factors associated with phonemic fluency performance in a sample of bilingual Hispanic undergraduate students. *Applied Neuropsychology: Adult*, *26*(4), 297-310.

Berroir, P., Ghazi-Saidi, L., Dash, T., Adrover-Roig, D., Benali, H., & Ansaldo, A. I. (2017). Interference control at the response level: Functional networks reveal higher efficiency in the bilingual brain. *Journal of Neurolinguistics*, *43*, 4-16.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*, 636–644. doi: 10.1111/1467-8624.00046

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

Bialystok, E. (2009). "Effects of bilingualism on cognitive and linguistic performance across the lifespan," in Streitfall Zweisprachigkeit–The Bilingualism Controversy, (VS Verlag für Sozialwissenschaften) 53–67.

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of experimental child psychology*, *110*(3), 461-468.

Bialystok, E., & Craik, F. I. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, 1-24.

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565.

Bialystok, E., Craik, F. I., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522-538.

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

Bialystok, E., Craik, F., & Luk, G. (2009). "Cognitive control and lexical access in younger and older bilinguals": Correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 828. https://doi.org/10.1037/a0015638

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103-119.

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and aging*, *29*(3), 696.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*(3), 610-629.

Blumenfeld, H. K., Bobb, S. C., & Marian, V. (2016). The role of language proficiency, cognate status and word frequency in the assessment of Spanish–English bilinguals' verbal fluency. *International journal of speech-language pathology*, *18*(2), 190-201.

Boersma, P., & Weenink, D. (2015) Praat version 5.4. 08. Doing phonetics by computer.

Bogulski, C. A., Rakoczy, M., Goodman, M., & Bialystok, E. (2015). Executive control in fluent and lapsed bilinguals. *Bilingualism: Language and Cognition*, *18*(3), 561-567.

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.

Botezatu, M. R., Miller, C. A., Johnson, J., & Misra, M. (2021). Event-related potentials reveal that bilinguals are more efficient in resolving conflict than monolinguals. *NeuroReport*, *32*(8), 721-726.

Brini, S., Sohrabi, H. R., Hebert, J. J., Forrest, M. R., Laine, M., Hämäläinen, H., ... & Fairchild, T. J. (2020). Bilingualism is associated with a delayed onset of dementia but not with a lower risk of developing it: a Systematic review with Meta-Analyses. *Neuropsychology review*, *30*(1), 1-24.

Caltagirone, C., Gainotti, G., Carlesimo, G. A., & Parnetti, L. (1995). Batteria per la valutazione del deterioramento mentale: I. Descrizione di uno strumento di diagnosi neuropsicologica. *Archivio di Psicologia, Neurologia e Psichiatria*.

Chabal, S., Schroeder, S. R., & Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Attention, Perception, & Psychophysics*, *77*(8), 2684-2693.

Chrysochoou, E., Kanaki, S., & Vivas, A. B. (2020). Executive functions in French-Greek early bilinguals: In search of the suggested bilingual advantage. *Psychology: The Journal of the Hellenic Psychological Society*.

Clare, L., Whitaker, C. J., Martyr, A., Martin-Forbes, P. A., Bastable, A. J., Pye, K. L., ... & Hindle, J. V. (2016). Executive control in older Welsh monolinguals and bilinguals. *Journal of Cognitive Psychology*, *28*(4), 412-426.

Coderre, E. L., Smith, J. F., Van Heuven, W. J., & Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Bilingualism: Language and Cognition*, *19*(3), 471-488.

Coderre, E. L., & Van Heuven, W. J. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in psychology*, *5*, 1070.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59-86.

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

Costa, F., & Guasti, M.T. (2021). Is Bilingual Education Sustainable? *Sustainability*, *13*. https://doi.org/10.3390/su132413766

Costumero, V., Rodríguez-Pujadas, A., Fuentes-Claramonte, P., & Avila, C. (2015). How bilingualism shapes the functional architecture of the brain: A study on executive control in early bilinguals and monolinguals. *Human brain mapping*, *36*(12), 5101-5112.

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.

