The Role of Bilingualism in Executive Control, Working Memory, and Episodic Memory: Evidence from Polish-English adults

by:
Marta Ciesielska

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

The University of Sheffield
Health Sciences School
Division of Human Communication Sciences

March 2022

Knowledge of languages is the doorway to wisdom

Roger Bacon

Abstract

This study sought to examine how bilingualism modifies long-term memory mechanisms, specifically episodic memory as measured by paired associate learning and verbal fluency. Additionally, it investigated whether there is a bilingual advantage in executive control and working memory which has been inconsistently found in previous studies (i.e., Bialystok, Craik, & Luk, 2012; but see Paap & Greenberg, 2013). A systematic review was conducted as a part of this project with 22 eligible studies included in the meta-analytic considerations.

Nineteen monolingual English adults and 28 Polish-English bilinguals were recruited to participate in three experiments. To assess their performance on episodic memory, a paired associate learning task was developed and piloted to ensure its integrity and reliability as an episodic memory measure.

Results obtained in this thesis are largely inconsistent with the results obtained from the systematic review. Working memory but not executive control bilingual advantage was found in the meta-analysis. No bilingual advantage was found for either working memory or executive control in Experiment One. The meta-analysis revealed a bilingual advantage in episodic memory for recall and recognition, but this was not replicated in Experiment Two. Finally, Experiment Three did not find a bilingual advantage in episodic memory as measured by verbal fluency which is consistent with research findings concerning low proficiency bilinguals. Indeed, performance on episodic memory and executive control is constrained by level of bilingualism and frequency of English usage.

Although no bilingual advantage was found in this study, there seem to be organisational differences in cognitive mechanisms between bilinguals and monolinguals. Also, it is important for future studies to ensure that bilingual sample is clearly defined and of the second language proficiency level that enables capturing a potential bilingual advantage in episodic memory, working memory, and executive control. The results are further discussed from the perspective of the findings from the systematic review of the available literature.

Acknowledgments

First and foremost, I would like to thank my amazing main supervisor, Dr Meesha Warmington, whose guidance and support helped me develop as a researcher. I genuinely could not have asked for a better mentor – Meesha, I am wholeheartedly grateful for the experience of being your PhD student.

I would also like to thank my second supervisor, Dr Jenny Thomson, for constructive feedback and encouragement throughout my PhD adventure.

Thank you to my fellow PhD students at Human Communication Sciences, especially Indri, Deema, Tanja, Alex, Sam, and Philip, for words of support and sharing the ups and downs of our PhD adventures.

To my parents, Henryka and Leszek; thank you for instilling passion for learning and always believing in me. To my siblings: Artur, Iza, and Piotr, for your continuous support whenever necessary.

I dedicate this thesis to my husband Richard and my little boy Leo for their infinite patience and support. Thank you for being there for me, I love you always.

Author's Declaration

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication. I declare that this thesis has not been submitted for a higher degree to any other University or institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in this thesis, published or otherwise, are fully acknowledged in accordance with standard referencing practices.

Presentations

The experiments reported in this thesis have resulted in the following presentations:

1) Ciesielska, M., Warmington, M. & Thomson, J. (2018). Does L1 vocabulary knowledge influence L2 lexical learning? Existing lexical knowledge in L1 shapes the ability to learn new words in L2. Poster presented at the Human Communication Sciences conference held at the University of Sheffield, July 2-3, 2018.

2) Ciesielska, M., Warmington, M. & Thomson, J. (2018). Does L1 vocabulary knowledge influence L2 lexical learning? Poster presented at Child Language Symposium held at the University of Sheffield, July 10-12, 2019.

List of Tables

List of Figures

Table of Contents

Introduction

Owing to globalisation, more and more people around the world can now be considered bilingual or multilingual. With a rising number of individuals able to communicate in two or more languages, there has been a growing interest in the role of bilingualism in modifying executive control and working memory. Indeed, research findings have accumulated empirical evidence in support of a bilingual advantage compared to monolinguals in tasks that measure memory and goal-directed behaviour or executive control (Kovacs & Mehler, 2009; Poarch & van Hell 2012; Pelham & Abrams, 2014; Bialystok, Craik, & Freedman, 2007; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012). However, there is also research that has failed to find bilingual advantage in these aspects (De Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2015; Polderman, Boomsma, & de Geus; 2005), and thus suggests that executive control and memory may not be in fact enhanced in bilinguals. Thus, there still remains little consensus concerning the role bilingualism plays in executive control and memory, and the extent to which these are interconnected in bilingual populations. Interestingly, executive control processes are not only linked to short-term memory, but they are also implicated in long-term memory mechanisms such as the ability to integrate and retrieve novel associations from long-term memory over time (as evidenced for example in paired associate learning and verbal fluency tasks). Research to date has focused on bilingual influence on executive control and working memory mechanisms. However, there is still very limited understanding of how bilingualism impacts long-term memory mechanisms.

Thus, this thesis aims to enhance the current level of understanding of the mechanisms underlying the bilingual advantage. More specifically, it investigates the impact of bilingualism on long-term memory, specifically episodic memory (as measured by paired associate learning and verbal fluency tasks). The thesis comprises four Literature Review chapters providing theoretical, empirical, and meta-analytic evidence and a chapter presenting rationale for further considerations. This is followed by a chapter which covers methodological considerations of the study and leads to three results chapters offering analysis and interpretation of findings obtained in experiments conducted with monolingual and bilingual adult participants. The thesis is concluded with the final chapter comprising general discussion of the key findings, strengths and limitations of the thesis, and final conclusions.

Chapter One provides a theoretical background to further considerations presented in this thesis. Thus, it reviews the literature concerning executive control and memory. Various relevant theoretical frameworks and research contributions in the field of executive control and working memory are presented in more detail. Next, conceptualisations and explanations of executive control as well as presentation of the individual differences within this domain are discussed. This is followed by presentations of several attempts to define memory along with the major theoretical frameworks. The Chapter then moves onto working memory models which are of the main importance in the context of this thesis. Here, the main models and their evaluations are also presented.

Chapter Two constitutes an introduction to the literature review regarding long-term memory mechanisms, the paired associate learning paradigm, and verbal fluency. In this Chapter, long-term memory processes are introduced, and the type of memory further emphasised here is episodic memory. Then, the paired associate learning paradigm is defined and research findings concerning associative learning are elaborated on. It presents and discusses relevant theoretical frameworks and research contributions in the field of paired associate learning in more detail. The Chapter closes with a presentation and discussion regarding verbal fluency.

Chapter Three offers a review of the literature concerning the current state of knowledge in the field of bilingualism, executive control, and working memory. It provides an introduction to various theoretical frameworks of bilingual language control and discusses three key models. Then, the Chapter continues with theoretical accounts of the bilingual advantage in executive control. Subsequently, empirical evidence regarding bilingual advantage in executive control as well as memory is given. The Chapter also covers key factors found to modulate the presence of bilingual advantage in research findings, and these are also discussed in more detail.

Chapter Four is a systematic review chapter that provides cumulative evidence regarding the bilingual advantage in episodic memory across the lifespan and across the languages. The summary of results from 22 eligible studies are included in this section. As this review is limited by substantial heterogeneity across studies, the results from conducted meta-analyses should be treated with caution and may be found difficult to interpret.

Chapter Five presents the rationale, research questions, and the thesis overview. Additionally, an overview of three experimental chapters is included as well as ethical considerations.

Chapters Six to Nine provide results of three main experiments conducted with monolingual and bilingual participants. Thus, they discuss findings in relation to bilingual advantage hypothesis in executive control (i.e., inhibition/attentional control), working memory, and episodic memory as measured by paired associate learning and verbal fluency.

Chapter Ten is the final chapter that provides a summary of main experimental findings along with their contextualisation within bilingual literature. In this Chapter, strengths and limitations of the study are also discussed and final remarks made.

CHAPTER ONE

Executive Control and Working Memory

1.1 Conceptualising Executive Control

According to Miyake, Friedman, Emerson, Witzki, Howerter, and Wager (2000), *executive control*, (i.e., *set-shifting*, *memory updating*, and *inhibition/attentional control*) refers to a set of higher-order executive control processes fundamental to maintaining and executing more complex tasks as well as regulating thoughts and behaviours. Research has demonstrated that executive control processes are multifaceted and various types of executive control are in fact separable although correlated (Friedman, Miyake, Young, DeFries, Corley, & Hewitt, 2008). Executive control processes have been claimed to play a crucial role in self-control and self-regulations in terms of one's cognition and actions and as such have a significant impact on many aspects of everyday life. Indeed, there are lots of individual variations in our behaviours and actions. Individuals possess various abilities to maintain their thoughts and behaviours – some can deal successfully with resisting temptations while for others this is much more troublesome and effortful. This is in line with the seminal paper by Miyake and colleagues (2000) who examined individual differences in executive control processes and derived the general conclusions: the concept of unity and diversity within executive control, substantial genetic contribution, and developmental stability.

The proceeding sections will begin by describing each of the processes (i.e., set-shifting, memory updating, and inhibition/attentional control) separately and discuss the framing conclusions of individual differences in executive control in more detail.

1.1.1 Set-Shifting

A relatively common definition of set-shifting refers to the ability to direct one's attention to the stimuli of interest (Diamond, 2013) and to concurrently take into consideration conflicting representations to execute relevant behaviour (Jacques & Zelazo, 2005). In other words, it is the ability to shift between one's cognitive resources to adjust to environmental demands (such as rule changes) (Dennis & Vander, 2010).

Set-shifting is therefore analogous to the concept of being able to flexibly switch between mental sets (i.e., cognitive flexibility; Ionescu, 2012. For a complete account of set-shifting and its neural underpinnings see Aron, 2008). Traditionally, set shifting can be assessed using tasks which measure participants' ability to efficiently switch between two mental sets. The difference between a set of items that requires them to switch between rules flexibly and a set that requires them to consistently use a rule (Vandierendonck, Liefooghe, & Verbruggen, 2010). The difference between these sets results in *switch cost*. The switch cost stems from the interplay between the time required for mental set reconfigurations and the time necessary for interference resolution (resulting from the previous set). The fact that the ability to shift takes time to develop manifests itself in studies with younger children whose performance on set-shifting tasks is indeed error prone (Crone, Bunge, van der Molen, & Ridderinkhof, 2006). In adult participants, the switch cost is noticeable immediately once the task has been switched (Monsell, 2003).

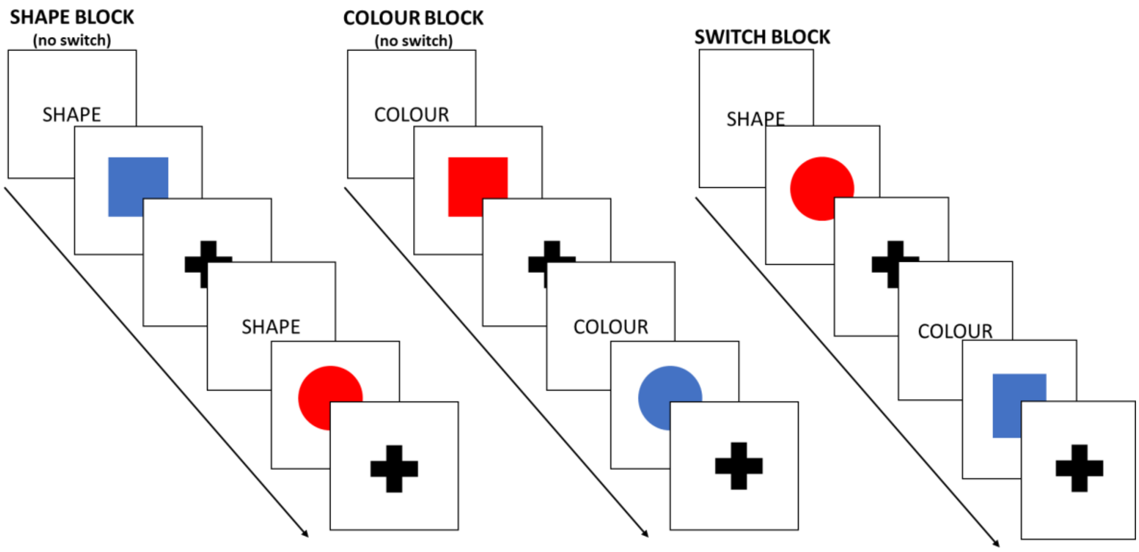


Figure 1. 1 An example of a colour-shape task (Miyake & Friedman, 2012)

Having introduced the concept of set-shifting, it is important to present some commonly accepted measures that assess its aspects. One set of tasks used in executive control research constitute set-shifting tasks (e.g., Prior & Gollan, 2011; Wiseheart, Viswanathan, & Bialystok, 2016). These tasks measure one's ability to switch between information or mental sets. They comprise switch and no-switch trials: in the former, participants are required to shift between one type of information to another in order to provide a correct response, whereas no-switching between information types is needed in the latter.

A typical task that measures set-shifting is the Colour-Shape task (Miyake & Friedman, 2012). In this task, participants are presented with a colour-shape combination (red circle) and are asked to identify the shape (task 1) or the colour (task 2). The Colour-Shape task (Figure 1.1) is an example of a set-shifting task: it comprises shapes that are presented in various colours. Here, for instance, in the switch trial participant's task is to name either the shape or the colour, while the prior trial required the opposite (naming the shape once the name of colour was required in the prior trial) or vice versa. In a non-switch trial on the other hand, the name of shape or the colour is again required and the preceding trial requires the same response.

Apart from this universal setup, there is also a linguistic version of the task (Prior & MacWhinney, 2010). In this version, instead of shapes and colours, two languages are used and participants' native and their second language are engaged.

1.1.2 Memory Updating

Memory updating refers to the process of maintaining and updating information based on its significance which also includes constant removal of no longer relevant information and replacing it with available relevant information (Baddeley, 2000). In other words, memory updating enables changes in a part of memory representation while leaving the rest of the representation unaltered (Artuso & Palladino, 2019). It is a flexible process of storing and updating content that allows combining already available mental content with newly available representations (Baddeley, 2000). Interestingly, it has been claimed to be the only of the three major executive control functions to reliably predict *fluid intelligence* (Chen & Li, 2007; but see Ecker, Lewandowsky, & Oberauer, 2014 for criticism) that is reasoning ability (Carroll, 1993).

Memory updating is typically assessed using an *n*-back task (Kircher, 1958). In this task, participants are presented with a sequence of stimuli and asked to indicate when a stimulus displayed matches the one shown *n* trials earlier in the sequence (Figure 1.2).

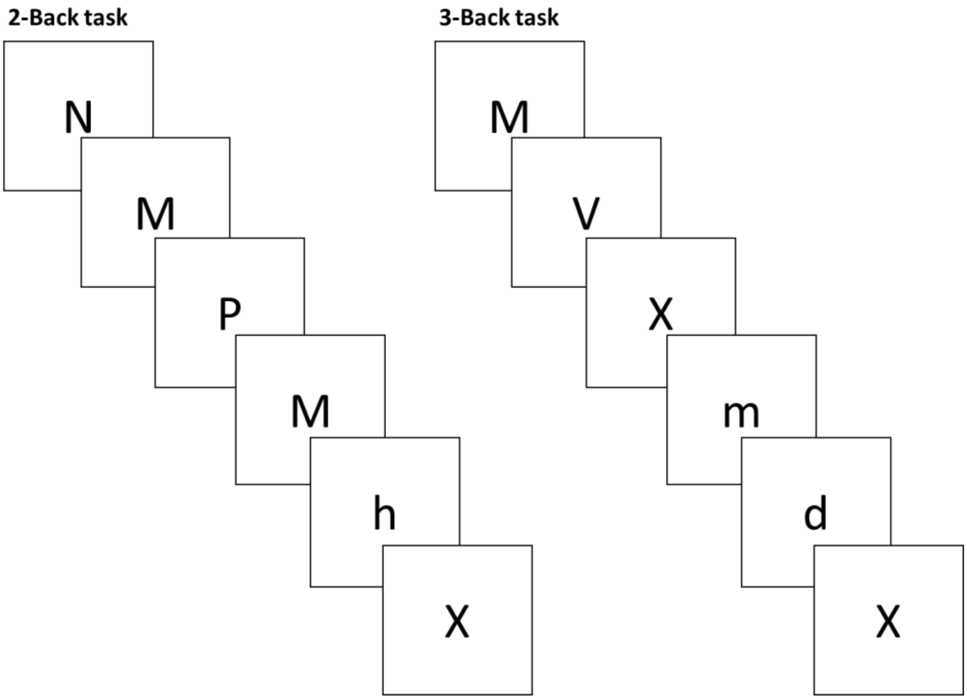


Figure 1. 2 An example of an *n*-back task (Kircher, 1958)

However, it has been questioned in terms of the extent to which memory updating tasks, for example an *n*-back task, additionally measure working memory (i.e., the ability to store and manipulate information for a task at hand, cf. Baddeley & Hitch, 1974) apart from executive control. Indeed, experimental tasks do not provide a measure of updating exclusively as they also engage other working memory processes such as memory storage and maintenance of information for instance.

Several previous studies propose that memory updating tasks are arguably a reliable measure of general working memory ability and thus can be employed to assess working memory (Chuderski, Taradaay, Necka, & Smolen, 2012). It is thus not entirely clear whether the correlation between memory updating and fluid intelligence as found by Chen and Li (2007), represents a true existing relation between the two or whether it reflects a relation between one's intelligence and their working memory capacity. Additionally, Schmiedek, Hildebrandt, Lovden, Wilhelm, and Lindenberger (2009) have offered additional evidence towards a strong connection between working memory updating and working memory while explaining that both are underpinned by common working memory abilities and thus these abilities impact individuals' performance.

Other studies have questioned the strong relation between updating and working memory (Radvansky & Copeland, 2001) offering an explanation that the relationship between situation model processing (in other words, complex mental representations that are involved in simulations of events presented in a text) and working memory is relatively weak. Additionally, Radvansky and Dijkstra (2007) also argued that memory updating stays relatively unaltered and preserved in old age while this cannot be said about other cognitive processes such as smaller working memory capacity (Craik & Byrd, 1982). Several age-related effects have been proposed to impact executive control processes, such as, worse state of health, arguably less formal schooling, as well as fewer opportunities for cognitive practice compared to younger adults. Indeed, one's physical strength and agility tend to decline while they get older as do the physiological and nervous system of the body as well as cognition (Craik & Byrd, 1982). A decline in working memory capacity has been proposed to stem from a more general cognitive decline concerning one's ability to suppress prepotent but irrelevant processing (Hasher & Zacks, 1988) and it often results in older people's difficulties in managing large amount of information (Zacks, Hasher, & Li, 2000).

1.1.3 Inhibition/Attentional Control

Inhibition/attentional control refers to an individual's ability to deliberately inhibit dominant and automatic responses when needed (Miyake et al., 2000). Inhibition/attentional control can be thought of as a unitary construct, but it has also been proposed to be composed of multiple components. Paap, Anders-Jefferson, Zimiga, Mason, and Mikulinsky (2020) proposed several hypothetical inhibition/attentional control factors including *resistance to proactive inhibition, resistance to distractor interference*, *inhibition of prepotent responses*, and *behavioural inhibition* that are present in non-verbal interference tasks. Resistance to proactive inhibition, one's ability to deal with intrusions from the information that was relevant in the previous task but is no longer relevant, has been proposed to be considered as a separate factor (Friedman & Miyake, 2004; Pettigrew & Martin, 2014). The issue of much disagreement relates to the concept of resistance to distractor interference (concerning stimulus-stimulus conflict resolution) and whether it is distinguishable from inhibition of prepotent responses (in stimulus-response conflict resolution). Indeed, the former describes an executive control mechanism that diminishes the competition between relevant and irrelevant task information, whereas the latter diminishes the competition between conflicting responses. As research findings suggest, these two are in fact correlated but distinct inhibitory forms (Friedman & Miyake, 2004). However, based on further studies by Rey-Mermet, Gade, and Oberauer

(2018), they suggested that "the non-verbal tasks used to assess inhibition/attentional control do not measure a common underlying construct but instead measure the highly task-specific ability to resolve the interference arising in each task" (Paap, Anders-Jefferson, Mason, Alvarado & Zimiga, 2018, p. 3). Thus, "[…] inevitable implication is that studies using a single laboratory paradigm for assessing or investigating inhibition do not warrant generalisation beyond the specific paradigm studied" (Rey-Mermet et al., 2018, p. 515).

Inhibition/attentional control is typically measured using the Antisaccade task, Simon task, Flanker task, Stop-Signal Reaction Time task, Stroop task, or Attention Network Task.

Antisaccade task (Roberts, Hager, & Heron, 1994) begins with a fixation cross being displayed in the middle of the screen for different time lengths across trials. It is then followed by a visual cue (a black square) on the side of the screen and a target object presentation (arrow inside a square) on the opposite side. Here, participants are required to decide on the direction of the arrow and react by a button press. The target object appears for only a short time before it is masked and thus participants have to inhibit their response of looking at the visual cue (the black square) to correctly identify the direction of the target. The dependent variable is the proportion of correct responses.

Simon task (Lu & Proctor, 1995) is typically used to assess the ability to resolve conflict by inhibiting a motor response. Here, coloured stimuli that are displayed either on the left or right side of the screen are associated with a left or right key press. The task comprises a set of trials in which the target response feature is either aligned with the target stimulus on the screen (i.e., congruent) or it is not (i.e., incongruent). In congruent trials (i.e., where both colour and position converge on the same response), the correct key press corresponds to the position of the stimulus on the screen. In incongruent trials (i.e., where the correct key and stimulus position conflict), participants need to ignore the position because the correct response is determined only by the colour of the stimulus. Additionally, the *Simon effect* (i.e., the difference in participant's accuracy (or reaction times) between congruent and incongruent trials) indicates how efficient participants are to resolve conflict by inhibiting irrelevant responses.

Flanker task (Eriksen & Eriksen, 1974) assesses the ability to inhibit irrelevant information and selectively attend to the direction of a central target while ignoring the congruent or incongruent distracters. The target, always displayed in the centre of the screen, constitutes the focal point of attention for participants. In the congruent trials, the distractors point in the same direction as the target, whereas in the incongruent trials the distractors point in the opposite direction. Participants are instructed to ignore the orientation of the distractors in both congruent and incongruent trials and respond to the direction of the target as quickly and accurately as possible by a button press. Additionally, the *Flanker effect* (i.e., the difference in participant's accuracy (or reaction times) between congruent and incongruent trials) indicates how efficient participants are to resolve conflict by inhibiting irrelevant information.