Damian, M. F., Ye, W., Oh, M., & Yang, S. (2019). Bilinguals as "experts"? Comparing performance of mono- to bilingual individuals via a mousetracking paradigm. *Bilingualism: Language and Cognition*, *22*(5), 1176-1193.

De Bruin, A., Bak, T. H., & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, *85*, 15-26.

de Leeuw, E., & Bogulski, C. A. (2016). Frequent L2 language use enhances executive control in bilinguals. *Bilingualism: Language and Cognition*, *19*(5), 907-913.

Desideri, L., & Bonifacci, P. (2018). Verbal and nonverbal anticipatory mechanisms in bilinguals. *Journal of psycholinguistic research*, *47*(3), 719-739.

Desjardins, J. L., & Fernandez, F. (2018). Performance on auditory and visual tasks of inhibition in English monolingual and Spanish–English bilingual adults: Do bilinguals have a cognitive advantage?. *Journal of Speech, Language, and Hearing Research*, *61*(2), 410-419.

Desjardins, J. L., Bangert, A., & Gomez, N. (2020). What Does Language Have to Do With It? The Impact of Age and Bilingual Experience on Inhibitory Control in an Auditory Dichotic Listening Task. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1581-1594.

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. *Psychonomic bulletin & review*, *26*(4), 1122-1147.

Dunn, A. L., & Tree, J. E. F. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: language and cognition*, *12*(3), 273-289.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *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. *Journal of cognitive neuroscience*, *14*(3), 340-347.

Fernandez, M., Acosta, J., Douglass, K., Doshi, N., & Tartar, J. L. (2014). Speaking two languages enhances an auditory but not a visual neural marker of cognitive inhibition. *AIMS Neuroscience*, *1*(2), 145.

Fernandez, M., Tartar, J. L., Padron, D., & Acosta, J. (2013). Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain and cognition*, *83*(3), 330-336.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189-198.

Friesen, D. C., Luo, L., Luk, G., & Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, cognition and neuroscience*, *30*(3), 238-250.

Garraffa, M., Obregon, M., & Sorace, A. (2017). Linguistic and cognitive effects of bilingualism with regional minority languages: a study of sardinian–italian adult speakers. *Frontiers in psychology*, *8*, 1907.

Gasquoine, P. G. (2016). Effects of bilingualism on vocabulary, executive functions, age of dementia onset, and regional brain structure. *Neuropsychology*, *30*(8), 988.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas-Guasch, N., ... & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from

preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in psychology*, *5*, 11.
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. *Frontiers in Psychology*, *11*, 2398.
Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, *1*(2), 67-81.
Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of cognitive psychology*, *25*(5), 515-530.
Green, D.W., and Kroll, J. F. (2019). "The neurolinguistics of bilingualism," in The Oxford Handbook of Neurolinguistics (Oxford: Oxford University Press), 261.
Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology, 23*(4), 574-586.
Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, 1-23.
Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, *163*, 42-55.
Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition, 23*(2), 283-294.
Guðmundsdóttir, M. D., & Lesk, V. E. (2019). Does the bilingual advantage extend to trilingualism?. *International Journal of Multilingualism*, *16*(4), 549-562.
Hakuta, K. (1986). Mirror of language: The debate on bilingualism. (New York, NY: Basic Books).
Heidlmayr, K., Hemforth, B., Moutier, S., & Isel, F. (2015). Neurodynamics of executive control processes in bilinguals: evidence from ERP and source reconstruction analyses. *Frontiers in Psychology*, *6*, 821.
Heidlmayr, K., Moutier, S., Hemforth, B., Courtin, C., Tanzmeister, R., & Isel, F. (2014). Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioural Stroop Colour Word task. *Bilingualism: Language and Cognition*, *17*(3), 630-645.
Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, *13*(3), 315-325.
Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic bulletin & review*, *18*(4), 625-658.
Hofweber, J. E., Marinis, T., & Treffers-Daller, J. (2020a). Experimentally induced language modes and regular code-switching habits boost bilinguals' executive performance: evidence from a within-subject paradigm. *Frontiers in Psychology*, *11*, 2897.
Hofweber, J., Marinis, T., & Treffers-Daller, J. (2020b). How different code-switching types modulate bilinguals' executive functions: A dual control mode perspective. *Bilingualism: Language and Cognition*, *23*(4), 909-925.
Hui, N. Y., Yuan, M., Fong, M. C. M., & Wang, W. S. Y. (2020). L2 proficiency predicts inhibitory ability in L1-dominant speakers. *International Journal of Bilingualism*, *24*(5-6), 984-998.
Jaimes-Bautista, A. G., Rodriguez-Camacho, M., Martínez-Juárez, I. E., & Rodriguez-Agudelo, Y. (2020). Quantitative and qualitative analysis of semantic verbal fluency in patients with temporal lobe epilepsy. *Neurología (English Edition)*, *35*(1), 1-9.

Kałamała, P., Drożdżowicz, A., Szewczyk, J., Marzecová, A., & Wodniecka, Z. (2018). Task strategy may contribute to performance differences between monolinguals and bilinguals in cognitive control tasks: ERP evidence. *Journal of Neurolinguistics*, *46*, 78-92.

Kavé, G., Kigel, S., & Kochva, R. (2008). Switching and clustering in verbal fluency tasks throughout childhood. *Journal of clinical and experimental neuropsychology*, *30*(3), 349-359.

Kazemeini, T., & Fadardi, J. S. (2016). Executive function: Comparing bilingual and monolingual Iranian university students. *Journal of psycholinguistic research*, *45*(6), 1315-1326.

Keijzer, M. C., & Schmid, M. S. (2016). Individual differences in cognitive control advantages of elderly late Dutch-English bilinguals. *Linguistic Approaches to Bilingualism*, *6*(1-2), 64-85.

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

Kmet, L.M., Lee, R. C., and Cook, L. S. (2011). Standard quality assessment criteria for evaluating primary research papers froma variety of fields. 2004. Edmonton: Alberta Heritage Foundation for Medical Research. 1–22.

Kohnert, K., Ebert, K. D., & Pham, G. T. (2020). *Language disorders in bilingual children and adults*. Plural Publishing.

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop-and Simon-like stimuli. *Psychological research, 56*(3), 130-135.

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society*, *10*(2), 164-172.

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *Quarterly Journal of Experimental Psychology*, *65*(2), 356-369.

Kousaie, S., & Phillips, N. A. (2017). A behavioural and electrophysiological investigation of the effect of bilingualism on aging and cognitive control. *Neuropsychologia*, *94*, 23-35.

Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, *8*, 250.

Kowoll, M. E., Degen, C., Gladis, S., & Schröder, J. (2015). Neuropsychological profiles and verbal abilities in lifelong bilinguals with mild cognitive impairment and Alzheimer's disease. *Journal of Alzheimer's disease*, *45*(4), 1257-1268.

Kramer, R., & Mota, M. B. (2015). Effects of bilingualism on inhibitory control and working memory: A study with early and late bilinguals. *Gragoatá, 38,* 309-331.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of cognitive psychology*, *25*(5), 497-514.

Kuipers, J. R., & Westphal, K. H. (2021). Auditory processing and high task demands facilitate the bilingual executive control advantage in young adults. *Journal of Neurolinguistics*, *57*, 100954.

Kuzyk, O., Friend, M., Severdija, V., Zesiger, P., & Poulin-Dubois, D. (2020). Are there cognitive benefits of code-switching in bilingual children? A longitudinal study. *Bilingualism: Language and Cognition*, *23*(3), 542-553.

Lange, K., Kühn, S., & Filevich, E. (2015). " Just Another Tool for Online Studies"(JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS one*, *10*(6), e0130834.

Lee Salvatierra, J., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*, *15*(1), 26-37.

Lehtinen, N., Luotonen, I., & Kautto, A. (2021). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult*, 1-13.

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*(4), 394.