Stop-Signal Reaction Time task (Logan, 1994) is a standard two-choice reaction task in which a stop-signal is activated occasionally and unpredictably to participants, and they need to inhibit the response to the choice signal. For instance, in the first block of the Stop-Signal Reaction Time task, participants are presented with words and their task is to assign the words to an animal or a non-animal category correctly as quickly as possible. In the second block, they are asked to continue with word categorisation unless they hear a tone on several randomly chosen trials which requires them to inhibit their response. The dependent variable is a proportion of correct responses. Logan and Cowan (1984) introduced the idea of the horse-race model that enables a quantitative interpretation of individual performance on the Stop-Signal task. The model assumes that both, the *go* processes and the *stop* processes, compete with each other to reach the finishing line first. Thus, if the *stop* processes finish before the *go*, it can be assumed that the response is inhibited. If the *go* processes finish before the *stop* processes, then the response escapes from inhibitory control and is executed.

Stroop task (Stroop, 1935) is often used in both academic and clinical research (Cohen, Dunbar, & McClelland, 1990) and it is an example of one of the earliest measures of inhibition/attentional control (Posner & Snyder, 1975). In the classic Stroop task, participants are asked to focus on one dimension of the target (such as the colour of the target word presented on the screen) while ignoring the other dimension (such as the word itself for instance). To illustrate this example, in Stroop tasks the subjects are shown coloured word stimuli and are asked to name the colour of the written word while suppressing the superior tendency to read the target word. In other words, participants are instructed to name the colour of the target word (i.e., target word is GREEN and it is displayed using blue font). Here, the subject needs to attend to the target word GREEN while disregarding its colour, blue, which is a competing dimension in this case. When the stimulus is incongruent (meaning that the print word and word colour do not match, i.e., the word GREEN displayed in blue), participants' performance has been found to be slower and less accurate as opposed to congruent trials (word GREEN is displayed in green). This so-called congruency effect reflects participants' inability to completely disregard distractors despite their willingness to do so.

Attention Network Task (Fan, McCandliss, Sommer, Raz, & Posner, 2002) is a computerised task that combines attentional and spatial cues with a Flanker task (Figure 1.3).

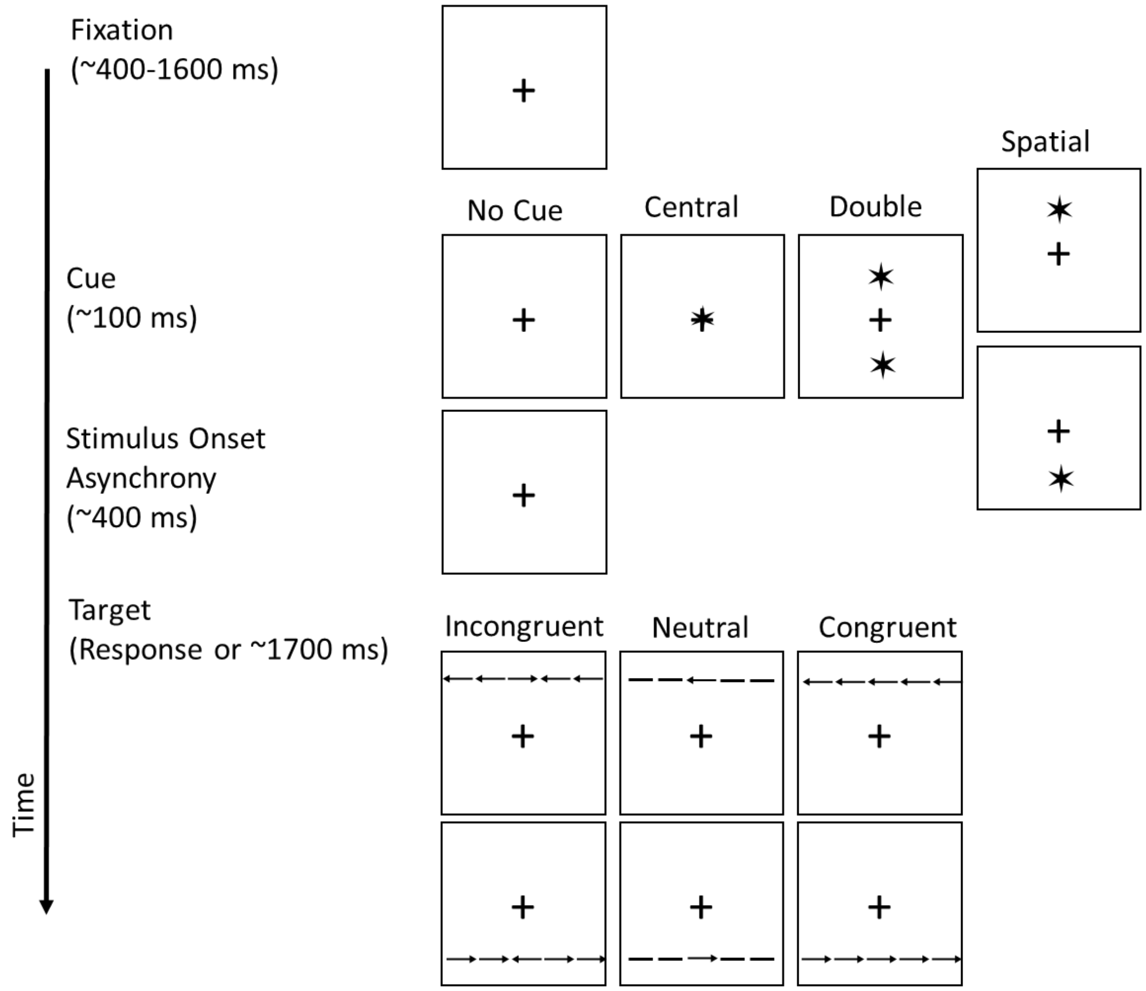


Figure 1. 3 An example of an Attention Network Task (adapted from Arora, Lawrence, & Klein, 2020)

At the beginning of every trial, a spatial cue is displayed followed by a presentation of the target arrows individually in an array of five. Participants are asked to decide on directionality of the target, the central arrow, as quickly as possible. The preceding cue can be non-existent, displayed centrally, a double cue (that is a prediction of an upcoming target presentation) or a spatial cue (that provides information about the location of the target).

1.2 Individual Differences in Executive Control

1.2.1 Unity and Diversity

Miyake et al. (2000) examined individual differences in executive control, i.e., set-shifting, memory updating, inhibition/attentional control, in a sample of 137 college students, by utilising multiple tasks to tap into each of the processes. Using the confirmatory factor analysis, they identified a three-factor model of executive control which illustrated commonality as well as disparity (i.e., unity and diversity, see Figure 1.4).

The notion of unity means that there is a *shared feature* (common across set-shifting, memory updating, and inhibition/attentional control) whereas diversity is about a *unique feature* that is indeed specific to each of the executive control processes (Miyake & Friedman, 2012).

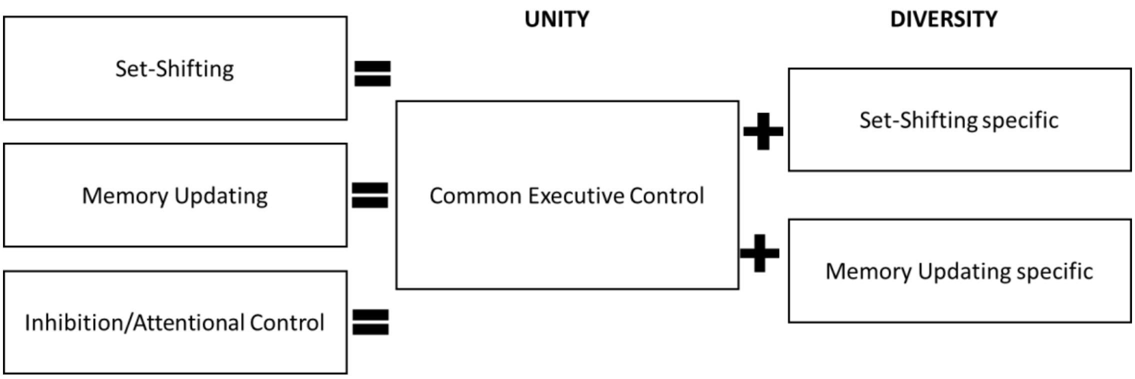


Figure 1. 4 Unity and Diversity framework by Miyake and Friedman (2012)

Indeed, individual differences in executive control can be characterised by both unity and diversity notions, namely, different executive control skills correlate with each other (and thus they share a common ability – unity notion), but they share some separate features at the same time (thus diversity notion). This unity-diversity notion has also been supported by neuroimaging studies pointing to brain localisation (Collette, Van der Linden, Laureys, Delfiore, Degueldre, Luxen, & Salmon, 2005; Sylvester, Wagner, Lacey, Hernandez, Nichols, Smith, & Jonides, 2003). For instance, Collette et al. (2005) discovered not only the common frontal and parietal brain areas where activation was shared by all three executive control processes, but also frontal and/or posterior brain areas where activation was unique to two functions, namely, set-shifting and memory updating. Research conducted by Hedden and Gabrieli (2010) also indicated that both inhibition/attentional control and set-shifting are strongly associated with a number of brain areas (i.e., dorsolateral prefrontal cortex, anterior cingulate cortex, and basal ganglia). Certain other areas observed more activation for inhibiting then they demonstrated for set-shifting. Interestingly, there was very little evidence of brain areas which received more activation for set-shifting than they received for inhibition/attentional control. Thus, this study provided further evidence to the shared brain areas associated with executive control. Moreover, it also sheds more light on the diversity of these functions.

Similarly, the evidence for the existence of unity and diversity notions has been provided by studies with preadolescent children. Rose, Feldman, and Jankowski (2012) tested 134 participants at the ages of 7, 12, 24, and 36 months on a series of measures assessing their information processing abilities (inhibition/attentional control, memory, and processing speed). These three abilities were then used as predictors of executive control at 11 years of age. The results revealed that in fact early processing abilities influence executive control skills in preadolescence. Structural equation modelling indicated that information processing ability in infants and toddlers accounted for 9 – 19% of variance in set-shifting, inhibition/attentional control, as well as working memory. Additionally, independent paths to working memory at the time point of 11 years of age (*ß* = .43 and *ß* = .23) were also revealed for processing speed and memory, as well as for processing speed to set-shifting at participants' age of 11 (*ß* = .33). Thus, this study provided further support toward the hypothesis of unity and diversity of core executive control processes, as well as them being longitudinally related to later executive control performance.

Similarly, Huizinga, Dolan, and van der Molen (2006) recruited four groups of participants (7-, 11-, 15-, and 21-year-olds) and examined the developmental trajectories of set-shifting, working memory, and inhibition/attentional control and participants' performance on a series of cognitively complex measures. The results of the study further supported the evidence of the developmental aspect of executive control processes into adolescence. Indeed, set-shifting and working memory were found to reach adult levels in participants tested between 11 and 15 years of age. Inhibition/Attentional control, however, was not found to reach adult levels until after the age of 15. Moreover, sequential equation modelling indicated that all three *latent variables* (i.e., variables not directly observed but inferred from other observed variables) are clearly separable and provided additional evidence toward the non-unitary nature of the three executive control processes (Miyake et al., 2000).

Vaughan and Giovanello (2010) conducted another study examining executive control, but in a group of 95 older adults aged between 60 to 90 years of age. They also investigated the relationship between set-shifting, memory updating, and inhibition/attentional control by means of structural equation modelling. Their three-factor model indicated that indeed all three processes can be considered as a non-unitary construct and that “inhibition/attentional control, set-shifting, and memory updating function as underlying distinctive yet convergent processes” (Vaughan & Giovanello, 2010, p. 351).

Fisk and Sharp (2004) recruited 95 participants aged between 20 to 81 to further decompose the unity and diversity notion of the three executive control processes. A crucial difference between Fisk and Sharp’s study and the results obtained by Miyake et al. (2000) is the small magnitude of the correlations between the three executive control skills which leads to a question whether there is in fact any significant unitary aspect of the system. However, it has been proposed that such results might indeed stem from research with elderly persons (Rabbitt & Lowe, 2000). Some yet interesting results were also obtained from studies with preschool children. For instance, Wiebe, Espy, and Charak (2008) tested a group of 243 children between 2 years 4 months and 6 years of age from low- and high-income socio-economic status on a battery of executive control tasks which were adjusted for the children’s age. Whilst there were no effects of socio-economic status, the results pointed to a unitary model of executive control. Specifically, Wiebe and colleagues explained that executive control might be unitary, in other words, that it may be considered a domain general construct that manifests itself in various ways depending on contextual demands. This is in line with other studies (Duncan & Miller, 2002; Duncan & Owen, 2000). For instance, tasks chosen to tap into working memory and inhibition/attentional control were found to in fact assess the same underlying executive control process in children (Rose, Feldman, & Jankowski, 2012; Huizinga, Dolan, & van der Molen, 2006). This result may stem from a limited behavioural repertoire of preschoolers and the fact that some of the tasks employed were simplified by necessity. Thus, they might have been rather homogeneous in executive control demands than measures used with other populations.

Friedman, Miyake, Corley, Young, DeFries, and Hewitt (2006) recruited 234 twins who were matched on their cognitive ability and tested on a battery of measures, the majority taken from Miyake et al. (2000), assessing the three executive control skills as well as intelligence. Participants were tested on executive control tasks between the age of 16 to 18 (i.e., *mean* = 17, *SD* = .27), and on intelligence measures at 16 to 17 (i.e., *mean* = 16, *SD* = .27). Fluid intelligence – that is reasoning ability (Carroll, 1993) – was tested by Raven’s Progressive Matrices Test (Raven 1960) where participants are asked to complete a series of complex patterns. Also, they were asked to complete the Wechsler Adult Intelligence Scale Block Design test (Wechsler, 1997) in which participants are instructed to reconstruct patterns using given blocks. Additionally, to assess their *crystallised intelligence* (i.e., acquired knowledge; Carroll, 1993), a multiple-choice vocabulary test (taken from DeFries, Plomin, Vandenberg, & Kuse, 1981) was employed where participants' task was to provide word definitions as well as identify synonyms. The second test for crystallised intelligence assessment was the Wechsler Adult Intelligence Scale Information subtest measuring one's factual knowledge. Friedman et al.'s sequential equation modelling analysis found that set-shifting, memory updating, and inhibition/attentional control were also differentially related to intelligence. Only memory updating out of the three executive control processes, was highly correlated with both fluid and crystallised ($r = .741$, $p < .001$) intelligence while the other two were not (set-shifting: $r = -.010$, $p = .841$ and inhibition/attentional control: $r = -.070$, $p = .639$).

From a clinical perspective, the notion of multicomponent character of executive control enables better description of executive control deficits that are specific to various clinical impairments. Willcutt, Pennington, Boada, Ogline, Tunick, Chhabildas, and Olson (2001) compared a total of 314 twins aged 8 to 16 years old divided into four groups; (1) with attention-deficit/hyperactivity disorder (ADHD), (2) with reading disability (RD), (3) with RD and ADHD, and (4) neither ADHD nor RD. Results revealed that each group demonstrated different executive control impairment: the ADHD children demonstrated inhibition/attentional control-related deficits, children with reading disorder presented deficits in working memory (related to memory updating), and the final group with both disorders combined showed deficits in terms of inhibition/attentional control, working memory, and set-shifting. Geurts, Verte, Oosterlaan, Roeyers, and Sergeant (2004) replicated these findings in a 6- to 12-year-old group of 54 children with ADHD and 41 children with high functioning autism. ADHD children demonstrated inhibition/attentional control and word production impairments while children diagnosed as high functioning autistics exhibited impairments in general executive control except inhibition/attentional control and working memory. Taken together, this evidence demonstrates that in clinical populations there are preserved executive control processes as well as individuals with selective impairments in different executive control processes which provides further evidence of the unity and diversity.

However, there has been evidence that contradicts the idea of unity and diversity of executive control. For example, a systematic review by Karr, Areshenkoff, Rast, Hofer, Iverson, and Garcia-Barrera (2018) found the unidimensional model of executive control to be the most common among children and adolescents, whereas the multidimensional model of executive control was found in adult samples. No model was found to be consistently applied to all samples. As already indicated by some research (Wiebe, Espy, & Charak, 2008), a greater unidimensionality of executive control can be indeed found in child and adolescent samples whereas both unity and diversity are most common among the adults. It might stem from a limited behavioural repertoire of preschoolers and the fact that some of the tasks tapping executive control need to be simplified by necessity. Moreover, Karr et al. also suggested that such results might be due to the possible publication bias resulting in the publication of underpowered studies but with well-fitting models.

As this section provided the empirical evidence for the unity and diversity framework supported by research findings, the following section elaborates on the substantial genetic contribution in individual differences in executive control.

1.2.2 Substantial Genetic Contribution

Friedman, Miyake, Young, DeFries, Corley, and Hewitt (2008) examined the role of heritability in individual differences in executive control. They conducted a twin study with 582 participants: 316 monozygotic and 266 dizygotic twins (age not reported). Results revealed that set-shifting, memory updating, and inhibition/attentional control are correlated as they are impacted by a highly heritable common factor (99%) which cannot be accounted for by perceptual speed or one's general intelligence (Friedman et al., 2008) and thus it has been proposed that executive control is one the most heritable psychological traits.

The rationale for conducting research with twins is the fact that they share on average half of their genes from their biological parents, and they share their home environment. According to results obtained from research with monozygotic and dizygotic twins, individual differences in executive control may stem from either genetic or environmental variances. Thus, monozygotic twin correlations are higher than dizygotic ones for a behavioural measure and indicate a genetic impact on that measure. In a situation when dizygotic correlations are in fact higher than expected, there is an indication of environmental impact, whereas when monozygotic correlations are lower than expected (less than 1.0), nonshared environmental influences are implicated (Friedman et al., 2008). It needs to be noted that in terms of the nonshared environment, a measurement error needs to be also considered at the level of individual tasks.

Such additive genetic impact is assumed to comprise the effects of a significant number of specific genes that work together in an additive way. Shared environmental influences are on the other hand influences that make contributions towards the similarity of twins such as family, friend- and peer groups, as well as mother's health during gestation (Friedman et al., 2008). Nonshared environment refers to influences that lead to uncorrelated experiences between the twins, for instance educational environment. While heritability constitutes a moderate proportion, the remaining variance has been claimed to be attributable to one type of environmental variance, namely, nonshared environment (comprising measurement error). However, when measurement error is minimised, which takes place at the level of latent variables, genetic variance is at an even higher level (over .75 as opposed to .25 - .55). Interestingly, significant genetic impact was observed at unity and diversity levels indicating that different sets of genes contribute to variability in *shared* and *unique* executive control processes.

Interestingly, research with twins indicate that one's general executive control ability has been found to be moderately heritable (approximately 50%), and the other half of variance is in fact split rather equally between shared and nonshared environmental influences (i.e., Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg, & Urbina, 1996). This trend is however not constant and changes with age with the major decrease observed for the role of shared environmental influences and the heritability factor raising to approximately 70% or even higher by the late adolescence and into adulthood (Pedersen, Plomin, Nesselroade, & McClearn, 1992). The remaining percentage of the variance is accounted for by non-shared environmental influences.

Friedman et al. (2008) conducted a multivariate study with twins investigating set-shifting, memory updating, and inhibition/attentional control, which were measured as latent variables. According to their findings (2008, p.1), the correlation between the three processes was in fact "influenced by a highly heritable common factor that goes beyond general intelligence or perceptual speed, and they are separable because of additional genetic influences unique to particular executive functions" Friedman et al. further claim that executive control is in fact the most heritable. Indeed, there is a potential for further research examining the role of genetic influences in individual differences within the executive control.

Having presented the evidence supporting the notion of substantial genetic contribution to the development of executive control, the following section elaborates more on the concept of developmental stability in the executive control system.

1.2.3 Developmental Stability

The last conclusion concerning individual differences in executive control refers to the development stability of executive control within an individual. To support this conclusion, the results from longitudinal studies with twins come in handy. Friedman, Miyake, Robinson, and Hewitt (2011) investigated executive control skills in 945 twins aged 17 and 23 who moved out of their family homes and started their independent lives. Results revealed that at both ages correlations are indeed substantial at the latent variable level (and equates as follows: r = .822 Common (shared) EC skills, r = 1.00 specific to updating, r = .932 specific to set-shifting). Additionally, twin studies provided identification of early measures predicting future individual differences in executive control being a measure of self-restraint in toddlerhood Indeed, in a task resembling delayed gratification task, a toddler was presented with a toy which they were asked not to touch for a certain period of time (Mischel, Ayduk, Berman, Casey, Gotlib, Jonides, Kross, Teslovich, Wilson, Zayas, & Shoda, 2011). Two different developmental trajectories emerged based on four test points at 14-, 20-, 24-, and 36-month of participants' age. In fact, at participants' age of 17 this group difference was still present. This suggests that although overall executive control skills improve substantially with age, they stay relatively stable within individuals.

1.2.3.4 Test-Retest Reliability

As proposed by Miyake and Friedman (2012), individuals vary in terms of their overall executive control performance as well as in terms of specific executive control components. In other words, individuals who demonstrate superior interference control in one task should also demonstrate great interference control in another task. Thus, scores obtained in various tasks measuring the same executive control component should indeed correlate and show convergent validity. However, this does not seem to occur consistently according to the literature. Research has found (Paap & Greenberg, 2013; Paap & Sawi, 2014) that cross-task Simon and Flanker correlation was negative and weak (r = -.01). What is more, Salthouse (2010) further presented that the letter and the arrow version of traditional Flanker task were not correlated with one another (r = .03).

Hedge, Powell and Sumner (2017) investigated test-retest reliability in a number of executive control tasks and found that although they are commonly used, their test-retest reliability is surprisingly low (i.e. from 0 to 0.82). They argue that such results stem from the low between-participants variability. In other words, the reliabilities of many widely used executive control measures are lower than might be expected by researchers as well as lower than some standards outline (i.e. Fleiss, 1981; Barch et al., 2008).

Although robust and replicable, some measures seem to fail to consistently capture individual differences within a given population and this makes them problematic as correlational tasks.

According to Hedge et al. (2018), between-session variance is linked to the systematic between-session changes across the sample whereas error variance is linked to non-systematic between-subject scores between sessions. This means that while the score will increase for some participants, it will decrease for others. In other words, the reliability of the measure decreases with higher measurement error while between-participant variance remains constant. Additionally, reliability also tends to decrease when the between-participant variance is smaller while error variance remains constant. Thus, if there are two measures with the same measurement error, the reliability will be lower for the one which exhibits more homogeneity. Hedge et al. (2018) further explains that measures that are characterised by low reliability should not be employed in correlational research as this compromises the ability to explore relationships with other constructs due to the issues with distinguishing the performance between individuals tested. Indeed, homogeneity is desired in certain experimental designs as it allows to investigate individual differences, however, it may come at disadvantage in the within-subject designs where minor variation is sought.

Although the multi-method approach to examining executive control yields many benefits, there is still very little research that focuses on evaluating the psychometric properties of the commonly used tasks. In fact, test-retest reliability of executive control measures is very limited and has focused largely on adults (i.e., Ingram, Greve, Ingram, & Soukup, 1999; Fan, McCandliss, Sommer, Raz, & Posner, 2002). Thus, it warrants a further investigation into the psychometric properties of these measures and their test-retest reliability.

1.3 Interim Summary

The aim of this section was to introduce and discuss the conceptualisations of executive control as well as three major executive control processes; namely, set-shifting, memory updating, and inhibition/attentional control. It also presented Unity and Diversity of executive control framework by Miyake et al. (2000) along with the major conclusions regarding individual differences in performance on executive control tasks.

Section 1.4 of this Chapter will focus on various conceptualisations of memory from the perspective of typical development in more detail. It will further cover the issue of working memory models and multi-store and unitary-store models will be discussed more extensively. This will be followed by a presentation and introduction of the working memory model by Baddeley and Hitch (1974) and the model evaluation.

1.4 Conceptualising Memory

The role memory plays in our everyday lives is invaluable. Without it, we would not be able to remember and recognise familiar objects and people. Moreover, we would not be able to communicate, read, and write. In other words, we would exist in a world of no recollection ability – born and living tabula rasa. Indeed, memory is a crucial component of everyday existence – it allows us to use the language to communicate, to decode print and read books, to remember phone numbers and people's names, as well as recognise family and friends' faces. Memory is present in almost every action we take on a daily basis.

1.4.1 Working Memory Models

Memory and learning are inevitably linked together. The initial phase of learning (i.e., presentation of to-be-learned material) is known as *encoding* and once encoded, the newly learned material is then stored in memory. Thus, *storage* and *retrieval* of the newly acquired learning material from memory storage are the next phases of learning. In other words, for the new material to be successfully learned, it needs to be encoded, stored in memory, and retrieved for further purposes.

One of the main distinctions proposed by theorists is the distinction between *short-term memory* and *long-term memory* with the distinction being that short-term memory is of a limited capacity and long-term memory is of unlimited capacity. Another dimension of difference is the duration of their ability to hold information: material in short-term memory is usually stored for a few seconds whereas in long-term memory material endures without conscious effort for decades.

A more comprehensive presentation of long-term memory and its role in paired associate learning will be elaborated on in more detail in Chapter Two.

1.4.1.1 Multi-Store Model

One of the first influential memory models in recent times was introduced by Atkinson and Shiffrin (1968). Their multi-store model comprised three types of memory retention: a sensory memory, short-term memory, and long-term memory storages. In line with the model, environmental input is initially processed by sensory memory which is modality-specific (i.e., there is a visual store, so called *iconic memory*, and an auditory store also referred to as *echoic memory*). The input is then held there for a short time to assess which information will be attended to and which information will stay unattended. The environmental input is further directed into short-term memory to be processed further and potentially be transferred to long-term memory storage. The multi-store model components are discussed below in more detail.

Sensory memory, also known as *the sensory register*, refers to the ability to store sensory information (limited to one modality) for a very brief period of time where it then decays and disappears. This store is constantly in receipt of various sensory input. Depending on the type of sensory input, two different sensory stores become engaged. Visual information is stored temporarily in the visual memory store and stays there for approximately 500 milliseconds (Sperling, 1960). This duration has been however argued to be longer as the mechanisms in the visual store always operate on the icon. According to Atkinson and Shriffin's model (1968), iconic memory is pre-attentive which means that it operates regardless of inhibition/attentional control. Moreover, they proposed that inhibition/attentional control resources are engaged only after the input has been held in the sensory stores. Recent research findings have accumulated evidence to the contrary, indicating that when completing an attentionally demanding task, participants' performance of iconic memory is significantly weaker and severely disrupted (Persuh, Genzer, & Melara, 2012). Echoic memory, on the other hand, is a sensory store where auditorily presented input is held for a few seconds. This type of memory has been investigated more in-depth by Ioannides, Popescu, Otsuka, Bezerianos, and Liu (2003) who looked at brain activation during tone presentations. They found that echoic memory was longer lasting (up to five seconds) in the left hemisphere than in the right hemisphere where it stayed for two seconds – a result that seems to reflect the left hemisphere specialisation for language processing.

Incoming visual information has been rather extensively researched (Sperling, 1960) however less is known about the information coming from other modalities. As registers for other senses are difficult to investigate experimentally, these are in fact under-researched and weakly understood.

Atkinson and Shiffrin (1968) employed the term *working memory* to refer to their concept of the short-term memory store. Short-term memory, in other words a memory system combining short-term information storage with the ability to process information, receives input selectively from both sensory memory as well as long-term memory storage. Information that reaches short-term memory storage stays there for about 30 seconds, then it decays and is lost. However, by means of rehearsal, the limited amount of information can in fact be stored in short-term memory for as long as the individual wishes. The short-term memory capacity has been claimed to be very limited and constitute 5 – 9 items which stems from Miller's magic number 7 +/- 2 (Miller, 1956). However, this number of seven items as an indicator of short-term memory capacity has been debated. Mathy and Feldman (2012) emphasise the difference between single item retrieval and remembering chunks of information. For instance, when presented with a word M E M O R Y in a letter-span task, participants see six letters but only one chunk (a word) and thus their performance on the task is errorless. They further argued that participants' performance on short-term memory tasks is often inflated due to repetition and their long-term memory. Indeed, in the absence of rehearsal, participants tend to produce between three to four items rather than suggested seven (Chen & Cowan, 2009).

Another assumption made regarding the short-term memory system within a multi-store memory model is the equal significance of all the given objects. This assumption has however been met with criticism with evidence (Nee & Jonides, 2013; McElree, 2006) pointing to varying levels of significance of items depending on their position in short-term memory. For instance, McElree (2006) suggested a privileged status of the item within this memory storage that is the current focus of one's attention. This proposal has been supported by experiments with probes in which responses to a probe are faster if it matches the most recently rehearsed item in a list of items.

There are two ways by which information disappears from short-term memory storage, namely, *decay* and *interference*. Decay refers to the disappearance of information over time due to lack of rehearsal. Interference on the other hand implies intrusion from previous experimental trials or other information present during the retrieval phase which results in losing the initial information of interest from short-term memory. Berman, Jonides, and Lewis (2009) investigated the ways in which information is forgotten and lost from short-term memory. They believed that interference plays a more important role in forgetting information than decay based on the seven experiments they conducted. In the study, participants' performance on short-term memory was disrupted due to the lexical items from the previous experimental trials which had not yet decayed. They argued that the effect of disruption could be minimised if the between-trials intervals were increased in time allowing for a sufficient decay of prior words. However, this strategy did not result in any changes and did not impact participants' short-term memory performance. Interestingly though, the disruption was eliminated when the interference from prior trials was minimised.

Campoy (2012) found certain methodological limitations of the study by Berman, Jonides, and Lewis (2009) and revised the study by reducing the time scale from an initial 3300 milliseconds (i.e., 600, 1200, 1800, and 2400 milliseconds). Indeed, the findings imply that there are in fact strong decay effects to be observed if time intervals are shorter than 3300 milliseconds. This suggests that decay effects can be noted mainly at short retention intervals whereas interference mostly occurs at longer time intervals.

Another memory store, *long-term memory*, differs from both previous systems and comprises a relatively stable and permanent storage of information that is moved there from the short-term memory. It can be characterised by unlimited capacity and it has been proposed that the information within long-term memory storage does not decay. This movement is also referred to as a *transfer* during which the information is copied from short-term store to long-term memory without this information decaying in the original memory store.

The multi-store memory model is still widely accepted with a valid distinction into three separate but interlinked memory storages. The most significant distinction into short-term and long-term memory stores is indeed heavily supported by evidence accumulated via research examining brain-damaged individuals. Indeed, amnesic patients demonstrate impairments in their long-term memory storage with the short-term store remaining rather unaffected (Spiers, Maguire, & Burgess, 2001). The opposite pattern has also been identified by research with individuals whose short-term memory deficits do not affect their long-term memory capabilities (Shallice & Vallar, 1990).

However, the multi-model store has been found to have some limitations. One of the key arguments against the model concerns its oversimplified approach to memory architecture. Indeed, the model assumes that both short-term and long-term memory stores are separate constructs and operate in a single and uniform manner. Also, it does not offer a sufficient description and explanation of both short-term and long-term memory concepts, as well as the linear character of the information flow which suggests that difficulties in short-term memory would prevent information transfer to long-term memory and vice versa. The linear relationship between these two memory storages has been indeed further challenged and claimed to be non-linear (Shallice & Warrington, 1977).

Moreover, the suggested gateway role of the short-term memory store to both sensory memory and the long-term store seems to be invalid. To illustrate this, let us imagine rehearsing a chunk of information in a form of letter cluster in short-term memory, the word NASA for instance. It is possible to rehearse it only because it has already accessed relevant information available in the long-term memory (Logie, 1999). In their memory model, Atkinson and Shriffin (1968) imply that only consciously acquired information can be transferred to long-term memory. This is contradictory to what research suggests with regards to implicit learning – the kind of learning that involves unconscious acquisition of learning material. Studies investigating implicit learning point to the information learned via this route to be successfully transferred to one's long-term storage. Moreover, rehearsal processes constituting the most crucial step in the process of transferring verbal information into long-term memory were also approached with uncertainty (Logie, 1999). In fact, information learned via rehearsal constitutes only a small proportion of all the information that is available in long-term memory storage. Thus, emphasising the role of rehearsal in information transfer from short-term to long-term memory is in fact considered an overstatement by other researchers.

In opposition to the presented multi-store model of memory, the proposal of a unitary-store model of memory has been put forward and it is presented in the following section.

1.4.1.2 Unitary-Store Model

An alternative to the previously described multi-store memory model (Atkinson & Shiffrin, (1968), is the unitary-store model proposed by Jonides, Lewis, Nee, Lustig, Berman, and Moore (2008). The unitary-store model combines both short-term and long-term memory stores, and at the same time rejects the idea of these being distinct memory systems. In other words, the two stores are not architecturally separate, and they constitute one unitary memory store of short-term memory with "temporary activations of long-term memory representations or of representations of items that were recently perceived" (Jonides et al., 2008, p. 198). Jonides et al. also suggested that the hippocampus and medial temporal lobes play an important role in forming novel associations and relations between items in both short-term and long-term memory stores.

Long-term memory can be manipulated in a number of ways using tasks such as backward digit recall task (i.e., participants are asked to recall a list of items in a reverse order that they were presented). Moreover, neuroimaging research has not supported the unitary-store model and there is very limited evidence for hippocampus to be engaged when short-term and long-term memory processes are being separated (Bergmann, Paulus, & Fikkert, 2012). For instance, Bergmann et al (2012) investigated word pairs and their processing in short-term memory as well as long-term memory stores. The study revealed the activation in the hippocampus when learning these pairs was not evident.

Both multi-store and unitary-store models have both their supporters and opponents. One of the most influential working memory models which has evolved over the years is the model introduced by Baddeley and Hitch (1974) and is discussed in more detail in the following section.

1.4.1.3 Working Memory Model

Baddeley and Hitch (1974) challenged Atkinson's unitary memory model and proposed another model of memory. Their model was extended into a new concept of working memory storage referring to a limited-capacity system, which is responsible for retention and manipulation of information required for a successful completion of a particular task (Baddeley & Hitch, 1974). Baddeley and Hitch's model is composed of central executive, a modality-free and limited capacity system, that controls two subsidiary systems, the phonological loop and the visuospatial sketchpad (the phonological loop in charge of processing and temporarily storing speech-based information whereas visuospatial sketchpad is a temporary storage of visual and spatial information).

The visuospatial sketchpad and the phonological loop are to a great extent independent of each other. Two main assumptions of the model state that: 1) in a situation where two memory components (i.e., visuospatial sketchpad and phonological loop) are engaged at the same time, the task cannot be successfully completed, 2) in a situation where two memory components are engaged at the same time, the task should be performed successfully together as if they were performed separately.

The way these assumptions work in real life was investigated by Robbins, Anderson, Barker, Bradley, Fearnyhough, Henson, Hudson, and Baddeley (1996) who recruited and tested chess players on their move choices while they were also completing a given task depending on the group they were assigned to. The following groups are described below: (1) control condition where participants were asked to tap repeatedly; (2) random letter generation (used to block central executive); (3) operations on a keypad (employed to block visuospatial sketchpad); and (4) repetitions of a word (to engage articulatory suppression via phonological loop). The results revealed that chess movement selection requires the engagement of the central executive as well as visuospatial sketchpad however phonological loop is not involved in this process. Indeed, suppressing the articulatory loop did not have any impact on participants' performance (similarly to the control condition), whereas visuospatial sketchpad and the central executive interference resulted in disruptive effects on participants' performance.

1.4.1.3.1 Phonological Loop

Phonological loop is a key speech-based component of working memory which is assumed to consist of two subcomponents: a phonological store where temporary storage and maintenance of auditory and verbal information occurs (Gathercole & Baddeley, 1993; Gathercole, Willis, Emslie & Baddeley, 1992). Thus, its task is to maintain rehearsal of verbal input to ensure it is available for mental processes until it is no longer needed or replaced by new incoming information. This subcomponent is directly linked to speech perception. The second one, on the other hand, is linked to speech production and refers to an articulatory process enabling access to phonological memory store.

The capacity of the phonological loop has been traditionally measured using a non-word repetition span as well as digit span task (i.e., being able to recall a list of items in order they were presented). As nonword repetition tasks rely on unfamiliar sound combinations that are not stored in one's mental lexicon, the mediation ought to be directed through the phonological loop which stores the unfamiliar representation for a short time. In contrast, digit span, which is often used to assess the capacity of verbal short-term memory employs familiar sound patterns which can be retrieved from existing phonological representations in the mental lexicon (Gathercole, Service, Hitch, & Martin, 1997). The ability to repeat nonwords has been found to be correlated with word knowledge (Baddeley, Gathercole, & Papagno, 1998) and hence it has been suggested that the ability to efficiently store temporary sound representations in verbal short-term memory, i.e., the phonological loop, determines the stability of long-term phonological structure of a new word (studies with adults, Papagno & Vallar, 1992).

The phonological loop is of limited capacity and it has been argued that people can remember from 5 up to 9 items. The limited capacity of the phonological loop has been demonstrated using the word length effect and the phonological similarity effect. Immediate memory span is believed to be word-length dependent and hence weaker for long words and greater for short ones (Baddeley, Lewis, & Vallar, 1984). However, Jalbert, Neath, and Surprenant (2011) indicated that the word length effect might be linked to words' orthographic neighbourhood size rather than the length of the word. Indeed, short words tend to be characterised by more orthographic neighbours than longer words and when they both are equated for the neighbourhood size, the word-length effect is no longer present (Jalbert et al., 2011). Phonological similarity effect, on the other hand, proposes that the more similar the words to be learned are, they seem to be more difficult to retain and hence may require more verbal short-term memory capacity to be stored than dissimilar ones (Sperling & Speelman, 1970; Conrad & Hull, 1964; Wickelgren, 1965). Indeed, this effect was further examined by Acheson, Postle, and MacDonald (2011) who proposed that the effect does not concern the phonological loop as such but rather the semantic processes.

All the above factors may negatively impact the functioning of the phonological loop and hence hinder new word acquisition. A paired-associate paradigm was utilised to investigate these variables on novel word learning in Italian students learning spoken Russian words and non-words (Papagno & Vallar, 1992) and confirmed the involvement of phonological short-term memory in unfamiliar word learning. Indeed, articulatory suppression in these participants led to slower rates of foreign vocabulary learning due to the reduction in the usage of the phonological loop (Papagno, Valentine, & Baddeley, 1991).

Baddeley (1998) has also emphasised the role of the phonological loop in language learning. They worked with a brain-damaged participant, known in the literature as PV, whose digit span equated two items. PV had a short-term phonological deficit which was linked to a specific impairment within their long-term learning of stimuli that were unfamiliar. The patient did not find it difficult to associate pairs of novel words that were unrelated to each other but their performance on matching related words and their translations in two languages (Italian – which was PV's native language, and Russian) was significantly weaker.

1.4.1.3.2 Visuospatial Sketchpad

The visuospatial sketchpad is a working memory subcomponent that provides temporary storage for and manipulates visual and spatial information. In other words, the visual element refers to remembering *the contents* whereas spatial processing relates to *the location of the contents*.

There has been some debate regarding the composition of this system with researchers proposing the existence of a single system with visual and spatial processing as one, and others

(Smith & Jonides, 1997) who suggested the presence of two separate subsystems: visual and spatial independently. Logie (1995) made a distinction into two main components: (1) a visual cache which is in charge of visual form and colour, and (2) an inner scribe which is a system that comprises spatial processing and details about movements. The latter also plays a role in transferring the input from the former, visual cache, to the central executive.

Smith and Jonides (1997) conducted a study where participants were asked to decide if the target object they were presented with, matched the initial spatial location of the target item or whether it shared visual features with the target. They found significant differences in terms of participants' brain activation patterns: the spatial processing element was found to be associated with more activity in the right hemisphere, whereas the visual task resulted in a bigger activity in the left hemisphere. More in-depth investigation into brain areas engaged in visual and spatial processing was employed by Zimmer (2008). They argued that the occipital and temporal lobes are the brain areas within which more activity is being observed during visual processing and the parietal cortex areas get more activation during spatial processing tasks.

Although there are differences to be considered between visual and spatial information processing, there are also some similarities and common features. In fact, both components have been argued to be in need of attentional control resources of the central executive. Thus, it has been proposed that the separability of the visual processing and the spatial processing systems are dependent upon the amount of attention needed for a given task (Vergauwe, Barrouillet, & Camos, 2009). The issues whether there is one single system or two separate systems within the visuospatial sketchpad remains largely unresolved and constitutes a focal point of interest for a number of studies (Baddeley, 2012).

1.4.1.3.3 Central Executive

Baddeley and Hitch's (1974) central executive component of working memory has been claimed to be the most crucial element in the whole system. Although the prefrontal cortex brain areas have been argued to demonstrate the most activation and engagement in the central executive, these are believed to be not the only brain regions involved in executive control (Stuss, 2011). Baddeley (1996) consistently proposed that the central executive is strictly linked to a number of executive control functions. To be more precise, four such executive processes were identified by Baddeley (2012): (1) "to focus attention on complex tasks"; (2) to divide attention between two important targets or stimulus streams"; (3) "switching between tasks for which […] there might be a specific control system"; and (4) "to interface with long-term memory" (Baddeley, 2012, p. 14).

1.4.1.3.4 Episodic Buffer

The term *episodic buffer* refers to a limited-capacity storage system located within working memory that is in charge of integrating information from different sources (Rudner & Rönnberg, 2008). The function of the episodic buffer includes information processing and storage. In other words, it aids assembling multidimensional representations in various codes from various sources (Baddeley, 2000) and further stores them temporarily. It also acts as a buffer linking working memory to perceptual memory and long-term memory. It has been proposed that the episodic buffer is controlled by the central executive that enables information retrieval, its manipulation and modification (Baddeley, 2000).

Indeed, the episodic buffer is believed to play a crucial role in transferring information into and from episodic long-term memory (Baddeley, 2000). The episodic buffer constitutes a storage for phonological input from the phonological loop and visuospatial input from the visuospatial sketchpad. An interesting finding by Baddeley, Vallar, and Wilson (1987) is the immediate recall ability of their participants: one's ability to immediately recall target words equals up to five unrelated lexical items, however, up to sixteen words presented in sentences. The latter has been argued to be beyond the capacity limits of phonological loop and thus might be proposed to be accounted for by the capacity of the episodic buffer. In their further investigations, Baddeley and Wilson (2002) proposed two factors that might influence ability to demonstrate an immediate recall of up to sixteen-word sentences, namely; (1) the episodic buffer's capacity, as well as (2) the efficiency of the central executive in integrating or separating sentence information.

Currently, the episodic buffer should be concerned as "an essentially passive structure on which bindings achieved elsewhere can be displayed" (Baddeley, 2012, p. 17). Although the episodic buffer serves as a store for both phonological and visuospatial information, there is little research concerning how the input from various modalities is being combined there into a single representation. Also, even less is known about information coming from other modalities, such as taste or smell for instance, and the process they come through in the episodic buffer.

Working memory is typically measured using Word Recall, Sentence Recall, Digit Recall, Backward Digit Recall, Block Recall, Listening Recall, Spatial Recall, Nonword Repetition, Dot Matrix, and Odd One Out tasks.

In *Word Recall* task, participants are presented with a list of words and asked to remember as many of them as possible. Similarly, the same procedure is applied in the *Sentence Recall* task with the difference that participants are shown a list of sentences to be remembered and recalled. *Nonword Repetition* is a task in which participants are asked to repeat nonwords (i.e., *gigogin*).

In *Digit Recall*, participants are exposed to a series of numbers which they are asked to remember and recall in the same order as presented. In *Backward Digit Recall* task, participants are asked to recall the numbers they were presented with in a backward order that they were shown (or played).

(Corsi) Block Recall (Corsi, 1973) requires participants to reproduce the sequence of block-tapping in the same order as presented. The task involves a backward condition in which participants are required to reproduce the sequence of block-tapping in the reverse order.

In *Dot Matrix*, participants are asked to remember and recall the location of a series of dots and the pattern in which all the dots were presented. Then, they are instructed to point to a block on the grid where the dot had been displayed one by one in an order they had been presented.

Odd One Out task requires participants to deduce about different features of several shapes to eliminate the one shape that does not fit in with the rest.

Having introduced the main components of the working memory model as well as common measures used to assess it, the next section will provide insight into the evaluation of the working memory model proposed by Baddeley (2012).

1.4.1.4 Model Evaluation

Although highly influential, the working memory model by Baddeley (2012) has been met with some criticism (Postle, 2006) concerning its oversimplification amongst others. The model consists of three main information storages including modalities such as verbal, visual, and spatial input. However, other modalities (i.e., touch, taste, smell), although mentioned briefly within the theoretical considerations, have not been given much consideration. Similarly, Postle (2006) argued that even though the spatial information processing is included in the model, it does not account for numerous types of spatial working memory (hand-, eye-, footcentered) as well as working memory for visual objects is arguably neurally separable (i.e., faces versus houses; see Ranganath, Cohen, Dam, & D'Esposito, 2004). Additionally, Postle argued that within verbal working memory, phonological, semantic, and syntactic information have been found to be separable (Martin, Hamilton, Lipszyc, & Potts, 2004). Also, working memory for loudness, pitch, and location can be dissociated from each other (Anourova, Rämä, Alho, Koivusalo, Kahnari, & Carlson, 1999; Clement, Demany, & Semal, 1999). In line with this argument, for the working memory model to account for all these dissociations, it would need to comprise "hundreds […] of domain-specific buffers, each responsible for the working memory processing of a different kind of information" (Postle, 2006, p. 25).