Leivada, E., Westergaard, M., Duñabeitia, J. A., and Rothman, J. (2020). On the phantom-like appearance of bilingualism effects on neurocognition:(How) should we proceed? *Bilingualism: Language and Cognition*, 1–14. doi: 10.1017/S1366728920000358

Li, X., Ng, K. K., Wong, J. J. Y., Lee, J. W., Zhou, J. H., & Yow, W. Q. (2021). Bilingual language entropy influences executive functions through functional connectivity and signal variability. *Brain and Language*, *222*, 105026.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., G_tzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6:e1000100. doi: 10.1371/journal.pmed.1000100

Ljungberg, J. K., Elbe, P., & Sörman, D. E. (2020). The bilingual effects of linguistic distances on episodic memory and verbal fluency. *Scandinavian journal of psychology*, *61*(2), 195-203.

Ljungberg, J. K., Hansson, P., Andrés, P., Josefsson, M., & Nilsson, L. G. (2013). A longitudinal study of memory advantages in bilinguals. *PloS one*, *8*(9), e73029.

Luk, G., Anderson, J. A., Craik, F. I., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and cognition*, *74*(3), 347-357.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29-41.

Marsh, J. E., Hansson, P., Sörman, D. E., & Ljungberg, J. K. (2019). Executive Processes Underpin the Bilingual Advantage on Phonemic Fluency: Evidence From Analyses of Switching and Clustering. *Frontiers in psychology*, *10*, 1355.

Marzecová, A., Asanowicz, D., Kriva, L. U., & Wodniecka, Z. (2013). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition*, *16*(3), 608-623.

Massa, E., Köpke, B., & El Yagoubi, R. (2020). Age-related effect on language control and executive control in bilingual and monolingual speakers: Behavioral and electrophysiological evidence. *Neuropsychologia*, *138*, 107336.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314-324.

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual. *Handbook of bilingualism: Psycholinguistic approaches*, 389-407.

Ministero della Pubblica Istruzione (MIUR). (1998). *Archivio delle news - Anno 1998*. Archivio dell'area istruzione. Retrieved January 14, 2022, from https://archivio.pubblica.istruzione.it/news/1998/seconda.shtml#:~:text=A%20partire%20dal%201962%2C%20anno,ore%20di%20insegnamento%20per%20anno.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8-14.

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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, *6*(7), e1000097.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: event related potential (ERP) evidence for experience-specific effects. *PloS one*, *9*(4), e94169.

Morrison, C., Farooq, K., & Taler, V. (2019). The influence of bilingualism on working memory event-related potentials. *Bilingualism*, *22*(1), 191-199.

Morrison, C., Kamal, F., Le, K., & Taler, V. (2020). Monolinguals and bilinguals respond differently to a delayed matching-to-sample task: An ERP study. *Bilingualism: Language and Cognition*, *23*(4), 858-868.

Nair, V. K., Biedermann, B., & Nickels, L. (2017). Effect of socio-economic status on cognitive control in non-literate bilingual speakers. *Bilingualism: Language and cognition*, *20*(5), 999-1009.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353-383.

Newman, J. P., & Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *Journal of abnormal psychology*, *95*(3), 252.

Novelli, G., Papagno, C., Capitani, E., & Laiacona, M. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su sogetti normali. *Archivio di psicologia, neurologia e psichiatria*.

Okada, K., He, G., & Gonzales, A. (2019). Monolinguals and bilinguals differ in performance on the Taboo Stroop task. *The Open Psychology Journal*, *12*(1).

Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From Bilingualism to Bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition*, *21*(4), 867-879.

Ouzia, J., Bright, P., & Filippi, R. (2019). Attentional control in bilingualism: An exploration of the effects of trait anxiety and rumination on inhibition. *Behavioral Sciences*, *9*(8), 89.

Paap, K. (2019). The bilingual advantage debate: Quantity and quality of the evidence. *The handbook of the neuroscience of multilingualism*, 701-735.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, *66*(2), 232-258.

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81-93.