Another crucial limitation is the conceptualisation of executive control per se (Postle, 2006). Indeed, the executive processes that are linked to the central executive are challenging to identify and list due to the fact that complex executive control tasks require more than one executive process and the involvement of each process is not always easy to disentangle.

1.4.2 Working Memory and Executive Control (Inhibition/Attentional Control)

The contemporary revised version of Baddeley's working memory model (2001) emphasises the crucial role of the central executive in the processing of temporarily stored information. For instance, Baddeley (1996) explains that performance on the digit-span task is rather determined by storage capacity than executive control skills due to its rather simple processing. Nevertheless, he further clarifies that as the load of the digit span task increases, there are indeed more demands being placed on the central executive (Baddeley, 1996). In fact, verbal memory capacity is dependent upon the phonological loop as well as the central executive – as the task load increases, individuals need to recruit their central executive resources. Thus, it can be suggested that to be able to maintain and store substantial working memory loads, certain strategic executive processes need to be employed (i.e., chunking of information).

A series of studies (Engle, Kane, & Tuholski, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003) offered a proposal assessing working memory capacity – the amount of information a person is able to store and process at the same time – are reflective of an individual's capacity of information maintenance despite interference even when it is highly active. The ability to maintain target information in a highly active state and so it is easily accessed when needed, reflects one's ability of attentional control since "coherent and goal-oriented behaviour in interference-rich conditions requires both the active maintenance of relevant information and the blocking or inhibiting of irrelevant information" (Kane, Bleckley, Conway, & Engle, 2001, p. 170). Working memory plays an important role in this as it contributes to the active task goal maintenance as well as maintenance of information that is relevant in this context.

The importance of working memory capacity can be displayed by its high correlation (r = .60) with fluid intelligence measures (Unsworth, 2010). Indeed, it has been argued that individuals who are characterised by a superior working memory capacity demonstrate superior executive control as well as inhibition/attentional control resources. Research by Engle and Kane (2004) resulted in a *two-factor theory* which proposes one factor related to the task goal maintenance whereas the second factor includes the resolution of either conflict or response competition. This is in line with other findings that high working-memory capacity individuals demonstrate better performance on task goal maintenance as well as conflict resolution. Moreover, Engle, Tuholski, Laughlin, and Conway (1999) asked 133 participants to complete 11 memory tasks believed to be tapping either working memory or short-term memory as well as general fluid intelligence tasks. The results revealed that short-term and working memories constitute separate but highly related constructs (Engle, Kane, & Tuholski, 1999). Interestingly, they also found that tasks commonly used in the literature to assess working memory tend to reflect a common construct. Indeed, short-term memory is not correlated with fluid intelligence whereas working memory demonstrates a strong connection to fluid intelligence.

Taking this into account, Engle, Kane, and Tuholski (1999) proposed that there might be a relationship between individual differences in inhibition/attentional control (as measured by the Antisaccade task or Stroop task) and working memory capacity. This study provided evidence that low working memory span individuals spent significantly more time on identifying targets and scored higher in terms of interference in the Stroop task when compared to the high working memory span participants, thus supporting Engle and colleagues' proposal. Indeed, it was claimed that working memory performance is reflective of one's capacity for inhibition/attentional control. In other words, high working memory span individuals have the ability to recall information by means of irrelevant stimuli suppression, and not necessarily due to having a better capacity for information storage.

Research examining working memory capacity has found that it is highly correlated with executive control measures (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Moreover, McCabe et al. argued that the two are both reflective of one's executive attention needed for task goal maintenance as well as interference resolution. They investigated the relationship between inhibition/attentional control and working memory in a sample of 200 adults. The results revealed a very strong correlation between working memory and executive control (r = .97) but weak correlations between processing speed and the two former constructs. Additionally, controlling for executive control and working memory removed the age effect on episodic memory, and executive control and working memory accounted for variance in episodic memory which was beyond the variance accounted for by processing speed. This is indicative of a common underlying inhibition/attentional control component being shared by working memory and executive control. Sorqvist (2010) examined distraction sound effects and found that participants with better working memory capacity, thus greater inhibition/attentional control, were less affected by the distractors than the lower capacity group.

Individuals with low working memory capacity have also been identified as having difficulties with efficient task goal maintenance (Unsworth, Brewer, & Spillers, 2012). In a series of experiments (including Antisaccade, Stroop adaptation, and Flanker task), Unsworth et al. (2012) found that high working memory capacity individuals are more efficient in their responses which is consistent with the finding that working memory capacity is related to active maintenance abilities. For instance, in an antisaccade part of the study, they found high-capacity participants to be better at inhibiting a most obvious predisposition to visually follow a flashing cue than their low working memory capacity counterparts. In line with these results, the individual differences in working memory capacity are partially linked to executive control operations (Unsworth et al., 2012).

In a series of studies by Hester and Garavan (2005), higher working memory load was found to weaken the ability of inhibition/attentional control (as measured by switching and inhibitory control paradigms). Their results supported the inhibition/attentional control theoretical framework stating that for executive control and working memory capacity, the active maintenance of competing task goals is crucial. Also, it offers further evidence of a decline in switching as well as exerting inhibition/attentional control, as a function of working memory load. Hester and Garavan (2005) managed to demonstrate that both the size and the contents of working memory have an impact on the control of inhibition/attentional control resources.

Indeed, in a number of theories concerning the relationship between working memory and inhibition/attentional control, attention has been featured as a limited resource that is in charge of maintaining information in an active (in other words available) state (Cowan, 2005). The idea of an attentional resource for storage and processing is also true for the same shared resource for both perception and memory. They both have the same assumption of the resource being allocated to object representations as well as to events that one stores in their working memory. In terms of inhibition/attentional control, this proposes a resource for control of our actions and thoughts.

Attention is also believed to be shared between storage and processing (Case, Kurland, & Goldberg, 1982; Cowan, Elliott, Saults, Morey, Mattox, Hismjatullina, & Conway 2005). In line with this proposal, it can be assumed that there is a competition between attention-demanding executive control processes and concurrent storage (Chen & Cowan, 2009) as attention is required for both: keeping information active as well as enabling basic executive control processes (i.e., stimuli response selection).

The different conceptualisations and approaches have various implications for working memory. One example is a situation when there is an overload of relevant and irrelevant information such as in the complex-span paradigm (Daneman & Carpenter, 1980). Complex span tasks involve processing and storage aspects of working memory (Daneman & Carpenter, 1983). For example, participants are asked to read sentences, which is a form of online processing, while simultaneously trying to store certain aspects of this processing for subsequent recall. Another example is a filtering paradigm in which participants are asked to pay attention to visual stimuli and retain a set of targets that was defined prior to the task commencement (Vogel, McCollough, & Machizawa, 2005). It can be assumed that working memory representations need attention allocated to them. Thus, attention imposes a limit on the amount of information that can be remembered and a separate filtering parameter monitors the amount of irrelevant information that is being kept out of working memory so it does not take up valuable space in its storage. This would lead to a conclusion that individuals characterised by a lower working memory capacity are able to keep smaller amount of both target and nontarget information and this proportion should not depend on their working memory capacity. Contrastingly, the controlled-attention approach suggests that the filtering efficiency is defined by the attentional resource. Thus, individuals with lower capacity of working memory are able to remember the same amount of information as individuals with higher capacity, however, the difference can be observed in the proportion of target to nontarget (distractors) information they can retain.

Two different approaches have been offered here. One variant proposes that keeping a target representation in working memory requires an assigned continuous share of attentional resource (Case, Kurland, & Goldberg, 1982) whereas the second proposal suggests that it is not directly required for storage, but it is indeed necessary for processing. The latter in fact points to the indirect contribution of attention to working memory maintenance as it is required for keeping working memory representations active to prevent their decay (Barrouillet, Bernardin, & Camos, 2004). The specific attentional resource for refreshing information stored in working memory has been claimed to be limited and needed for central processes (Barrouillet, Bernardin, Potrat, Vergauwe, & Camos, 2007).

In some theories supporting the relationship of working memory with attention, the latter is considered as a single content-general resource. Here storage and processing exist in competition regardless of mutually shared content or the lack of it. Based on this assumption, it can be predicted that working memory storage and processing demands from non-shared contents are combined and result in dual-task costs (Chein, Moore, & Conway, 2011; Morey & Bieler, 2012; Vergauwe, Barrouillet, & Camos, 2010). Additionally, the capacity of working memory is limited by attention. Research has also supported the notion that storage and processing are in competition for central processing capacity and "the extent to which maintenance in working memory is impaired by concurrent processing is a monotonic function of executive control load, defined as the proportion of time during which central attention is engaged by the processing demand" (Barrouillet et al., 2007, as cited in Oberauer, 2019, p.3).

An issue with the assumption of a shared attentional resource for storage and processing is that dual-task cost reduces considerably during the initial few seconds of the retention interval (Vergauwe, Camos, & Barrouillet, 2014; Souza & Oberauer, 2019), even though working memory load diminishes the efficiency of concurrent response-selection tasks, and thus it is indeed not always noticeable in case of an unfilled few second period between memory set encoding and the beginning of the target processing task (Hazeltine & Witfall, 2011). This problem steered further research (Klapp, Marshburn, & Lester, 1983) into the notion of shared attentional resource for storage and processing. To maintain this proposal, it would need to be assumed that the attentional resource demand of maintenance diminishes gradually to an insignificant level within a period of a few seconds. This would also be in line with the assumption that a central processing resource is indeed needed for short-term information consolidation in working memory (Ricker & Hardman, 2017) but at the same time incompatible with another assumption of resource being required for maintenance during the retention interval.

The concept of shared resources for both storage and processing, as discussed earlier, can be divided into two options: the first variant proposes that for a representation to be active in working memory, there needs to be a resource assigned to it. However, the same resource is also required for some basic executive control processes to take place. This variant is not without limitations. It has been claimed (Jonides, 2008) that maintenance and processing compete with others at the same time for a shared resource until the completion of the task. Once this task is finished, the entire resource can be assigned again to the representations in working memory.

The second variant offers another explanation. Here, maintenance processing and executive control operations share this limited attentional resource (Barrouillet, Bernardin, & Camos, 2004). Moreover, this proposal is based on the assumption that representations in working memory decay if they are not refreshed. Oberauer and Lewandowsky (2014) state that the idea of decaying representations is very likely not valid, at least when verbal stimuli is concerned.

According to Oberauer (2019), one should prepare for substantial dual-task costs if working memory and attention are believed to share a limited resource, and both the demand for attention control and the maintenance of working memory are combined. Research that implies such dual-task costs include studies such as the Flanker task in which irrelevant (distractor) stimuli compete for selection when working memory load increases (Kelley & Lavie, 2011; Lavie, Hirst, de Fockert, & Viding, 2004). However, it needs to be emphasised that only verbal working memory load leads to an increase in the Flanker effect, and this has not been observed for visual load. In fact, with visual load the opposite effect has been noted (Konstantinou, Beal, King, & Lavie, 2014; Konstantinou & Lavie, 2013). One explanation has been proposed to account for this dissociation, namely, the fact that visual load triggers a visual perceptual resource – and in fact increasing information load on perceptual resources has been evidenced to reduce Flanker interference (Lavie, 2005). To the contrary, verbal working memory is reliant upon rehearsal to maintain target representations, and it competes with the visual attention control for a shared attention control resource. This assumption has been questioned by other specialists in the field proposing that rehearsal does not need any such resource (or very little if it does in fact) (Baddeley, 1986; Camos, Lagner, & Barrouillet, 2009).

Indeed, maintaining representations in working memory makes individuals more distractible. However, it has been suggested that different types of working memory load impact the attentional process of selection depending on whether working memory load is in line with mechanisms engaged in target or non-target (distractor) processing. Kim, Kim, and Chun (2005) investigated the effect of simultaneous working memory load in a series of Stroop tasks (tasks which examine executive control and inhibition/attentional control, see Chapter One for a more detailed description of the task). Results revealed that when there was an overlap between the type of working memory load and the type of target information necessary for the task, Stroop interference increased. However, when the type of information changed to distractor processing, Stroop interference decreased. Based on the findings, the authors (Kim, Kim, & Chun, 2005) proposed that parallel working memory load does not necessarily lead to an executive control impairment; individual's performance is rather impacted by the way working memory load and target/nontarget information overlap. The results highlight how dissociable components of working memory interact with perception and executive control.

1.5 Summary

The aim of the Chapter was to systematically review the current state of knowledge regarding executive control and working memory. Thus, key theoretical frameworks for each of the listed domains were discussed here and their strengths and limitations were presented.

The first section contributed to explanations of executive control by presenting various attempts to define this phenomenon. Here, the influential Unity and Diversity model (Miyake & Friedman, 2012) was discussed in more detail. The subsequent section comprised an introduction to memory definitions presenting the multi-store and unitary-store models of memory. Working memory models were presented and the similarities and differences between the various accounts proposed over the years were reflected upon. Finally, the relationship between working memory and executive control was presented.

The following Chapter investigates long-term memory mechanisms and introduces two types of measures of episodic memory, paired associate learning tasks and verbal fluency, that are also discussed in more detail.

CHAPTER TWO

Long-Term Memory and Paired Associate Learning

2.1 Long-Term Memory Systems

The previous Chapter discussed executive control and memory. From the evidence presented in the previous Chapter, executive control and working memory are separable processes, however, working memory is implicated in executive control processes (Engel, Kane, & Tuholski, 1999). Executive control processes are not only linked to short-term memory, but they are also implicated in long-term memory mechanisms particularly from the point of view of how they interact with episodic memory processes as well as the ability to learn associations, integrate them into long-term memory, and retrieve at a later time.

Long-term memory storage and short-term memory storage vary greatly in terms of their capacity and their ability to hold information (James, 1890). Long-term memory storage is composed of declarative and procedural memory. *Declarative memory*, also known as explicit memory, refers to one's ability to store and retrieve general knowledge and personal information (Baddeley, 1995). It comprises memory for personal events, so called episodic memory, and semantic memory for facts. *Procedural memory*, in contrast to declarative memory, encompasses skills and is often defined as the result of motor, perceptual or cognitive learning (Lechevalier & Habas, 2021). Moreover, procedural memory is acquired by means of practice and repetition and is commonly known as muscle or body memory. Also, as skills are deeply embedded, procedural memory performs without awareness and is thus commonly referred to as implicit memory (i.e., riding a bike, ice skating). All the types of long-term memory stores are indeed related and interlinked (Fuster, 2003).

Craik and Lockhart (1972) propose that the information that reaches long-term memory is arguably dependent upon attentional and perceptual processes that take place during learning. The deeper the analysis of the meaning of the target, the deeper the stimulus processing level. Indeed, what happens at learning significantly impacts long-term memory (Roediger, 2008). Two processes, learning and remembering, have been suggested to be considered as by-products of a combination of perception, attention, as well as one's comprehension skills.

Of interest to this thesis are declarative long-term memory mechanisms, specifically episodic memory, which is focused on a conscious recollection of personal experiences.

2.2 Episodic Memory

Episodic memory, a term first introduced by Tulving (1972), is a long-term memory system that is best described as "a recently evolved, late-developing, and early-deteriorating past-oriented memory system, more vulnerable than other memory systems to neural dysfunction" (Tulving, 2002, p. 5). In other words, episodic memory enables an individual to access mental information about the location and timing of personal (biographical) events in their life. It differs from semantic memory by concerning personal experiences while semantic memory is claimed to be detached from personal happenings and thus is conceptual, generalised, and not linked to any actual experience (Binder & Desai, 2011). Arguably, some believe that episodic and semantic memory constitute two separate memory systems, whereas others propose that they are interdependent.

Episodic memory store has been usually assessed by means of two types of measures tapping recognition and recall. In basic recognition measures, participants are presented with stimuli (i.e., words or pictures) and asked to decide whether they had already seen the items in the previous trials. When considering recognition memory, there are two distinct processes (i.e., recollection and familiarity). *Recollection* is a process in which an individual can recognise an object taking into consideration various specific details regarding its context. *Familiarity* refers to the process of recognising something on the basis of its perceived memory rather than any specific contextual details (Diana, Yonelinas, & Ranganath, 2007). In such a setting, target items (usually words) that were presented to participants on a list are then mixed with non-targets that were not shown. The items are randomly presented to participants one by one, and participants are asked to provide an *old item* or *new item* response. Arguably, items can be recognised by employing either recollection or familiarity process (Mandler, 1980). The difference between the two is that recollection comprises conscious retrieval of contextual details that a person made associations with at encoding, whereas familiarity refers to the process of knowing (being sure) that the item was presented as a target without any additional and specific information about it. This approach has been indeed supported by a number of dual-process frameworks of recognition memory; however, it was met with scepticism and argument that recollection and familiarity can in fact rely on the same factor (Diana, Van den Boom, Yonelinas, & Ranganath, 2011). Dunn (2008) conducted a meta-analysis of 37 studies that investigated whether *remember* (recollection) and *know* (familiarity) responses reflect various mechanisms or whether they are underpinned by the single mechanism. Results showed that both recollection and familiarity can be explained by a single factor of memory strength.

Interestingly, Dunn's findings are not in line with the event-related potential research (Addante, Ranganath, Olichney, & Yonelinas, 2012) which suggests that recollection and familiarity are associated with different event-related potential. Moreover, Diana, Yonelinas, and Ranganath (2007) indicated by neuroimaging research that both recollection and familiarity are associated with activation in different brain areas. In essence, the process of recollection seems to require an access to more information, mainly contextual, which is not observed for familiarity.

Recall measures, on the other hand, include retrieving targets from memory. There are three common types of recall tests, namely, (1) free recall, where participants are asked to produce the names without any stimuli presentations; (2) serial recall, in which items should be produced in order of their presentation, and (3) cued recall, where cues are presented and participants' response to them expected. In order to investigate whether recognition and recall are similar processes that are based on similar mechanisms, Staresina and Davachi (2006) conducted an experiment in which they employed three different memory assessments: free recall, object recognition (tapping familiarity), and associative recognition (tapping recollection). They found an increased activation in brain areas (left hippocampus and left ventrolateral prefrontal cortex specifically). Indeed, the activation was higher for the free recall performance and much lower for object recognition tasks. They argued that in this experiment free recall comprises item-colour association formation which is not needed in recognition memory.

Short-term memory and long-term memory are believed to be linked together by means of the episodic buffer. The episodic buffer, as presented earlier in Chapter One, refers to a limited-capacity storage system located within working memory that is in charge of integrating information from different sources (Rudner & Rönnberg, 2008). Its main function is to process incoming input and store it temporarily. The buffer is believed to be controlled by the central executive and it plays a crucial role in information transfer into and from episodic long-term memory (Baddeley, 2000).

Schrauf and Rubin (2000) propose that during the process of retrieving information about an important event, people often tend to retrieve some aspects of such an event linguistically. Thus, researchers suggest that language potentially impacts episodic memory performance (Loftus & Palmer, 1974). Fernandes, Craik, Bialystok, and Kreuger (2007) asked participants to listen to words and recall as many of them as they were able to. In fact, as bilingual participants demonstrated weaker performance by recalling fewer items than monolinguals, Schroeder and Marian (2012) propose that bilingual episodic memory is at a disadvantage when decoding and encoding purely linguistic information. Bilinguals have been evidenced to possess smaller vocabularies in both of their lexicons and are slower in picture-naming tasks (Bialystok, Craik, Green, & Gollan, 2009). However, Bialystok, Craik, Klein, and Viswanathan (2004) found that bilingual non-linguistic recall is superior to monolinguals' which may be explained by the bilingual advantage in cognition stemming from bilinguals' need to switch between two languages (see also Costa, Hernandez, & Sebastian-Galles, 2008).

It has been also proposed that the bilingual advantage may be reflected in better performance of episodic memory. Bilinguals have been found to exhibit an advantage in executive control arguably due to their extensive experience of managing two languages (Costa, Hernandez, & Sebastian-Galles, 2008). As executive control is closely linked to episodic memory (i.e., it is engaged via controlled searches through one's memory), it can be proposed that enhanced bilingual executive control may lead to enhanced episodic memory performance in bilinguals (Schroeder & Marian, 2013). Schroeder and Marian (2012) employed a picture recall task conducted by with mono- and bilingual participants, a linguistically-reduced version of the experiment (with a number of strategies employed to encourage visual encoding and discourage employment of verbal remembering), in which participants were instructed to remember a number of pictures presenting complex scenes with no information that they would be later asked to recall what they saw. Results revealed that bilinguals recalled more pictures than their monolingual counterparts with factors such as early acquisition and more extensive bilingual experience associated with better episodic memory performance.

Additionally, Ullman (2001) proposed that the medial temporal lobe system which is crucial for episodic and spatial memory can be improved by bilingualism. Since second language acquisition involves the engagement of this system, bilinguals might exhibit its improved functioning and thus enhanced episodic memory. Yet another proposal suggests that the bilingual advantage in episodic memory experiments may stem from the fewer number of memories being encoded in their second language (as they were encoded in their first language) thus leading to a more effective retrieval due to decreased competition between memories (Ullman, 2001).

Indeed, bilinguals seem to demonstrate enhanced episodic memory performance when non-verbal stimuli are involved as well as decreased age-related cognitive decline. With this in mind, bilinguals diagnosed with Alzheimer's dementia tend to exhibit memory-related impairments much later than monolingual counterparts, which suggests that bilingualism may contribute to memory improvements and delays in cognitive decline (Bialystok, Craik, & Freedman, 2007).

There is growing evidence within the aging literature that episodic memory mechanisms remain more intact in bilinguals and this is indeed demonstrated in studies which looked at how bilingualism modifies aging. For instance, Bialystok, Craik, and Freedman (2007) looked at episodic memory longitudinally and found that those mechanisms were preserved in bilinguals as they got older. Studies to date have largely focussed on executive control as opposed to episodic memory more generally and thus there is a precedent to investigate it in more detail.

As discussed in this section, recognition and recall play an important role in episodic memory assessments and they tap into different underlying executive control mechanisms. The next section introduces the paired associate learning paradigm and explains how it is linked to episodic memory and word learning.

2.3 Paired Associate Learning

The classic paired associate learning task (Papagno & Vallar, 1995) is an episodic memory paradigm which includes learning as well as remembering associations between stimuli (for instance novel word and novel object) that are paired randomly. In a paired associate learning task, participants are shown the pairings (i.e., flower-table) and at test they are instructed to recall the word from the pair based on a presented cue (i.e., flower-?). Thus, it enables examining both short-term memory and long-term memory storages as it requires accessing and recalling lexical items over a period of time. There are two possible stimuli combinations used in paired associate learning tasks (Figure 2.1): unimodal associations where both stimuli belong to one type (either visual and visual stimuli; i.e., two novel objects; or verbal-verbal with two novel words), or cross-modal where the stimuli used is from two distinct modalities (i.e., visual-verbal stimuli; novel word and a novel object).

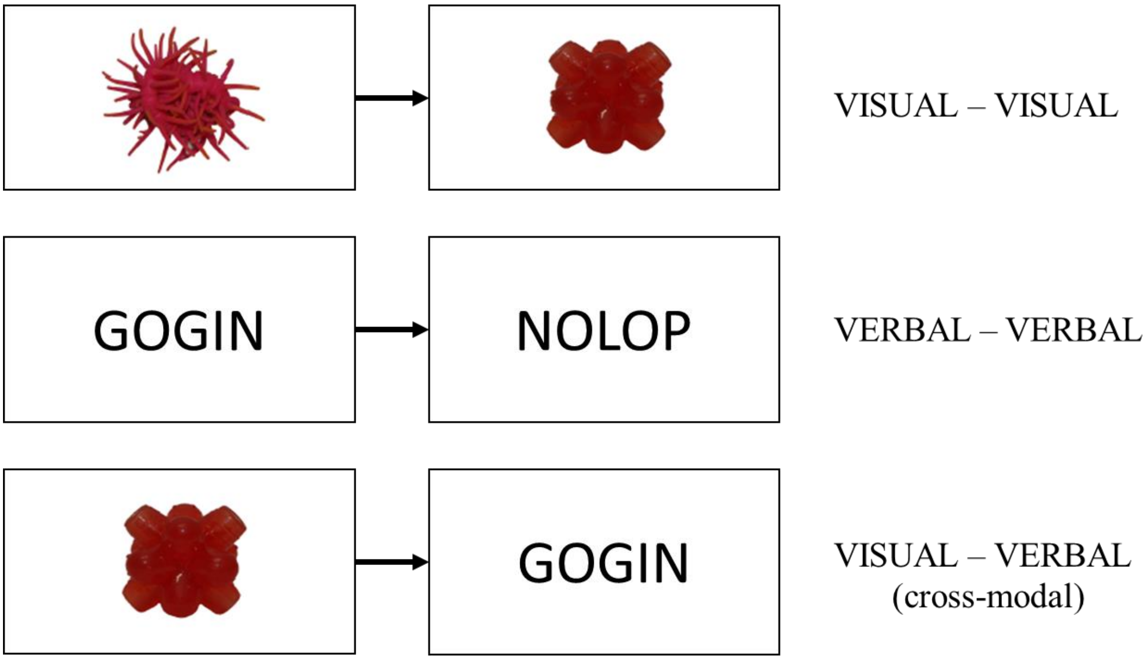


Figure 2. 1 An example of stimuli combinations in a paired associate learning task

The paradigm is often applied in novel word learning experiments (Reynolds & Romano, 2016) as it resembles the way we learn new labels for objects.

It can be concluded that novel word learning is a typical example of paired associate learning where participants learn a novel spoken word for a visual referent (novel object). Individual differences in novel word learning can also be measured by employing various strategies (looking at how many novel words participants can recognise and how many they are able to recall the target word). These two assessments, although both used equally often in research, rely on different mechanisms, and differ substantially in terms of the specific systems that they use to complete the task (Reynolds & Romano, 2016). By examining one's ability to recall a target word, their ability to retrieve semantic information of a given picture from their memory storage is assessed. Additionally, the phonological representation of the target needs to be activated in order to enable its articulation. Recognition, on the other hand, relies strongly on decoding the orthographic form of a given word. Once the corresponding item has been found in one's mental lexicon, it gets activated and directed to its semantic representation. In this sense, recall is a more effortful process and thus individuals' performance on recall tends to be worse than on recognition.

An immediate pairing created between a novel object and novel word (so called fast mapping), has been widely researched in children (i.e., Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992) and has been claimed to play an important role in early word learning. Horst and Samuelson (2008) put this statement to the test in a series of four experiments with 2-year-old toddlers. Consistently with prior research findings, children were able to successfully choose a novel object when surrounded by familiar objects and hearing a novel word. However, they were unable to name these novel objects 5 minutes later. Although participants correctly identified novel objects as referents to novel words, this did not seem to lead to automatic success in learning and retaining the names of these objects. This stands in contrast to the assumption that a successful initial selection of the referent results in a successful word acquisition and its retention. Pomper and Saffran (2019) suggest that presenting a child with a familiar object when learning a new word can in fact hinder vocabulary acquisition. The fact that children tend to choose a novel object for a novel word may be a result of "a preference for novelty, socio-pragmatic reasoning, or a combination of all three" (p. 246). Nevertheless, Horst and Samuelson (2008) indicated that the correct selection of novel referent does not imply that a child will successfully retain and retrieve the word-object pairing for future use. Establishing appropriate links between novel referents and novel words, such as in a paired associate learning task, can be a gradual and lengthy process.

It has been suggested that humans learn new words by relying on associative learning mechanisms (i.e., Pavlov, 1927) like the ones that can be observed in some animals. Similarly, to dogs associating the sound of the bell with food, we are believed to utilise associations between new lexical items in the very same way. Although in vocabulary acquisition other crucial skills come into play, such as social context, knowledge of syntax and pragmatics among others, associative learning has been proposed to account for at least some stages of vocabulary acquisition (Merriman, 1999; Plunkett, Sinha, Moller, & Strandsby, 1992), in particular, the initial ones when learning occurs slowly before the age of two years old. After this age, children start becoming quicker and more effective in learning new words. However, a number of accounts propose that learning a new word (i.e., novel word learning) cannot be an associate learning process and that the underlying mechanisms are in fact social (Singer, Golinkoff & Hirsh-Pasek, 2006), referential or conceptual (Waxman & Gelman, 2009), or based on constraints (Woodward & Markman, 1998). The evidence accumulated in favour of these accounts is not entirely obvious as these terms are very often not defined clearly. Indeed, the initial stages of learning a novel word differ significantly from the subsequent ones as both lexical and social skills gradually develop in children – this suggests that novel word learning may in fact be associative (Nazzi & Bertoncini, 2003).

Miller and Unsworth (2020) explains that paired associate learning measures have been found to correlate with each other and thus constitute a general paired associate factor that has also been observed to be linked with other long-term memory factors. From the perspective of this thesis, it is of great importance to consider the relationship between the paired associate learning paradigm and long-term memory mechanisms.

Tirre (1991) investigated the paired associate learning paradigm in 714 army recruits and found that both fluid and crystallised intelligence are predictors of performance on the task. Moreover, almost 68% of the variance in paired associate learning performance was accounted for not only by both the types of intelligence, but also working memory, and learning strategies employed by participants. Kyllonen, Tirre, and Christal (1991) tested over 2000 participants in a series of four experiments. They found that general knowledge (i.e., knowledge of a wide range of facts from various fields and subjects) was a strong predictor of long-term memory and was also related to paired associate learning. Thus, they concluded that individual differences in using general knowledge to create associative links between target items can in fact account for the variation in paired associate learning task performance. Indeed, general knowledge as well as processing speed were found critically attributable to individual differences in paired associate measures.

Kyllonen and Tirre (1988) also investigated individual differences in terms of recall ability in the paired associate learning task in a large sample of 710 military recruits aged 16 to 23 years old. In their study, they employed a paired associate learning task including 13 pairings, and tested participants on their short-term memory, verbal learning ability, and reasoning ability. They found that participants who were fast learners were also performing better on a series of memory measures thus suggesting that individual differences in performance on the paired associate learning task can be accounted for by general associative learning proficiency. General learning speed, which can be predicted by general knowledge, has been also found to predict word retention and relearning. This is an interesting result indicating that learning speed is a predictor of one's retention abilities. Additionally, participants who learned the pairings faster, were also better at recalling them later on.

Similarly, Zerr, Berg, Nelson, Fishell, Savalia, and McDermott (2018) further investigated individual differences in the ability to learn and recall pairings in the paired associate learning task in a longitudinal task of 281 and 92 participants. Results revealed that participants who learned pairings faster, were better at retention across days. Thus, Zerr et al. propose that participants who were more efficient at learning were also characterised by faster processing speed, enhanced memory performance, as well as higher scores on intelligence measures. They also suggest that one mechanism that is potentially accountable for efficient learning is inhibition/attentional control. Indeed, better learners are believed to be more efficient at allocating attention when they learn new things. Unsworth and Spillers (2010) also found that participants who are better at allocating their attention to task-relevant information are less affected by interference and demonstrate less forgetting. As discussed in Chapter One, inhibition/attentional control is closely linked to the capacity of working memory and thus impacts the efficiency of one's learning ability.

By means of a paired associate learning task, Miller and Unsworth (2020) examined the consistency and intensity of attention. Consistency refers to one's ability to consistently allocate attention to a target task whereas intensity describes the amount of attention to the task of interest. Indeed, these two aspects have been found to be crucial predictors of one's working memory and their attention control (Miller & Unsworth, 2020). Miller and Unsworth examined the intensity of attention via pupillary responses and found that the best performing participants on the paired associate learning task are those who directed more attention to the target objects during the encoding phase than those who devoted less attention to the targets. Additionally, Seibert and Ellis (1991) found that more successful learners are those who do not demonstrate tendencies to mind wandering during the important phase of encoding information (also Xu & Metcalfe, 2016). More crucially, pupillary responses have been supported by research findings as a reliable indicator of the intensity of attention paid to the target stimulus (see Kahneman 1973 for more detail). Pupil dilation in participants as a measure of the intensity of attention allocation has also been employed in other attention-demanding tasks (i.e., Hess & Polt, 1964 – the description of mathematical problems). Taken together, research suggests that target objects that receive more attention at the encoding stage – as measured by pupil dilation - are indeed better recalled at a later stage.

It is however not of a rarity that mind wandering instances occur during experiments and the amount of attention devoted to the task decreases as a result of off-task thinking. Of particular interest to this study is the fact that attentional lapses seem to occur often during encoding (Unsworth & Miller, 2017, Unsworth, McMillan, Brewer, & Spillers, 2012). Garlitch and Wahlheim (2020) also found that they are associated with weaker performance of working memory on subsequent trials. Individual differences in learning abilities as well as memory predispositions lead to differences in performance on attention-related tasks due to varying degrees of individual susceptibility to mind wandering. Thus, according to Seibert and Ellis

(1991), participants who demonstrated the ability to consistently maintain their attention on the target object were also found to possess superior memory skills compared to those individuals who were not so efficient at this task. Xu and Metcalfe (2016) replicated the study and proposed that consistency of attention is a crucial factor for being a successful learner. Craik, Govoni, Naveh-Benjamin, and Anderson (1996) found that participants tend to perform weaker on memory tasks in situations where their attention is divided and not fully devoted to the task of interest. Collectively, these research findings indicate that learning is a process that requires a significant amount of attention.

One of the very early studies that employed paired associate learning task was conducted by Papagno and Vallar (1995). They proposed that the individual differences in phonological working memory impact the easiness of learning foreign languages and indicated that the multilingualism may confer memory enhancements. Bilinguals tend to outperform monolinguals when paired associate learning tasks are employed in their dominant language (Kaushanskaya, 2012; Kaushanskaya & Marian, 2009; Kaushanskaya & Rechtzigel, 2012; Bogulski, Bice, & Kroll, 2018). Interestingly, these results do not seem to be associated with either the degree of similarity between the languages of a bilingual person or their familiarity with second language phonology. It has been argued that enhanced bilingual performance on paired associate learning might be linked to better phonological working memory or enhanced phonological long-term memory (Papagno & Vallar, 1995; but see Kaushanskaya, 2012). This suggests that bilinguals may develop a more general cognitive enhancement which leads to more efficient encoding and retrieval ability. Kaushanskaya and Rechtzigel (2012) suggest that the bilingual cross-language activation results in a wider activation of lexico-semantic system than in monolinguals which in turn leads to a more efficient novel word learning. Alternative explanation proposed by Kaushanskaya and Marian (2009) argues that bilinguals are better at allocating their selective attention to task-relevant information and are less affected by interference, and that enhanced bilingual executive control results in more efficient learning in general (Kaushanskaya, 2012). Indeed, bilinguals may employ more efficient ways of encoding new information and have greater skills of making novel associations which is directly linked to their bilingual experience.

There is a growing body of evidence that demonstrates that there is a bilingual advantage in paired associate learning and thus there is precedent to examine it further. Bilingual research to date has focused largely on explicit memory mechanisms while much less is known about the implicit memory mechanisms. Language learning is in some ways a form of implicit learning and is largely dependent on unconscious knowledge. Paired Associate Learning task is a task that resembles real-life language learning. In this task, participants know that they learn the novel pairs but they are not explicitly aware of how they do it. Of the major interest to this thesis is the question of how these novel associations are integrated into long-term memory and retrieved from there. As discussed earlier in this Chapter, retrieval processes indicate how well newly acquired information has been integrated and consolidated into memory. These mechanisms are assessed especially when people get older and allow understanding of how long-term memory and executive control degrade over time (i.e. Paired Associate Learning tasks are often used in assessments of mild cognitive impairment). Bilingual literature focuses largely on the immediate short term memory tasks and studies tend to employ a specific task. What is more, it also reveals difficulties in defining cognition as well as the lack of understanding of its diversity. This thesis aims to investigate a much broader aspect of cognition and employs a suite of various measures that tap into various cognitive aspects. As research employing paired associate learning tasks is still rather limited, it warrants further investigation.

2.4 Verbal Fluency

Verbal fluency is another measure commonly used to assess individual differences in long-term memory mechanisms as it involves the lexical access and the retrieval of lexical information from episodic memory. Verbal fluency measures one's ability to produce as many unique words as possible that fall into a given category within a period of one minute and it can be divided into two types: letter fluency that requires production of words that begin with the same letter, and semantic fluency, where participants are asked to produce words that fall into the same category, i.e., animals, clothes, things to eat. The final score is counted by the number of words produced correctly.

Verbal fluency is often employed in research as a measure of executive control and long-term memory mechanisms (see Figure 2.2). By asking individuals to produce words that meet the task requirements, one is able not only to assess their executive control, but also test the verbal ability comprising lexical knowledge and ability to retrieve lexical representations from episodic memory. Indeed, the process of word production in verbal fluency is underpinned by executive control processes (Fisk & Sharp, 2004) and has also been suggested to constitute a reliable screening tool to measure general verbal functioning (see Table 2.1).

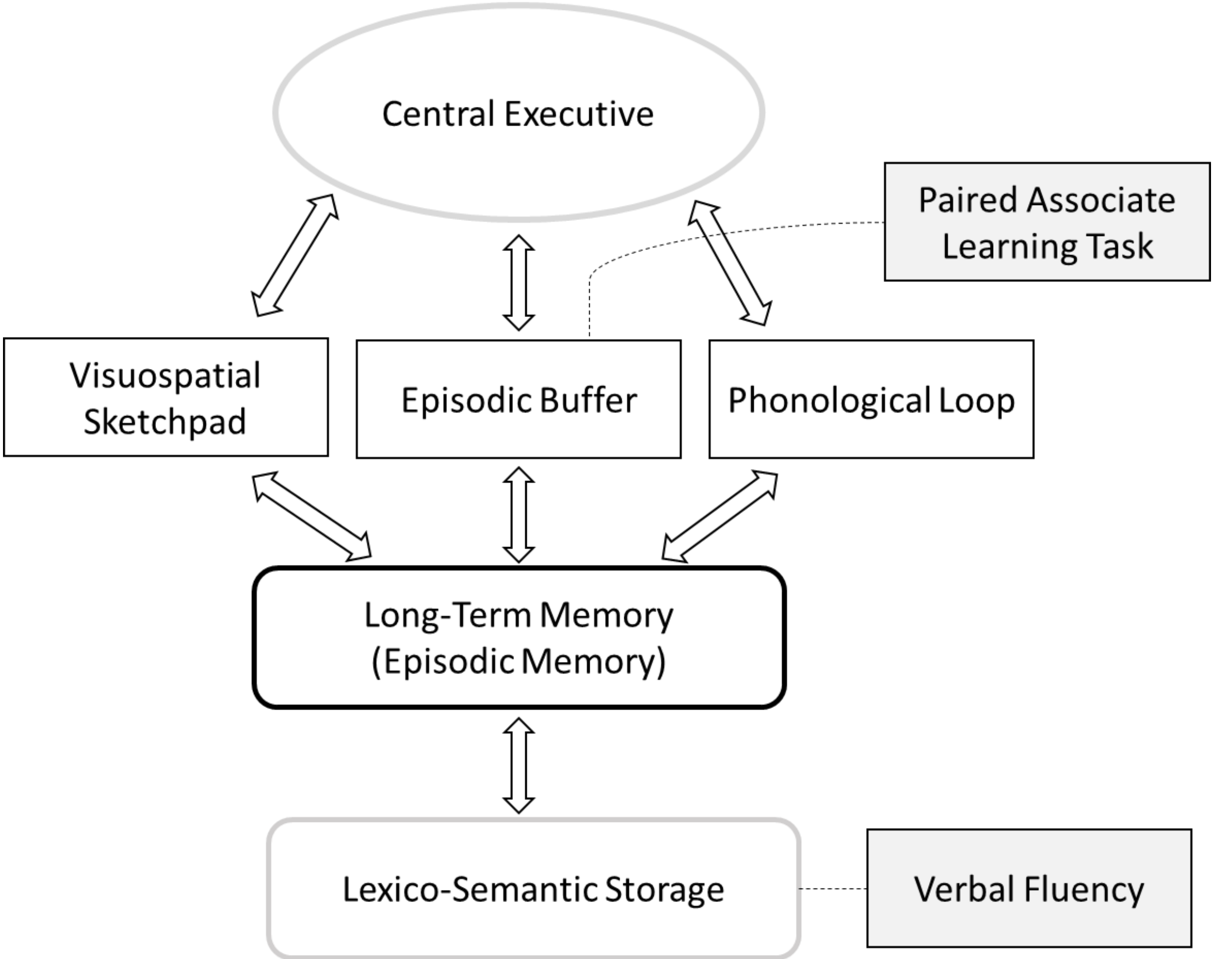


Figure 2. 2 The relationship between long-term memory, executive control, paired associate learning, and verbal fluency (adapted from Vandek, Gabrić, Kužina, Erdeljac, Vlasta, & Sović, 2018)

When producing target words in verbal fluency, participants need to follow a number of steps: first, they need to retrieve words and to perform this task they need to access available representations present in their long-term memory. Additionally, while staying focused on the task at the same time, participants need to choose words that meet certain criteria (depending on the type of the fluency task) and avoid repetition and production of words that count as errors (i.e., proper names). Taking this into consideration, any verbal ability or executive control impairments are likely to manifest themselves through the usage of fluency tasks with poor performance indicating serious executive control deficits (Shao, Janse, Visser, & Meyer, 2014).

Table 2. 1 Verbal fluency analyses and their contribution to language and executive control (adapted from Patra, Bose, & Marinis, 2020)

	Description	Language	Executive control
Quantitative analysis			
Number of correct answers	Number of correctly produced words (excluding errors) is indicative of word retrieval abilities.	✓	✓
Time-course analysis			
First response latencies	Preparatory time taken to initiate the first response.	✓	
Subsequent response latencies	Estimated mean retrieval latency that is indicative of the time point at which half of the total responses have been generated.		✓
Slope	The shape of the curve that reflects how resources are monitored and deployed over time during retrieval. It is largely determined by executive control.		✓
Qualitative analysis			
Cluster size	Ability to generate words within a subcategory that relies on accessing mental representations of words within subcategories.	✓	
Number of switches	Ability to shift efficiently to a new subcategory when a previous one is exhausted.		✓

Verbal fluency tasks are commonly used in various fields. For instance, as a neuropsychological measure in assessing individuals with attention deficit hyperactivity disorder (Andreou & Trott, 2013), as well as executive control disorder such as Parkinson's disease for instance (Pettit, McCarthy, Davenport, & Abrahams, 2013). There is a great deal of research that employs verbal fluency measures to examine individual differences in executive control ability, more precisely, set-shifting, memory updating, and inhibition/attentional control. For instance, Mahone, Koth, Cutting, Singer, and Denckla (2001) tested children diagnosed with ADHD and found they tend to score lower when compared to typically developing peers. Takács, Kóbor, Tárnok, and Csépe (2013) replicated such findings with a group of children aged 8 to 12 that were either typically developing (N = 22) or had a diagnosis of attention deficit hyperactivity disorder (N = 22). Cohen, Morgan, Vaughn, Riccio, and Hall

(1999) examined the performance on verbal fluency measures in 130 children aged 6 to 12. They found differential performance in a group of typically developing children, and children with attention-deficit hyperactivity disorder and two subtypes of dyslexia. Verbal fluency tasks were proposed to be clinically useful in differentiating between these conditions. Also, studies with participants characterised by impaired frontal brain lobes demonstrated the association between that damage and poor verbal fluency performance (Schwartz & Baldo, 2001).

However, what is less clear, is the question of which executive control components have the biggest impact on one's performance on fluency measures. To successfully complete the fluency task, the subject needs to hold both the instructions and their previous responses in their working memory while also aiming to suppress irrelevant representations and repetition. One tactic that people tend to employ during verbal fluency performance is clustering, in other words, producing series of words in succession that are somehow related (i.e., when listing animals, one may start with farm animals [hen, chicken, pig, cow, horse], and then move to another cluster of animals [i.e., tiger, lion, zebra, etc.]). In fact, clustering ability relies heavily on a systematic memory. Another strategy on the other hand, so called switching, refers to the ability to transition from one cluster to another. According to Henry and Crawford (2004), one's performance on verbal fluency tasks is reflective of their working memory (Rosen & Engle, 1997), as well as inhibition/attentional control (Hirshorn & Thompson-Schill, 2006) and effortful self-initiation (Shao, Janse, Visser, & Meyer, 2014). However, other proposals comprised the importance of one's ability to switch (Abwender, Swan, Bowerman, & Connolly, 2001) which may depend on the efficiency of response suppression mechanisms (Hirshorn & Thompson-Schill, 2006). Interestingly, research has also found verbal fluency linkage to fluid intelligence (Roca et al., 2012).