Paap, K. R., Johnson, H. A., and 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. doi: 10.1016/j.cortex.2015.04.014

Paap, K. R., Mason, L. A., Zimiga, B. M., Ayala-Silva, Y., Frost, M. M., Gonzalez, M., & Primero, L. (2019). Other Language Proficiency Predicts Unique Variance in Verbal Fluency Not Accounted for Directly by Target Language Proficiency: Cross-Language Interference?. *Brain Sciences, 9*(8), 175.

Paplikar, A., Alladi, S., Varghese, F., Mekala, S., Arshad, F., Sharma, M., ... & ICMR-NCTB Consortium. (2021). Bilingualism and Its Implications for Neuropsychological Evaluation. *Archives of Clinical Neuropsychology*.

Patra, A., Bose, A., & Marinis, T. (2020). Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition, 23*(1), 204-218.

Peal, E., and Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs General and Applied, 76*, 1–23. doi: 10.1037/h0093840

Pekkala, S., Goral, M., Hyun, J., Obler, L. K., Erkinjuntti, T., & Albert, M. L. (2009). Semantic verbal fluency in two contrasting languages. *Clinical linguistics & phonetics, 23*(6), 431-445.

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 313.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, *22*(3), 415-422.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society: JINS*, *17*(4), 682.

Rainey, V. R., Stockdale, L., Flores-Lamb, V., Kahrilas, I. J., Mullins, T. K. L., Gjorgieva, E., ... & Silton, R. L. (2021). Neural differences in the temporal cascade of reactive and proactive control for bilinguals and monolinguals. *Psychophysiology*, e13813.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Standard progressive matrices*. Oxford: Oxford Psychologists Press.

Rhys, M., & Thomas, E. M. (2013). Bilingual Welsh–English children's acquisition of vocabulary and reading: implications for bilingual education. *International Journal of Bilingual Education and Bilingualism*, *16*(6), 633-656.

Rodrigues, L. R., & Zimmer, M. C. (2016). Inhibitory and attentional control: the interaction between "professional activity" and bilingualism. *Psicologia: Reflexão e Crítica*, *29*.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, *2*, 388.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied neuropsychology*, *7*(1), 17-24.

Rosselli, M., Ardila, A., Lalwani, L. N., & Vélez-Uribe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *19*(3), 489-503.

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., LUIS, M., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience*, *112*(6), 759-776.

Sadat, J., Pureza, R., & Alario, F. X. (2016). Traces of An Early Learned Second Language in Discontinued Bilingualism. *Language Learning*, *66*(S2), 210-233.

Sala, A., Malpetti, M., Farsad, M., Lubian, F., Magnani, G., Frasca Polara, G., ... & Perani, D. (2021). Lifelong bilingualism and mechanisms of neuroprotection in Alzheimer dementia. *Human brain mapping*.

Sanchez-Azanza, V. A., López-Penadés, R., & Adrover-Roig, D. (2020). More similitudes than differences between bilinguals and monolinguals on speeded and demand-varying executive tasks. *Language, Cognition and Neuroscience*, *35*(8), 992-1009.

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism*, *13*(2), 231.

Schroeder, S. R., & Marian, V. (2017). Cognitive consequences of trilingualism. *International Journal of Bilingualism*, 21(6), 754-773.

Seçer, I. (2016). Skills of cognitive flexibility in monolingual and bilingual younger adults. *The Journal of general psychology*, *143*(3), 172-184.

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

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In *Advances in psychology* (Vol. 65, pp. 31-86). North-Holland.

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99-105.

Simson, R., Vaughan Jr, H. G., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual Go/NoGo tasks. *Electroencephalography and clinical neurophysiology*, *43*(6), 864-875.

Soares, S. M. P., Ong, G., Abutalebi, J., Del Maschio, N., Sewell, D., & Weekes, B. (2019). A diffusion model approach to analyzing performance on the Flanker task: The role of the DLPFC. *Bilingualism: Language and Cognition*, *22*(5), 1194-1208.