Indeed, although looking relatively similar, both types of verbal fluency tasks differ in important ways. The semantic fluency task relies on existing links between concepts (words) that are related in some way, and in this aspect, it is based on mechanisms that are employed during daily production tasks such as a shopping list preparation. In opposition to this stands the second type of the task, phonemic fluency, which is not often utilised on a daily basis; here, for the relevant words to be retrieved, one needs to first suppress other semantically related words that get activated automatically and they must stick to the novel strategies to retrieve the targets (Luo et al., 2010; Katzev, Tüscher, Hennig, Weiller, & Kaller, 2013).

Indeed, verbal fluency tasks enable the assessment of participants' word retrieval speed, vocabulary size, and executive control performance (Luo et al., 2010). Luo et al. explain that the role of executive control and verbal ability in verbal fluency measures can be investigated not only by the number of the correct responses that participants produce, but also by the timing of these responses. They distinguished two types of mean retrieval latencies (Rohrer, Wixted, Salmon, & Butters, 1995): first response and an average of subsequent responses. First response timing is measured as the time interval calculated from the beginning of a task until the onset of the very first word that the subject produces, whereas the mean subsequent response is measured by averaging the onset of the first response and the beginning of each following word produced. It is worth mentioning that the mean latency does not represent the retrieval speed. The length of the mean latency is in fact indicative of the rate of recall with a short mean latency suggesting a fast-declining rate: this is explained by the higher number of words being indeed retrieved and produced at the onset of the trial. A long mean latency, on the other hand, is often characterised by either a slow decline in word retrieval or a rather slow production of subsequent words throughout the verbal fluency task.

Another possibility of investigating verbal fluency is via slope analysis. Slope refers to the shape of the curve and represents the decay occurring over time. It is associated with an increased use of controlled processes in verbal fluency that appear towards the end of the task. In fact, the beginning of the task relies more on automatic processes that allow word spurts and are often indicative of one's vocabulary size (Luo et al., 2010). Overall, slope represents the dynamic processes engaged in one's ability to monitor resources over time which are largely determined by their executive control. As Perret (1974) explains, "as a trial progresses, the initial resources deplete as a function of increased processing demand, such as monitoring, searching, and the need to resist the interference from a habitual semantic-based word searching strategy for letter fluency in particular" (as cited in Luo et al., 2010, p. 33).

Luo et al. (2010) propose that executive control contribution to verbal fluency performance indeed increases with the length of the trial as later in the task participants (1) need to bear in mind the words that they have already produced, and thus (2) suppress interference from the earlier responses, as well as (3) employ the ability to switch between word subgroups (i.e., exhausting the list of pets and starting to recall farm animals).

Clinical research employing verbal fluency tasks have yielded interesting results in terms of letter versus category type of the task. Meijer, Schmitz, Nieman, Becker, van Amelsvoort, Dingemans, Linszen, and de Haan (2011) examined 37 individuals with psychosis, certain types of focal lesions, or Alzheimer's disease and found that they demonstrate more significant impairments in the latter type of task than the former one which is in line with the deficient access to semantic information. Overall, neuroimaging studies with typically developing subjects demonstrate that there are in fact two brain circuits that overlap in the two verbal fluency tasks. Indeed, Katzev et al. (2013) found an association between the semantic fluency task as well as activation more anterior-ventrally while phonemic fluency task was associated with an activation more posterior-dorsally in left inferior frontal gyrus. Shao et al. (2014) proposes that this means that verbal ability seems to be reflected in semantic verbal fluency measures whereas executive control ability in phonemic fluency tasks.

Although verbal fluency measures have been found to be reliable assessments of one's verbal ability and their executive control skills (Ettenhofer, Hambrick, & Abeles, 2006), it needs to be noted that these scores do not constitute a pure measure of either of them. Ettenhofer et al. tested 118 older adults on a series of executive control measures, including two verbal fluency tests, and found modest reliability of individual measures employed. This needs to be taken into consideration when employing verbal fluency tasks to examine individual differences in verbal fluency tasks and executive control skills.

2.5 Summary

Chapter Two explained the theoretical underpinnings of long-term memory systems with more in-depth presentation of episodic memory. It also explained the links between short-term and long-term memory storages and presented the differences between recognition and recall measures that are often used to investigate episodic memory performance. Additionally, differences in monolingual and bilingual performance on episodic memory tasks were presented in more detail. The Chapter then focused on the introduction and presentation of the paired associate learning paradigm and its importance in studying novel word learning. Moreover, paired associate learning was also presented from the perspective of associative learning and word learning and this section further explained how paired associate learning tasks are often employed by research to understand the structure of long-term memory mechanisms. Lastly, another measure of individual differences in executive control and memory, verbal fluency task was introduced along with relevant research findings. Thus, the next Chapter will review the evidence related to bilingualism, executive control, and working memory across the lifespan.

CHAPTER THREE

Bilingualism, Executive Control, and Working Memory

3.1 Conceptualising Bilingualism

The term *bilingualism* is often explained as the ability to communicate effectively in two languages (Bloomfield, 1956). As this phenomenon can be influenced by a number of factors, there has been an ongoing debate regarding the term *bilingual* and how it should be applied to relevant bilingual speakers. One variable that is often mentioned to facilitate different types of bilingualism is that there is a critical period for when we acquire language, typically referred to as *the critical period hypothesis.* Critical period hypothesis, popularised by Lennenberg (1967), refers to a period in life before puberty when it is arguably easier to acquire language and after which learning a new language is more effortful and challenging due to biological milestones. The younger the child starts to be exposed to two languages (or live in a country where two languages are used with the same frequency), be it at home or at school, the more proficient and competent language users they become. Children who are exposed to two languages (i.e., in home and/or school environments) are referred to as *simultaneous* bilinguals due to their simultaneous acquisition of two different linguistic systems. Those who start learning the second language later than their native language are commonly called *sequential* bilinguals as they have already mastered one language and begin to learn the second language subsequently. These two language acquisition pathways vary developmentally; as *simultaneous bilinguals* learn both languages at the same time, their pathway is similar to the way monolinguals learn their native languages (Kohnert, 2008). *Sequential bilinguals*, on the other hand, need to learn certain features that are language specific features and hence their pathway differs (August & Shanahan, 2006).

Kovelman, Baker and Petitto (2008) argued the existence of the critical period for the ability to achieve native-like second language. They indicated that being born in a country where the one majority language persists (i.e., Polish in Poland) does not necessarily mean that an individual is systematically exposed to this language. In fact, children might receive more exposure to another language that is in use within their family or home environment first and learn this language more efficiently. This results in a number of individual differences in terms of executive control, social, and language development of bilingual children. Additionally, such instances also lead to difficulty in labelling a person *bilingual* due to many ways of becoming bilingual. Factors such as various degrees of bilingual language proficiency (i.e., how well a person has mastered a language) as well as patterns of languages used at home (i.e., when both languages are spoken at home with each parent speaking a different language with the child, or only one language being spoken (either first language or the majority language)) need to be taken into consideration.

There are also additional factors that influence bilingualism such as learners' innate features (Chomsky, 1965), language learning settings (Skinner, 1957), along with a combination of both: environmental aspects and learner's individual characteristics (Vygotsky, 1986). Continued exposure to target language, levels of proficiency in both languages, as well as sociocultural factors have been found to impact the bilingual status (Calvo & Bialystok, 2014).

The next section will address the different theories of language control in bilinguals in terms of activating one language while having to suppress the other. It also presents a number of theoretical models that have been put forward, specifically the Bilingual Interactive Activation model, Bilingual Interactive Activation Plus model, and Inhibitory Control model.

3.1.1 Theoretical Models of Language Control in Bilinguals

The fact that there might be two independent mental lexicons, a separate one for each language of a bilingual person, does not imply that language-selective access should be taken for granted. In fact, there have been proposals that both mental lexical storages get activated non-selectively and simultaneously during word recognition until the input matches target representations (Grainger & Dijkstra, 1992). Costa, Miozzo, and Caramazza (1999) recruited Catalan-Spanish bilinguals (aged between 18 to 25 years) to investigate lexical selection by means of a series of picture-word interference tasks. The study comprised experiments with identical, semantically related, and phonologically related distractors. They found evidence that there is indeed a language-specific selection in bilingual lexical access meaning that only target language words are considered for selection and thus compete with each other. Additionally, it can be proposed that non lexical phonological processes play a role in the phonological facilitation effect. This is in line with the Bilingual Interactive Activation model (Costa et al., 1999) that makes the following assumptions:

> "(a) the semantic system sends activation in parallel and to equal extents to the lexical entries in the two lexicons of a bilingual; (b) only the lexical nodes in the lexicon which

is programmed for response are considered for selection; and (c) there are nonlexical mechanisms that allow a written word to activate its phonological segments" (p. 387).

Bilingual Interactive Activation model (Dijkstra & van Heuven, 1998) is an initial version of one of the highly influential models of bilingual word recognition. The model consists of a proposal of an integrated mental lexicon for both first and second language, as well as a non-selective lexical access mechanism. The Bilingual Interactive Activation model comprises three interconnected levels of representation, namely, letter, word, and language. The incoming input feeds forward starting from letter through word units to language. The visual input, namely the target word in the form of a letter string, activates certain features at each letter position which further activates letters that match these features. At the same time, a simultaneous inhibition of the rest of the letters takes place as they do not match the target features. In the next step, the selected letters further activate words in both languages of a bilingual speaker and inhibit all the irrelevant words at the same time. It has been proposed that at this level irrelevant words are inhibited regardless of whether they are first or second language vocabulary.

To further extend the initial model, the Bilingual Interactive Activation Plus model was proposed by Dijkstra and van Heuven (2002). The model comprises two systems, namely, the first one aims to identify a word, while the second system is so-called a task schema (task/decision) system. The former system is similar to the Bilingual Interactive Activation model but it has been enriched with phonology and semantics. Thus, it offers an explanation to cross-linguistic overlaps in terms of word orthography, phonology, and semantics. Language nodes serve a function of identifying the language to which the target word belongs to, but they no longer serve as language filters. Task schema was also implemented in the revised model as it sets the order of specific processing steps that need to be completed to perform a target task (Green, 1998). It also receives input coming from the word identification system.

Bilingual Interactive Activation Plus (Dijkstra & Van Heuven, 2002) assumes that there is an interconnection between words across the mental lexicons of a bilingual person. This interlanguage is indeed a basis for integration of first language and second language words and thus it enables further interconnections between the two languages of a bilingual person. The model explains that once representations are triggered in an individual's two languages (first and second language) at the same time, they are then mapped onto both phonological and semantic associations with the visual input. Comesaña, Ferré, Romero, Guasch, Soares, and

García-Chico (2015) recruited 20 Catalan-Spanish bilingual students to investigate the processing of cognates – words that share orthographic and phonological overlap, but which differ in meaning. They employed lexical decision tasks with identical, non-identical cognates, and non-cognates. Results from the study revealed that the type of word and language context play an important role in bilingual language processing. The degree of orthographic and phonological overlap was found to modulate cognate processing: the more similarly looking the words are, word recognition reaction times have been found to be much faster. Thus, the orthographic similarity of the words influences the extent of semantic activation as well as the accuracy and time of recognition (Comesana et al., 2015).

As discussed above, in terms of bilingual visual word recognition models, the Bilingual Interactive Activation (Dijkstra & Van Heuven, 1998; Van Heuven, Dijkstra & Grainger, 1998) and the Bilingual Interactive Activation Plus (Dijkstra & Van Heuven, 2002) are often employed as a reference framework for conducting and understanding further research in this field. Both models assume that visual presentation of a lexical item results in non-selective co-activation of word candidates that are similar to the target, from all available languages. Although both Bilingual Interactive Activation and the Bilingual Interactive Activation Plus models share some common assumptions, there are also a few major differences between them such as (1) Bilingual Interactive Activation Plus assumes bottom-up directionality, indicating that the task schema system cannot influence the word identification system. In the initial Bilingual Interactive Activation model, the two systems were proposed to interact; and (2) in Bilingual Interactive Activation Plus the language nodes do not serve any functional role and do not influence the activation levels in word identification system, while they were inhibitory in the Bilingual Interactive Activation model.

A highly influential model of language production proposed by Green (1998), the Inhibitory Control Model on the other hand, proposes yet another explanation to the resolution of cross-linguistic lexical activation. The model is based on the assumption that mental representations of words in both languages are active and a series of underlying subsystems gets activated to enable the bilingual speaker to use the target language and suppress interference from the language not in use.

The Inhibitory Control Model assumes the existence of *the bilingual lexico-semantic system, the conceptualiser, the supervisory attentional system,* and *language task schemas,* a set of actions based on one's previous experiences, whose purpose is to support bilingual language users in selecting the target language by means of suppression of another competing language. The bilingual lexico-semantic system refers to the main linguistic mental storage that contains word forms and their meanings in all languages of a person. The role of the conceptualiser is to use the already available information retrieved from long-term memory and create conceptual representations, which enables activation of words in the bilingual system. The Supervisory Attentional System (Green, 1998) has been claimed to inhibit the irrelevant language tags via its connection with the conceptualiser. Additionally, this System can in fact activate the relevant language task schemas as well.

As Green (1998) explains, task schemas are not simply derived from long-term memory as a set of pre-planned actions but they are indeed adaptable on the spot and depending on the situation. These mechanisms; activation, modification, creation of schemas are monitored by the supervisory attentional system. This system is in charge of planning, regulating, and monitoring the efficiency of language task schemas, and it also adjusts their activation according to a situation: whether a conversation in one language is conducted or both languages are engaged. This model proposes that inhibition/attentional control is engaged on two levels. The first level is the lemma level where specific representation management takes place at a local level, and the supervisory attentional system where the native or the second language system is managed at a global level. Inhibition on both these levels work together where the need for a language task arises. In other words, when one language is progressing, the activation of the second language is inhibited. This is indeed possible due to the language tags of words which enable the system to identify the relevant items to be activated or deactivated, and also the supervisory attentional system is then guided as to whether first or second language task schemas should be activated or not.

In summary, language task schemas are in charge of regulating input and output from the bilingual lexico-semantic system for as long as it is required, namely, until it is inhibited by another task schema, or the particular goal is achieved or changed. The Supervisory Attentional System relies on the particular goal and works through the task schemas rather than direct input and output.

As discussed above, Green's model (1998) proposes that both languages of a bilingual speaker compete for selection with a non-target language being suppressed by inhibition. Thus, it has been further suggested that this mechanism can be improved with practice resulting in smaller interference effects in tasks requiring conflict resolution (such as the Simon task which is described in more detail in Chapter One), and lower interference cost consequently, in bilinguals (Bialystok, Craik, Klein, & Viswanathan, 2004).

In light of the considerations regarding executive control in bilingual population (Martin & Bialystok, 2003; Hilchey & Klein, 2011; Von Bastian, Souza, & Gade, 2016), the Inhibitory Control Model makes some key assumptions. First and foremost, the model proposes that both languages of a bilingual speaker are interlinked and remain active, thus additional inhibitory mechanisms are needed to solve the interference problem. Secondly, language control in bilinguals is managed by the supervisory attentional system which oversees the interconnected linguistic systems and their performance. Lastly, the Supervisory Attentional System is a general control system and is claimed to be language external (Abutalebi & Green, 2008). In line with the Inhibitory Control Model, inhibition/attentional control is crucial to allow regulation of bilingual language production. Although monolingual language output also requires the engagement of inhibition/attentional control to control one language, bilinguals manage twice as much information for their two languages. Not only do they need to manage interference within their native language but also between-language interference by using the limited resources of inhibition/attentional control. Thus, this is where the bilingual executive control advantage has been suggested to appear.

3.1.2 Interim Summary

Monolingual and bilingual language learning environments differ in many aspects. One of the fundamental differences is the number of languages being used by speakers on a daily basis and the amount of exposure to each language. In contrast to monolinguals, bilinguals operate in a dual-language environment where they are faced with the necessity to switch between the languages. This has been accounted for in various theoretical models of language control in bilinguals which were discussed more extensively in the previous section.

The next section presents the evidence for bilingual advantage in executive control in more detail.

3.2 Bilingual Advantage in Executive Control and Working Memory

3.2.1 The Theoretical Accounts of the Bilingual Advantage in Executive Control

As noted earlier, one of the arguments that has been put forward for the bilingual advantage in executive control is because of the fact that bilingualism involves managing two languages in terms of activating target language and inhibiting non-target language given the context

(Hilchey & Klein, 2011). There have been a number of theoretical accounts proposed which attempt to account for the bilingual advantage.

Martin-Rhee and Bialystok (2008) recruited primary-school aged children to take part in a series of experiments examining the development of two types of inhibitory control in monolingual and bilingual children. Results showed that the bilingual advantage was replicated for tasks that require attentional control (where competing cues are presented; like the Simon task) but not for tasks requiring response inhibition (where competing responses are included, like the Stroop task) (see Chapter One for the description of Simon task and Stroop task). For the Simon task, the bilingual advantage was noticeable for reaction times in both congruent and incongruent trials, however, the size of the Simon effect did not differ significantly between the language groups. This means that monolingual and bilingual children were equally efficient at resolving conflict by inhibiting irrelevant information. Martin-Rhee and Bialystok indicated that attentional control constitutes a part of executive control and is often employed by more proficient bilinguals to selectively attend to the cues of interest in situations of conflict. They also explain that the bilingual advantage does not simply concern inhibition but rather stems from the constant need to switch between the languages of a bilingual person. Indeed, bilinguals need to control attention between the two linguistic systems in order to be able to ensure fluent communication in a target language.

This bilingual advantage of the ability to resolve conflict more efficiently in comparison to their monolingual counterparts has been conceptualised within the Bilingual Inhibitory Control Advantage which assumes that bilinguals have more efficient inhibitory abilities compared to monolingual speakers as they need to manage language interference (Hilchey & Klein, 2011). The second hypothesis, the Bilingual Executive Processing Advantage refers to the advantage in an overall executive control system which is more advanced than inhibitory abilities on their own. As presented, both hypotheses offer different explanations and propose that different executive control mechanisms may constitute foundations for the bilingual advantage.

Von Bastian, Souza, and Gade (2016) investigated bilingual executive control advantage by approaching bilingualism from the dimension of age of acquisition, language proficiency, as well as language usage and nine executive control skills being measured by numerous tasks. They recruited a sample of 118 young adults to test the four hypotheses regarding the bilingual control advantage: inhibitory control advantage (Bilingual Inhibitory Control Advantage as defined by Hilchey & Klein, 2011), conflict monitoring advantage

(Bilingual Executive Processing Advantage), set-shifting advantage, and generalised cognitive advantage (Table 3.1). They have further divided the Bilingual Executive Processing Advantage hypothesis into three additional components in an attempt to account for the bilingual advantage.

Table 3. 1 Bilingual advantage hypotheses (adapted from Vīnerte and Sabourin, 2019)

Hypothesis	**Predictions**	**Mechanism tested**	**Representative tasks**
Bilingual Inhibitory Control Advantage (BICA)	Faster reaction times (reaction times) in incongruent trials, resulting in less interference (improved inhibition)	Inhibition	Flanker, Simon, Stroop
Bilingual Executive Processing Advantage (BEPA)	Similar performance across all task types (neutral, congruent, incongruent)	General mechanisms	Nonlinguistic Flanker
Conflict Monitoring Advantage (CMA)	When conflict present (incongruent trials), faster reaction times in congruent and neutral trials	Conflict resolution and monitoring	Flanker, Simon, Stroop
Shifting Advantage (SA)	Faster reaction times in switch trials, leading to smaller switch costs	Set shifting	Colour-shape switching
Generalised Cognitive Advantage (GCA)	Regardless of task, better performance compared to monolinguals	General mechanisms (inc. working memory and reasoning)	N-back, Complex digit span

The first of these is the Conflicting Monitoring Advantage which proposes the bilingual executive control advantage is in tasks that require resolution of conflict. As bilinguals are in constant need to monitor and resolve conflict within and between languages, they have been claimed to have superior conflict monitoring skills which apparently seem to be also transferred to non-conflict situations (Von Bastian et al., 2016). The second hypothesis by Von Bastian et al. is the Shifting Advantage which stems from bilinguals’ need to shift between their two languages. In line with the Shifting Advantage, this language switching results in more efficient and arguably faster switching between mental lexicons. Thirdly, the Generalised Cognitive Advantage which assumes that a bilingual's executive control system is more flexible than the one of monolingual speakers and that the overall executive control is rewired as a result of bilingual experience.

It is worth reporting that the above discussed hypotheses also make various predictions. The Bilingual Inhibitory Control Advantage hypothesis for instance states that bilingual language users' responses are faster than monolinguals and that they also demonstrate smaller interference in incongruent conditions (Hilchey & Klein, 2011). The Bilingual Executive Processing Advantage on the other hand predicts that bilinguals' reaction times are faster and thus the bilingual advantage can be noted across all trials (whether they are congruent or not). The Conflict Monitoring Advantage hypothesis makes yet another prediction regarding reaction times in bilinguals, namely, that they are faster in paradigms with mixed trial types (Costa et al., 2008). The Shifting Advantage hypothesis that refers to the set-shifting ability, assumes that bilingual speakers demonstrate smaller switch costs and are indeed faster in conditions involving switching between the task sets. Finally, the Generalised Cognitive Advantage offers a much broader perspective predicting that bilinguals demonstrate an advantage in a number of tasks assessing executive control; not only inhibition/attentional control and set-shifting, but also working memory and memory updating (Von Bastian et al, 2016).

Interestingly, a study by Von Bastian, Souza, and Gade (2016) indicated that there is no consistent bilingual advantage in executive control. Although a number of findings have supported the idea of the bilingual executive control benefit, their study indicates that such benefits from being bilingual are not as broad and robust as suggested by previous research. This might result from the methodological features such as single tasks being used to assess executive control skills separately. Indeed, considering a number of biases (publication bias; c.f. De Bruin, Treccani, & Della Sala, 2015; Donnelly, Brooks, & Homer, 2019), there is a great need for bilingual advantage research to be conducted with much larger and heterogeneous samples. Also, Hilchey and Klein (2011) reviewed eight studies investigating inhibitory control in bilinguals and they found that an inhibitory control advantage, measured as performance on conflict resolution tasks, is not consistently found in bilinguals. Thus, they also proposed that there is limited evidence in support of a Bilingual Inhibitory Control

Advantage but there is greater evidence towards a more general processing benefit in bilingual population.

3.2.2 The Empirical Evidence of the Bilingual Advantage in Executive Control and Working Memory

3.2.2.1 Non-Verbal Executive Control

Bilingualism has been claimed by numerous studies to offer a wide range of benefits far beyond the obvious ones such as the ability to communicate socially in more than one language. Positive aspects of being bilingual has been found across the lifespan in research with infants (Kovacs & Mehler, 2009) through young age (Poarch & van Hell, 2012), and also with adults (young adults: Bialystok, Craik, & Luk, 2008; Pelham & Abrams, 2014; middle-aged adults: Bialystok et al., 2004; and older adults: Bialystok et al., 2004; Bialystok et al., 2008, and for evidence of neuroprotective effects of bilingualism on dementia onset see Bialystok, Craik, & Freedman, 2007). Indeed, attentional skills have been reported to be superior in bilingual population (for a review see Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012: however, see De Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2015).

Very influential work on bilingual executive control has been conducted by Bialystok and colleagues (Bialystok, Craik, Green, & Gollan, 2009; Bialystok et al., 2012; Bialystok, 2017; Kroll & Bialystok, 2013). In one of their studies (Bialystok, Craik, Klein, Viswanathan, 2004), the Simon task (described in more detail in Chapter One) was used in three experiments to examine differences between younger (*mean* age = 43, *SD* = 7.3) and older (*mean* age = 71, *SD* = 8) monolingual and bilingual participants. The results revealed that bilinguals demonstrated better interference suppression – the ability to ignore more prominent cues while attending to less prominent ones – suggesting an advantage in inhibitory control. Also, they seemed better at handling an increasing load on working memory than their monolingual counterparts. Moreover, bilinguals were found to be faster in terms of reaction times in not only incongruent trials, where interference is present, but also congruent trials where there is no interference to be found. This finding was true for middle-aged and older adult participants (with age ranging from 30 to 88).

To further investigate this phenomenon in other age groups, Bialystok, Martin and Viswanathan (2005) conducted a follow-up study in which they employed magnetoencephalography (MEG, i.e., a functional neuroimaging technique that measures the magnetic fields produced by electrical currents in the brain). The results revealed strong bilingual advantage for children and again older adults, however, it was not existent for the group of young adults. Subsequent research with undergraduate students, however, yielded results in favour of a bilingual advantage for this age group as well. Costa, Hernández, and Sebastián-Gallés (2008) recruited Catalan-Spanish young bilinguals (*mean* age = 22 years, range 19 to 32) to participate in an attentional network task (see Chapter One for a detailed description of this task). Results showed a bilingual advantage in faster reaction times as well as more efficient conflict resolution compared to monolingual counterparts. Based on the findings from their work with bilingual individuals, Bialystok and her colleagues (2004, 2008) proposed that it is not necessarily that bilingualism enhances inhibition, but it is inhibition that improves the ability to efficiently manage attention in complex tasks in a bilingual population.

Costa, Hernández, Costa-Faidella, and Sebastián-Gallés (2009) investigated yet another aspect of the bilingual executive control advantage in 122 young adult bilinguals (*mean* age = 19.9) and 122 monolinguals (*mean* age = 19.5). In the Flanker task, bilingual participants demonstrated less interference in incongruent trials and their global reaction times were faster. Similar results were also obtained in studies by Hernández, Costa, Fuentes, Vivas, and Sebastián-Gallés (2010). It has thus been proposed that individuals characterised by superior executive control monitoring and resolution remain unaffected by highly-demanding tasks, such as incongruent trials for instance, and their performance on congruent trials is unaltered and relatively high (Costa et al., 2008). According to Costa et al. (2008), bilinguals demonstrate a Conflict Monitoring Advantage which manifests itself as an ability of the attentional system to identify conflict, determine the degree of it, and take appropriate actions by adjusting responses accordingly. Costa et al. (2008) observed that in the Flanker task, bilingual participants were faster than monolinguals on both congruent and incongruent trials. Superior inhibitory control is expected to manifest itself in faster response to incongruent trials only, thus the authors argued that the superior bilingual performance is associated with their enhanced monitoring skills. This, on the other hand, stems from the bilingual need to monitor the current environmental demands to be able to select the relevant target language.

Similarly, Grundy, Chung-Fat-Yim, Friesen, Mak, and Bialystok (2017) found bilingual advantage in 31 bilingual young adults (*mean* age = 19.3, *SD* = 1.9) over 28 monolingual peers (*mean* age = 19.1, *SD* = 1.5). They adapted the Flanker task by creating three blocks: the first aimed to assess one's ability to perform two conflicting tasks, the second assessed the ability to ignore distractors while maintaining attention to the target item, and the third assessed underlying executive control processes engaged in selective attention, inhibition, and rule-switching. The findings showed that, although no between-group differences were found in performance on the task nor in the Flanker effect, bilinguals experienced less influence from congruency present in prior trials and were more efficient in their ability to disengage attention from the prior trial to enable focus of attention on the target trial.

However, interesting contradictory findings regarding the bilingual advantage being robust in young adults, were obtained from the study by Salvatierra and Rosselli (2011) who used two types of the Simon task, a simple and a more complex version, to assess group differences between 108 monolinguals (younger: *mean* age = 25.88, *SD* = 6.4; and older: *mean* age = 63.40, *SD* = 8.4) and 125 Spanish-English bilinguals (younger: *mean* age = 26.67, *SD* = 6.6; and older: *mean* age = 64.84, *SD* = 7.3). The study revealed that bilinguals' performance was better only in a simple version of the task, which was explained by an enhanced selective attention when demands on working memory are relatively low. When these are heightened, the performance between the two language groups does not significantly differ.

Another crucial executive control component being investigated in studies is set-shifting. Miyake and Friedman (2012) offered a hypothesis that bilingual experience enhances the ability to shift between mental sets of information and the evidence in favour of set-shifting comes largely from set-shifting tasks. Prior and Gollan (2011) compared performance of two groups of bilinguals 41 Spanish-English (*mean* age = 20.0, *SD* = 1.6) and 43 Mandarin-English (*mean* age = 19.4, *SD* = 1.2) speakers, with 47 English speaking monolinguals counterparts (*mean* age = 20.2, *SD* = 1.5). The former bilingual group reported more daily between-language switching and demonstrated a smaller task-switching cost when compared to the monolingual group, whereas the latter reported less language switching and no advantage in task-switching was found in this group relative to monolingual participants. In terms of between-bilingual group performance, Spanish-English speakers again exhibited a smaller switching cost than Mandarin-English bilinguals which indicates a direct link between the ability to switch languages and bilingual advantage in set-shifting performance (Prior & Gollan, 2011). Another study (Wiseheart, Viswanathan, & Bialystok, 2016) revealed smaller mixing costs in bilinguals as opposed to monolinguals indicating the bilingual advantage in better ability to alter stimulus-response associations.

However, in a study by Prior and MacWhinney (2010), 45 monolingual (*mean* age = 18.7, *SD* = .9) and 47 bilingual (*mean* age = 19.5, *SD* = 1.5) young adults were tested using single- and mixed-task blocks. Smaller switch costs were found for bilingual participants, but no language group differences in both blocks. Also, no differences between monolingual and bilingual groups were found in a series of studies by Paap and colleagues (Paap & Greenberg, 2013; Paap & Sawi, 2014; Paap et al., 2017). Thus, it has been suggested that the bilingual experience of using more than one language leads to a greater mental flexibility but studies have not yielded consistent results in terms of bilingual advantage in task-switching performance.

In a recent meta-analysis by Ware, Kirkovski and Lum (2020), 170 studies were analysed to examine the bilingual advantage and whether it is dependent on the executive control measure as well as the participants' age. Results revealed that the executive control advantage in bilinguals is both task- and age-specific. For instance, bilingual performance on four out of seven measures was significantly faster than monolingual performance and their responses were also significantly more accurate than those of the latter group. Additionally, the effect of age was noted in this meta-analysis with greater bilingual advantage found for research comprising samples of 50 plus years of age when compared to studies with 18- to 29-year-old samples (i.e., young adults). A number of additional measures were then employed to assess the extent of the publication bias in the investigation of the bilingual advantage. Publication bias is defined as a type of bias present in academic research being published based on whether research study results in significant or nonsignificant results. Indeed, publishing research with significant findings while ignoring research with nonsignificant results leads to imbalance of the real character of the findings and at the same time introduces bias in favour of positive results (Ware et al., 2020). Ware et al.'s meta-analysis suggests that the bilingual executive control advantage is modulated by participants' age as well as the task employed for assessment. The role of the task employed to examine a bilingual advantage plays a crucial role in identifying it at all.

Van den Noort, Struys, Bosch, Jaswetz, Perriard, Yeo, Barisch, Vermeire, Lee, and Lim (2019) also found a bilingual executive control advantage in their systematic review of 46 studies with participants across the age span. More precisely, 54.3% of eligible studies reported the presence of a bilingual advantage with 56.4% comprising adult participants and the rest examining children. Van den Noort et al. proposed that brain differences between the children and adults may contribute to the fact that bilingual executive control advantage is not consistently found during childhood. They explain that as children's brain areas responsible for executive control are still developing at this point, the bilingual advantage may not be as consistent and clear in this age group. In terms of the publication bias, it was not investigated in this systematic review.

The issue whether the bilingual executive control advantage exists causes an undeniably heated debate. Although a great number of studies conducted over the years has indicated the bilingual executive control advantage, there is still research that failed to replicate it and hence the bilingual advantage hypothesis has been constantly challenged. Hernández, Costa, Fuentes, Vivas, and Sebastián-Gallés (2010) conducted a review of nonlinguistic executive control research comparing monolingual and bilingual samples. The findings revealed that in a significant number of experiments the magnitude of the conflict effect for both populations, bilingual and monolingual, remains similar. Twenty five of 37 experiments reviewed employed mixed designs where congruent and incongruent conditions are presented randomly and 6 out of the 25 pointed to a bilingual advantage being observable in the magnitude of the conflict effect. Nevertheless, in 12 out of the 25 studies, a significant between-group difference was revealed in terms of the global reaction times effects. They also argued that the differences attributed to bilingualism are difficult to disentangle when experiments employ blocked designs (namely, where the task comprises the same trials, either congruent or incongruent, in a single block).

Paap and Greenberg (2013) recruited between 86 to 110 participants across their three experiments and conducted three studies that investigated bilinguals' and monolinguals' performance on a series of executive control measures: antisaccade, Simon, Flanker, and colour-shape switching type of tasks. Participants were matched on their nonverbal IQ. The researchers did not find a consistent pattern of cross-task interactions that would support the idea of the bilingual advantage. Moreover, even though the Simon task and the Flanker task are both claimed to assess one's ability to suppress interference, they did not reveal bilingual advantages in this study (Paap & Greenberg, 2013). The authors suggested that there is limited evidence that the tasks employed typically in studies assessing bilinguals' versus monolinguals' inhibitory control do tap into the same general ability. Stins, Polderman, Boomsma, and de Geus (2005)' study with a group of 12-year-old participants revealed that the correlations between the Flanker, Simon, and Stroop tasks were all relatively small and not significant (exact values not reported), although these are claimed to be measuring the same executive control mechanism, interference. Similar results were found by Fan, Flombaum, McCandliss, Thomas, and Posner (2003) in their behavioural and fMRI study.

No differences between bilingual and monolingual participants were found in other studies for the Stroop task and the Flanker task (Kousaie & Phillips, 2012) as well as the Simon task (Kousaie, Sheppard, Lemieux, Monetta, & Taler, 2014). Paap and Sawi (2014) recruited 58 bilingual (*mean* age = 24.4, *SD* = .78) and 62 monolingual (*mean* age = 24.8, *SD* = 1.1) university students and employed the same set of executive control measures (Antisaccade, ANT, Simon, and Color-Shape task) but this study yielded no consistent results concerning bilingual advantage.

Hilchey, Saint-Aubin, and Klein (2015) found that additional research published between 2011 and 2015 yielded even more inconsistent findings regarding a bilingual advantage for global reaction times and interference cost, and thus the bilingual executive control advantage is not unequivocally supported by evidence.

Lehtonen, Soveri, Laine, Jarvenpaa, de Bruin, & Antfolk (2018) conducted a meta-analysis on 891 effect sizes derived from 152 studies with healthy adults and offered an empirical overview of research to date on the bilingual advantage. Overall, in this publication no systematic support for a bilingual executive control advantage was found based on the available evidence and moderator analyses did not support it either. Prior to estimate corrections, the authors reported a small bilingual advantage for working memory, inhibition/attentional control, and set-shifting which disappeared, however, once the correction for bias was employed. No advantage was found for attention or monitoring. Additionally, moderator analyses revealed bilinguals who acquired the second language before the age of 6 demonstrated a small bilingual working memory advantage which was not found in those who learned the second language after that age. This, however, also vanished after bias correction was implemented. Also, the study did not find any moderating effects of working memory for second language proficiency. The task unity might have been compromised, however, as the meta-analysis comprised both linguistic and non-linguistic measures and these tend not to correlate at all times. In other words, even if the bilingual advantage was there, it might have been concealed due to a high number of uncorrelated tasks (Miyake & Friedman, 2012).

Paap, Johnson, Sawi (2015) argued that the bilingual advantage should be evident in at least two different experimental measures tapping the same executive control skill if it is to be considered coherently present. Additionally, such tasks should also correlate but research to date indicates that this is not always the case (Paap & Greenberg, 2013). One of the methodological limitations proposed by Paap et al. (2015) is the lack of convergent validity of executive control measures utilised in the bilingual research which may result in an identification of a task-specific mechanism rather than domain-free executive control ability. Paap, Johnson, and Sawi (2016) argued that some studies indicating bilingual advantage might have inappropriately matched samples in terms of the demographics (education, socioeconomic status, occupation, etc.). Some types of bilinguals may not exhibit executive control advantage due to insufficient bilingual experience as well as late age of the second language acquisition resulting in lower linguistic competence (Cummins, 1976). To examine the relationship between bilingualism and executive functioning, Paap et al. (2016) provided a summary of interference control and set-shifting research comparing monolingual and bilingual performance. They found that the majority of bilingual advantage tests (around 80%) that took place after 2011 resulted in no group differences with research that found significant bilingual advantages (20%) characterised by relatively small sample sizes. In fact, studies with small samples were more likely to result in a bilingual advantage than research with larger sample sizes.

Bilingual executive control advantages in bilinguals' ability to resolve interference are relatively large and noticeable within middle-aged and elderly sample (Hilchey & Klein, 2011) yet are not consistently found in children and young adults (Costa, Hernandez, & Sebastian-Galles, 2008; Costa, Hernandez, Costa-Faidella, & Sebastian-Galles, 2009). Thus, it has been argued that the bilingual advantage in the latter reduces significantly with practice and hence suggested that the inhibitory control mechanism may not be enhanced in bilinguals. Moreover, another plausible explanation has been offered with bilinguals and monolinguals engaging different executive control modules in resolving conflict and interference, as opposed to possession of the same module but trained by extensive linguistic experience (Hilchey & Klein, 2011).

Valian (2015) suggested that a number of activities that actively engage certain executive control components, such as playing music or taking exercise for instance, can contribute to enhanced executive functioning performance. She further explains that although many detailed definitions of executive control can be found in literature (i.e., Miyake & Friedman, 2012), these are likely not capturing all possible underlying identifiers of the phenomenon. Additionally, the lack of a unified framework with the components of executive control and their role in linguistic contexts contributes to further complications in the field. Indeed, if this issue is not clear, the executive control advantages for bilinguals are also not so transparent (Paap & Sawi, 2014; Treccani & Mulatti, 2015).

Kroll and Bialystok (2013) as well as Morales, Gomez-Ariza, and Bajo (2016) propose that bilingualism cannot be divided into separate executive control mechanisms but rather should be considered as a composite phenomenon and thus examined by a set of measures instead of single tasks. This has been reflected in the studies by Stocco & Pratt (2014) as well as Morales, Gomez-Ariza, and Bajo (2016) or Warmington, Kandru-Pothineni, Hitch (2019) where multiple tasks were employed to investigate complex interactions sufficiently. For instance, Stocco and Prat (2014) employed the Rapid Instructed Task Learning paradigms (RITL) that require participants to adapt their behaviour to every trial according to the instructions they are given at the very beginning of the experiment. They used it in a study with young adult bilinguals whose characteristics included early age of acquisition as well as high second language proficiency. To test whether the bilingual advantage in executive functioning comes from the need to choose and adapt rules flexibly when using two languages, the RITL involving ever-changing basic mathematical operations was employed. Indeed, the study revealed a bilingual executive control advantage as the bilingual group was faster than monolinguals on the execution of the novel rules and this activation was associated with greater modulation of activity in the basal ganglia (Stocco & Prat, 2014).

Warmington, Kandru-Pothineni, and Hitch (2019) recruited a well-defined sample of 23 Hindi-English bilinguals (*mean* age = 23.2) and 23 monolingual English speakers (*mean* age = 23.3) and tested them on a number of executive control measures. They found that bilinguals outperformed monolinguals on tasks tapping response inhibition, working memory tasks, as well as novel word learning. Interestingly, both language groups performed similarly on selective attention measures. Further analyses revealed that novel word learning performance in the bilingual group was linked to verbal working memory ability as well as inhibition, which was not replicated for monolinguals.

Due to the mixed results being yielded by bilingual advantage research, researchers have identified the need to investigate factors contributing to such findings. For instance, Valian (2015) looked more in-depth into the factors impacting executive control abilities and distinguished three probable causes for such inconsistencies within the literature, namely; (1) the broad and perhaps incomplete definition of what executive control encompasses, (2) *the task impurity control*, and finally (3) the possible influence of other executive control experiences.

Although some studies used a number of measures to investigate relationships between executive control mechanisms more in-depth, they are still relatively sparse and further research with a multicomponent approach to the issue is desperately needed.

3.2.2.2 Verbal Executive Control

Although a significant percentage of research being conducted in the field of enhanced executive control mechanisms utilises non-linguistic measures, there are also studies that employ language-related batteries of tasks. Bilinguals are often suggested to be in possession of smaller vocabulary sizes in both of the languages they use compared to monolingual counterparts (Bialystok & Luk, 2012). Additionally, they have also been reported to have slower language processing skills (Poarch & van Hell, 2012; Pelham & Abrams, 2014) which result in them performing worse on vocabulary assessments, but when considering their inhibition/attentional control, they tend to outperform their monolingual peers (Pelham & Abrams, 2014).

Bialystok, Craik, and Luk (2008) employed the Stroop Task with linguistic stimuli in English to investigate bilingual executive control in 96 monolingual and bilingual participants. The study showed that indeed both younger (*mean* age = 20.0) and older adults (*mean* age = 68.0) bilingual participants demonstrated a smaller Stroop effect when compared to monolinguals. They were also faster overall and particularly in the incongruent trials. Another study by Bialystok, Poarch, Luo, and Craik (2014) has further tested this hypothesis indicating that word interference in providing the name of the colour is larger for the monolingual group, and these were greater in older bilingual adults but also evident in the younger participants.

To further test these findings, Heidlmayr, Moutier, Hemforth, Courtin, Tanzmeister, et al. (2014) chose the Stroop Task combining the first, French, language and the second language being German, to test 65 participants (34 French-German bilinguals (*mean* age = 26.8, *SD* = 3.7) and 17 French monolinguals (*mean* age = 32.4, *SD* = 5.2). Their results supported previous findings of a larger Stroop effect in monolinguals. It also revealed that the effect observed was larger for participants' first language, French, than for their second language, German, suggesting that first language processing skills are in fact more automated and efficient.

Coderre, Van Heuven, and Conklin (2013; and also Coderre & Van Heuven, 2014) investigated two groups of young bilinguals (15 English-Chinese (*mean* age = 21.8, *SD* = 2.4) and 24 Chinese-English (*mean* age = 21.0, *SD* = 1.6)) and 24 English monolingual adult speakers (*mean* age = 23.0, *SD* = 4.1) in the Stroop Task in their first and second language.

They found that although they did not observe any group differences in their performance, the findings revealed that there is a smaller interference found in Chinese-English participants in both their native language and the second language, when compared to the interference for first language, English, in monolinguals. For the second bilingual group of English-Chinese speakers, this interference was smaller in the second language compared to the monolingual group but not in their first language. Coderre explained that although the bilingual executive control advantage was to be noted in the study, it was highly dependent on a number of factors such as proficiency, language immersion, as well as script. It is not clear whether the interference in the bilingual groups results from the bilingual executive control advantage or whether it is caused by processing the Chinese script that differs significantly from English. Coderre et al. (2013) also proposed that higher language proficiency in the second language indicates more efficient interference controlling abilities and thus superior executive control compared to low-proficient second language speakers.

In a follow-up study with the Stroop Task by Coderre et al. (2014) employing both behavioural and electrophysiological methodology, the results revealed faster reaction times and higher accuracy of responses in bilinguals. This was clearly indicative of a bilingual advantage in a recruited sample of monolinguals and bilinguals although a reduced Stroop effect was not present. Similarly, Kousaie and Phillips (2012) noted faster reaction times but no inhibitory control advantage in bilingual young adults when tested with monolinguals.

In summary, it can be concluded that the findings from Coderre (2014) as well as Kousaie and Phillips (2012) are in favour of the Bilingual Executive Processing Advantage hypothesis, assuming faster overall reaction times, and those of Heidlmayr et al. (2014) seem to support Bilingual Inhibitory Control Advantage, where smaller Stroop effect and faster responses in incongruent conditions are assumed. Bialystok et al.' study (2008) provided evidence of interference suppression advantage in bilinguals as well as faster responses, while research by Coderre et al. (2013) revealed that only Chinese-English bilingual speakers demonstrated the former as well. This is in line with the Generalised Cognitive Advantage hypothesis, suggesting that there must be more general executive control underpinnings that come to light under different measures. Although the processes engaged in the Stroop task are more complex to disentangle, they do reveal some significant differences between bilingual and monolingual performance.

Stroop type tasks tend to provide support for the existence of bilingual advantage (i.e., Bialystok, 2006; Costa et al., 2008, 2009) and this finding is often interpreted as bilinguals possessing a more efficient conflict monitoring system (Costa et al., 2009) defined as the executive control system's ability to adapt itself to deal with task-demands by means of appropriately adjusting the amount of attention assigned to distractors and the stimuli of interest (Botvinick, Braver, Barch, Carter, & Cohen, 2001).

Bilingualism has been found to yield executive control benefits in nonverbal executive control performance (i.e., Bialystok et al., 2007) but a bilingual disadvantage has also been identified by research particularly in the area of verbal ability. For instance, bilinguals have been found at a disadvantage in tasks requiring lexical access due to their arguably smaller vocabularies in both languages. They were also observed to be slower than monolinguals in responses to picture naming tasks even though the target language used was their dominant language (Gollan, Montoya, Fennema-Notestine, & Morris, 2005). The study by Gollan et al. (2005) was replicated by Ivanova and Costa (2008) who observed that even though both language groups, bilinguals and monolinguals, performed similarly in terms of accuracy, bilingual participants' lexical access was indeed much slower.