Soltani, M., Moradi, N., Rezaei, H., Hosseini, M., & Jasemi, E. (2021). Comparison of verbal fluency in monolingual and bilingual elderly in Iran. *Applied Neuropsychology: Adult*, *28*(1), 80-87.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.
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*(2), 401-415.
Taler, V., Johns, B. T., Young, K., Sheppard, C., & Jones, M. N. (2013). A computational analysis of semantic structure in bilingual verbal fluency performance. *Journal of Memory and Language*, *69*(4), 607-618.
Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: the role of age of acquisition. *Frontiers in psychology*, *2*, 123.
Tao, L., Taft, M., & Gollan, T. H. (2015). The bilingual switching advantage: Sometimes related to bilingual proficiency, sometimes not. *Journal of the International Neuropsychological Society: JINS*, *21*(7), 531.
Tao, L., Wang, G., Zhu, M., & Cai, Q. (2021). Bilingualism and domain-general cognitive functions from a neural perspective: A systematic review. *Neuroscience & Biobehavioral Reviews, 125*, 264-295.
Thompson-Reuters. (2012). *Web of science*. Retrieved 29 June 2021 from http://www.isiknowledge.com
Treffers-Daller, J., Ongun, Z., Hofweber, J., & Korenar, M. (2020). Explaining individual differences in Executive Functions performance in multilinguals: the impact of code-switching and alternating between Multicultural Identity Styles. *Frontiers in Psychology*, *11*.
Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *neuropsychology*, *11*(1), 138.
Van den Noort, M., Vermeire, K., Bosch, P., Staudte, H., Krajenbrink, T., Jaswetz, L., ... & Lim, S. (2019). A systematic review on the possible relationship between bilingualism, cognitive decline, and the onset of dementia. *Behavioral Sciences*, *9*(7), 81.
Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*, 40-46.
Vivas, A. B., Ladas, A. I., Salvari, V., & Chrysochoou, E. (2017). Revisiting the bilingual advantage in attention in low SES Greek-Albanians: does the level of bilingual experience matter?. *Language, Cognition and Neuroscience*, *32*(6), 743-756.
Waldie, K. E., Badzakova-Trajkov, G., Miliivojevic, B., & Kirk, I. J. (2009). Neural activity during Stroop colour-word task performance in late proficient bilinguals: A functional Magnetic Resonance Imaging study. *Psychology & Neuroscience*, *2*(2), 125.
Warmington, M. A., Kandru-Pothineni, S., & Hitch, G. J. (2019). Novel-word learning, executive control and working memory: A bilingual advantage. *Bilingualism: language and Cognition*, *22*(4), 763-782.
Wei, L. (Ed.). (2000). *The bilingualism reader* (Vol. 11). London: Routledge.
Wong, B., Yin, B., & O'Brien, B. (2016). Neurolinguistics: Structure, function, and connectivity in the bilingual brain. *BioMed research international*, 2016.
Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1579.
Woumans, E., Van Herck, S., & Struys, E. (2019). Shifting gear in the study of the bilingual advantage: Language switching examined as a possible moderator. *Behavioral Sciences*, *9*(8), 86.
Xie, Z., & Dong, Y. (2017). Contributions of bilingualism and public speaking training to cognitive control differences among young adults. *Bilingualism: Language and Cognition*, *20*(1), 55-68.
Xie, Z., & Zhou, S. (2020). Bilingualism, demographics, and cognitive control: a within-group approach. *Frontiers in psychology*, *11*, 94.
Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology*, *146*, 121-136.
Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in psychology*, *2*, 309.

Zarino, B., Crespi, M., Launi, M., & Casarotti, A. (2014). A new standardization of semantic verbal fluency test. *Neurological Sciences*, *35*(9), 1405-1411.

Zeng, Z., Kalashnikova, M., & Antoniou, M. (2019). Integrating bilingualism, verbal fluency, and executive functioning across the lifespan. *Journal of Cognition and Development*, *20*(5), 656-679.

Zhang, H., Wu, Y. J., & Thierry, G. (2020). Bilingualism and aging: A focused neuroscientific review. *Journal of Neurolinguistics*, *54*, 100890.

Zhou, B., & Krott, A. (2018). Bilingualism enhances attentional control in non-verbal conflict tasks–evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition*, *21*(1), 162-180.