Michael and Gollan (2005) proposed a *weaker link hypothesis* which explains that bilingual disadvantage in linguistic measures is directly linked to the less frequent usage of each language of a bilingual which consequently results in weaker connections between both of their languages. Another account, proposed by Hernandez and Li (2007) points to the later acquisition of the second language in a bilingual person to be the factor accounting for the bilingual lexical disadvantage. Moreover, bilinguals deal with greater lexical competition than monolinguals (Costa & Caramazza, 1999) as their two languages get activated non-selectively during language processing. Indeed, their underperformance on lexical tasks can be explained by the higher level of lexical competition (Green, 1998).

Thus, the bilingual advantage has been found in nonverbal tasks, whereas the bilingual disadvantage refers to bilingual performance on verbal measures. However, performing a linguistic task often requires the engagement of executive control. To further examine the relation between bilingual advantageous and disadvantageous performance on verbal and nonverbal processing, a task tapping verbal memory performance is commonly utilised. Here, processes such as effective search strategies (Tulving, 1983), ignoring irrelevant information, and controlling output for accuracy are engaged.

Initial research investigating bilingual performance on verbal memory indicated a bilingual disadvantage. Bilinguals tended to produce fewer words than monolinguals (Gollan, Montoya, & Werner, 2002; Lehtonen et al., 2018) and recalled fewer newly acquired items

(Fernandes, Craik, Bialystok, & Kreuger, 2007) in experiments employing free recall. However, the between-language group difference was no longer present when controlling for vocabulary size difference suggesting that smaller vocabulary sizes in bilinguals lead to the bilingual disadvantage (Fernandes et al., 2007). Interestingly, in another study (Bialystok, Craik, & Luk, 2008), bilinguals were found to be able to recall more lexical items when their lower vocabulary size was accounted for. Bilingual participants who performed equally well as monolinguals on vocabulary size scores demonstrated similar levels of achievement in semantic fluency (Bialystok, Craik, & Luk, 2008). This finding is thus suggestive of a mediating role played by bilingual vocabulary knowledge in their ability to perform on lexical access tasks. Whether the bilingual disadvantage on verbal measures is to be noted depends on a number of factors such as language proficiency, vocabulary size, and arguably executive processes that are engaged in an experimental task. As proposed by Luo et al. (2010), in tasks that are mainly based on lexical access and where executive control is not much involved (i.e., semantic fluency), the bilingual disadvantage is arguably not present if both groups are matched on some language proficiency measure (i.e., vocabulary size). In fact, the bilingual advantage is likely to manifest itself where higher executive functioning is required and language proficiency differences in monolinguals and bilinguals are accounted for meaning that their vocabulary sizes are similar (Pino Escobar, Kalashnikova, & Escudero, 2018).

Verbal fluency tasks are measures used to investigate both vocabulary size as well as executive control. They comprise two main categories: letter fluency as well as semantic fluency (see Chapter Two). Bilinguals seem to respond slower and take more time to retrieve words due to the competition between their first language and second language lexical entries that needs to be resolved, and arguably weaker word form-word meaning connections stemming from certain words being used less often than by monolinguals for instance (Gollan et al., 2002).

Research accumulated evidence for monolinguals outperforming bilinguals on correct responses in semantic fluency (Gollan, Montoya, & Werner, 2002; Sandoval, Gollan, Ferreira, and Salmon, 2010). This is however no longer the case once both groups are matched on their receptive vocabulary (Luo et al., 2010). Letter fluency tasks yield even more inconsistent results with bilinguals performing worse, equally well, and better than their monolingual counterparts (Bialystok, Craik, & Luk, 2008; Luo et al., 2010, Paap et al., 2017)

Luo et al. (2010) investigated language group differences in performance on verbal fluency tasks between young 20 monolingual English speakers and two groups of 20 young bilingual speakers whose English vocabulary sizes differed. By means of a time-course analysis as well as retrieval analysis, they found no differences in their performance on semantic verbal fluency. However, they observed the best performance in word production within the bilingual group with larger vocabulary size as opposed to the two remaining groups. Interestingly, both bilingual groups demonstrated longer subsequent mean reaction times than their monolingual counterparts. This result suggested that bilingual participants were better at executive control than monolinguals (Luo et al., 2010). To further examine bilinguals' performance in verbal fluency, Paap, Myuz, Anders, Bockelman, Mikulinsky, and Sawi (2017) replicated the study but did not replicate the findings by Luo et al. (2010). They also argued that better performance on letter fluency as compared to semantic one should not be treated as strong evidence towards the existence of enhanced executive control abilities and should be in fact investigated more in-depth by employing independent and direct measures of executive control ability (Paap et al., 2017).

Research focusing on bilingual and monolingual differences on verbal fluency are limited and further investigations are needed to answer questions about monolingual and bilingual differences in performance on letter and semantic fluency and whether these difference stem from specific aspects of their executive functioning.

Patra, Bose, and Marinis (2020) compared their participants' performance not only on the number of correct responses but also on the time-course analysis, as well as on their clustering and switching ability. The study comprised a set of assessments of inhibition/attentional control, set-shifting, and memory and how these are related to participants' performance on verbal fluency tasks. Luo et al. (2010) found that bilinguals with large vocabulary sizes (that were matched with monolinguals) demonstrated a profile of greater number of accurate responses, longer subsequent reaction times, and the slope for their performance was flatter than for monolinguals for the letter fluency task. In other words, they take longer to recall subsequent words and demonstrate superior executive control skills as they are more efficient at resource-monitoring and usage over time during the retrieval process. This finding was also replicated by Friesen, Luo, Luk, and Bialystok (2015) where in letter fluency bilinguals produced more correct responses than speakers of only one language. Again, no differences were noted for semantic fluency performance (Friesen et al., 2015).

On the other hand, another set of studies (Sandoval, Gollan, Ferreira, & Salmon, 2010) indicated longer subsequent reaction times but fewer accurate words produced by bilinguals in the letter fluency task. Thus, it has been suggested that the bilingual disadvantage in word production stems from cross-linguistic interference which also results in slower ability to retrieve target words and hence longer subsequent responses are to be observed. Additionally, with high-vocabulary bilinguals (Luo et al., 2010) outperforming monolinguals on letter fluency task, it has been proposed that the bilingual advantage stems from bilingual's enhanced executive control which is a result of constant dealing with cross-linguistic interference from both language that bilingual speakers need to face (Luo et al., 2010; Friesen et al., 2015, Abutalebi & Green, 2008).

Shao, Janse, Visser, and Meyer (2014) recruited older Dutch participants and assessed their verbal fluency. The results of the study revealed that updating of working memory was a significant predictor of the number of correct responses produced in both letter and semantic fluency, while vocabulary size predicted the speed of the initial response in the time-course analysis. Additionally, the speed of lexical access was found to predict the speed of the initial response but only for the semantic fluency performance.

Patra, Bose, and Marinis (2020) conducted a study with Bengali-English bilinguals and English monolinguals matched on their vocabulary sizes. They found similar between-language group performance in the more linguistically demanding semantic fluency. In letter fluency, which is believed to engage higher order executive control skills, bilinguals demonstrated better performance than their monolingual counterparts. Bilinguals do not exhibit a disadvantage when they are matched on their vocabulary sizes with monolingual participants. Research to date (Patra et al., 2020; Paap et al., 2017) suggests that several independent measures of executive control as well as a number of verbal fluency tasks allow explanations of bilingual advantage or disadvantage.

3.2.2.3 Working Memory

Working memory and executive control are related processes and the former has been found to be a mediator in executive control such as planning (Morris, Downes, Sahakian, Evenden, Heald, & Robbins, 1988), learning (Hasher & Zacks, 1988), or decision-making (Bechara & Martin, 2004). The processes that take place within working memory, such as information manipulation, memory updating or reorganisation, are monitored and regulated by the central executive system (Baddeley & Logie, 1999).

There has been some debate whether the bilinguals executive control advantage reflects enhanced working memory ability rather than superior executive control skills (Namazi & Thordardottir, 2010). It is difficult to distinguish these two and consider them as separate entities as working memory and executive control are closely linked together (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Also, common definitions of working memory and updating tend to overlap in terms of functions they define: both referring to one's ability to manipulate information that is held in memory (thus they seem to be considered as equal (Bialystok, 2015), although it has been evidenced that they are related but separate components (Paap & Sawi, 2014)).

While bilinguals need to manage two languages that are in constant competition for being chosen, for which process they require working memory resources, the assumption of superior working memory performance seems rational. Research that investigates bilingual advantage in working memory has yielded inconsistent findings. It is thus of great importance to further investigate the role bilingualism on both working memory and executive control in the bilingual population.

A review investigating the associations between bilingualism and superior performance on executive control tasks in children and young adults was conducted by Adesope et al. (2010). The meta-analysis examined the monolingual versus bilingual differences in performance on working memory and inhibition/attentional control among others. Based on 63 studies, the bilingual advantage in working memory was revealed, and an even stronger one for attentional control (see Table 3.2 for more details). The results also indicated that the publication bias seemed not to impact their meta-analysis findings.

In Grundy and Timmer's (2017) meta-analysis of 27 studies of children and adults, a small bilingual advantage was found in working memory when compared to monolingual speakers. Additional moderator analyses revealed the greatest working memory advantage in children when compared across the age groups (younger and older adults). Also, a greater bilingual advantage was found for verbal working memory performance in their first language rather than a second language. However, Grundy and Timmer (2017) did not investigate the impact of participants' socioeconomic status and education on their working memory performance which are significant factors shown to affect working memory (Hackman & Farah, 2009), but they are often not reported in research studies.

Von Bastian, De Simoni, Kane, Carruth, and Miyake (2017) conducted a meta-analysis on 88 studies and found a small and positive effect of bilingualism on working memory performance however high between-study heterogeneity in data suggests that this result should be treated with caution.

In another study by Bialystok, Craik, and Luk (2008), no language-group differences were found on their performance on self-ordered pointing tasks and only slight differences were noted for the Corsi task with bilinguals recalling more target items. Self-ordered pointing task (Petrides & Milner, 1982) is a non-spatial executive working memory task in which participants are shown a booklet with several paper sheets and on each page identical set of items varying in their location is presented. Participants are then asked to point to a different target on every page while remembering not to touch an item that they have already touched. Corsi task, on the other hand, assesses visuospatial working memory and here the researcher taps a sequence of the nine square blocks to then ask participants to tap the same block in the correct order. Bialystok, Craik, and Luk did not report participants' socioeconomic status and reported a wide range of second languages of their participants indicating that they came from different cultural, linguistic and potentially ethnic backgrounds.

Luo, Craik, Moreno and Bialystok (2013) recruited 157 young adults to test their performance on forward and backward verbal and spatial tasks investigating working memory and found better spatial performance in bilinguals and better verbal performance in monolinguals. However, in this study on the other hand, participants from both groups differed in terms of their level of English vocabulary knowledge as well as non-verbal intelligence. Additionally, the study recruited uneven sample sizes with 99 bilinguals with various languages and thus ethnic differences, and 58 monolinguals. Here, the study did not report details about socioeconomic status.

Hansen, Macizo, Duñabeitia, Saldaña, Carreiras, Fuentes, and et al. (2016) examined the bilingual advantage in working memory in a group of Spanish speaking children. Participants were either monolingual or emergent bilinguals with Spanish as their first language and English as a second language at immersion school. The groups were additionally matched on socioeconomic status, intelligence, and they also came from the same city. The results revealed that bilingual children outperformed monolinguals on n-back tasks (see Chapter One for a description of the task), whereas monolinguals outperformed them on reading span tasks. This result was found to be the case in younger bilinguals as older ones were observed to be at an advantage in the reading span task too. This finding suggests that bilingualism may enhance domain-general executive control ability but on the other hand it seems to impair linguistic processing (Hansen et al., 2016). Age seems to alter the developmental course of working memory, while bilingualism seems to modulate the development of working memory in younger bilinguals but the presence of this modulation is only detectable in certain tasks.

Bialystok, Craik, Klein, and Viswanathan (2004) employed the Simon task to investigate the influence of age of bilinguals on executive control. The results revealed that bilinguals performed better than monolinguals when working memory manipulation was greater but their performance on lightly loaded working memory tasks were comparable. Thus, it indicates that the bilingual advantage is to be observed within the working memory performance further suggesting an interrelation between working memory and executive functioning.

For instance, Morales, Calvo and Bialystok (2013) recruited bilingual primary school children for a modified Simon-task and a visuospatial span task. They found the bilingual advantage in terms of faster and more accurate reaction times in Simon, and in a visuospatial working memory overall. Bilingual children demonstrated more efficient information processing ability than their monolingual counterparts. In another study, Hernandez, Costa, and Humpreys (2012) assessed bilingual executive control ability in Catalan-Spanish speakers in three visual search experiments. They found that bilinguals demonstrated faster responses than their monolingual counterparts in all three conditions. Interestingly, they seemed to be less impacted by irrelevant stimuli in their working memory but equally impacted by visual priming and unique target objects suggesting that:

> "bilingualism influences performance on non-linguistic tasks that demand executive control mechanisms to prevent top-down effects on attention from irrelevant information in WM [working memory]. On the other hand, mechanisms that control attention in the face of distracting bottom-up signals do not differ between bilingual and monolingual participants" (Hernandez, Costa, & Humphreys, 2012, p.49).

Blom, Kuntay, Messer, Verhagen, and Leseman (2014) tested Turkish-Dutch bilinguals and Dutch monolinguals on their verbal and visuospatial working memory. They found the bilingual advantage in low socioeconomic status Turkish-Dutch children in both working memory types when controlled for vocabulary and socioeconomic status. This finding suggests that bilingual advantage in executive functioning can be found in socioeconomically deprived bilingual individuals and the extent of this advantage is modulated by bilingual speaker's language proficiency. It also further supports claims that bilingual advantage occurs beyond inhibition and the bilingual experience enhances the central executive control that regulates processing in a number of cognitively demanding tasks.

However, Engel de Abreu (2011) did not identify a bilingual advantage in visuospatial working memory, even though she employed highly similar tasks to the ones used by Morales et al (2013) and her participants were of the same age and socioeconomic status as children in the previous study. She recruited 44 bilingual and monolingual children aged 6 to 8 years old that were matched on a number of factors: age, gender, socioeconomic status, as well as fluid intelligence and their language ability. Monolinguals outperformed bilinguals on language measures with no language group differences for working memory performance.

Contrary findings have also been found by Ratiu and Azuma (2015). They recruited 53 monolingual speakers of English as well as 52 English-Spanish bilinguals to examine between group differences in working memory by using complex span tasks. The results showed that bilingual participants did not exhibit a superior verbal or visuospatial working memory performance even though they had a significantly higher level of education and started using both their languages before they were 4 years old. Thus, in this study where participants were tested by both single and complex span tasks to examine their working memory performance, the bilingual advantage did not occur. Interestingly, working memory performance was found to be a strong significant predictor of participants' performance on other working memory measures, but their bilingual status was not.

Buac, Gross, and Kaushanskaya (2016) recruited English monolingual children (*mean* age = 6.34, *SD* = .84) and Spanish-English bilinguals (*mean* = 6.24, *SD* = .76) to investigate their performance on a set of processing-based tasks engaging short-term memory, working memory, and novel word learning task. Monolinguals performed better than bilinguals only on the short-term memory measure. Additionally, socioeconomic status and vocabulary size was predictive of children's performance only in the bilingual group of children.

Anton, Carreiras, and Dunabeitia (2019) investigated a bilingual advantage in executive control and working memory with large and well-matched adult samples of monolinguals (*mean* = 21.84, *SD* = 3.05) and bilinguals (*mean* = 22.29, *SD* = 2.87) from the same country. To be able to isolate the bilingual advantage within working memory and executive control, they employed tasks tapping into these separately thus excluding the need for participants to activate inhibition or switching processes. No differences in group performance were found when bilinguals and monolinguals were matched on their socio-demographic factors.

Similarly, Lopez (2021) examined the role of language proficiency, number of languages spoken as well as age of acquisition in memory performance of university students. Phonological short-term memory was investigated by means of the digit span task, visuospatial memory by the Corsi task, and semantic memory by employing the word span task. No group differences were found on semantic memory indicating that it is language independent. To be more precise, concepts (i.e., word meanings, lexical items, factual information) stored in semantic memory are not associated with the number of languages used by a person. Intermediate second language speakers as well as multilingual speakers performed significantly better than monolinguals on visuospatial memory tasks, while intermediate and advanced second language speakers as well as multilinguals were also better on phonological short-term memory measures when compared to simultaneous bilinguals. These findings propose that second language acquisition may enhance phonological short-term memory as well as visuospatial memory.

Research on bilingual advantage in working memory has been mainly conducted by means of simple span tasks while fewer studies employed more sensitive measures tapping working memory more efficiently. Moreover, it has been suggested that bilinguals may be at disadvantage in their verbal task performance arguably due to their reduced vocabulary knowledge in both languages as well as a mild deficit in verbal processing which might be linked to their need to resolve conflict between two non-selectively activated languages (Luo, Craik, Moreno, Bialystok, 2013).

3.2.3 Modulating Factors of the Bilingual Advantage Research

One of the reasons that has been argued to result in inconsistencies within the field of bilingual advantage research is the difficulty in defining bilingualism and recruiting a homogenous sample. The majority of studies have recruited speakers who learned both languages simultaneously or immigrants who came to the country of their second language and became balanced bilingual speakers. Less is known about sequential bilinguals who differ in their language proficiency levels in both languages. It has been proposed that the degree to which bilingualism impacts cognition may be dependent upon bilinguals' proficiency level – with bilingual advantage appearing gradually in higher proficient bilinguals. In line with the threshold hypothesis offered by Cummins (1976), a certain proficiency level in both first and second language needs to be achieved in order for the bilingual executive control advantage to occur. However, it should be also taken into consideration that there is not much research to date that would indicate the potential relationship between bilingual executive control advantage and their language proficiency levels.

Table 3. 2 Meta-analyses of bilingual advantage in executive control

Paper	**No of studies**	**Effect size (Hedges' g)**
Hilchey & Klein (2011)	13	Not reported
Adesope, Lavin, Thompson, & Ungerleider (2010)	63	Bilingual advantage found in working memory (*g* = .48, *p* < .01) and in attention (*g* = .96, *p* < .01)
van den Noort, Struys, Bosch, Jaswetz, Perriard, Yeo, et al. (2019)	46	Not reported
Ware, Kirkovski, & Lum (2020)	Not reported	Bilingual advantage found: faster (.23 - .34) and more accurate (.18 - .49) reaction times.
Donnelly, Brooks, & Homer (2019)	80	Bilingual advantage found initially (*g* = .11, *p* = .007) but when corrected for publication bias becomes nonsignificant (*g* = -.17, *p* = .067).
Lehtonen, Soveri, Laine, Järvenpää, de Bruin, & Antfolk (2018)	152	Bilingual advantage found initially (*g* = .06, *p* < .05) but when corrected for publication bias becomes nonsignificant (*g* = -.08, *p* = .099). Smaller effect for verbal tasks (*g* = .01) than for non-verbal (*g* = .30) when corrected for bias non-verbal one becomes nonsignificant.
Paap (2019)	109	Bilingual advantage found initially (*g* = .11, p < .001) but when corrected for publication bias becomes nonsignificant (*g* = -.02, p = .708)
Paap, Anders, Mikulinsky, Mason, & Alvarado (2017)	101	Not reported
Von Bastian, De Simoni, Kane, Carruth, & Miyake (2017)	88	Bilingual advantage found (*g* = .11, *p* = not reported but sig.)

De Bruin, Treccani, and Della Sala (2015) pointed to a publication bias as a crucial factor indicating the bilingual advantage due to the lack of published research findings to the contrary. In their first study investigating a bilingual advantage in any executive control measure, they found that 63% of conference abstracts that reported an advantage were published, whereas only 36% of those abstracts where a bilingual advantage was not found. The second study comprised a meta-analysis on papers submitted for conferences that resulted in publication: here, a medium effect size was identified (d = .30) and the analysis of funnel plots pointed to the strong evidence of publication bias being present.

Meta-analysis by Donnelly et al. (2019) included 80 studies of monolingual and bilingual performance on executive functioning tasks. The results revealed a bilingual advantage of a small effect size (see Table 3.2). Initially, publication bias was not observed, however, it was revealed to be present after the effect size was corrected by means of an alternative model (PET-PEESE, Stanley & Doucouliagos, 2014 which presentation is beyond the scope of this research). The implementation of this model resulted in the bilingual advantage not being significant anymore. Further moderator analyses conducted in this meta-analysis revealed that no effect of age or task (i.e., Simon, Flanker, Stroop, and ANT) was found to modulate the effect sizes found.

Similarly, another meta-analysis (Lehtonen et al., 2018) presenting 152 studies on monolingual and bilingual executive control performance, revealed a bilingual executive control advantage. The advantage was of a small effect size and it was no longer present after correcting for publication bias. Moderator analyses conducted as a part of the meta-analysis resulted in task and age not being significant moderators of the effect sizes of interest. Thus, the main finding from this research was the proposal that bilingual advantage might in fact be attributed to publication bias (Lehtonen et al., 2018).

A similar conclusion stemmed from the study by Paap (2019) who closely examined 109 eligible experiments employing Simon, Flanker, Stroop, or ANT. The initially present bilingual advantage became nonsignificant when corrected for publication bias.

Taking the above presented research into consideration, what needs to be noted is the fact that publication bias is in fact omnipresent in psychological research and does not negate the existence of the real effect itself.

Another important investigation into the available literature regarding the bilingual executive control advantage was conducted by Sanchez-Azanza et al. (2017) who examined 139 studies and separated them into four categories based on their findings: research supporting the bilingual advantage (43), research reporting mixed results (ambiguous towards the advantage; 42), studies that challenged the advantage (29) as well as the ones that did not mention it at all (25). They revealed that in 2014 and 2015 there was an increase in publications that challenged the bilingual advantage. Moreover, in June 2016 the studies from 2014 accumulated a higher number of citations than the studies in favour of the bilingual advantage. The authors suggest that according to the available evidence-based bibliometric information there seems to be a turn in publication trends within the current state of literature of interest that has followed the influential publication by Paap and Greenberg (2013). Recent research demonstrates evidence pointing in both directions (Anton, Fernandez, Carreiras, & Duñabeitia, 2016; Colzato, Bajo, van den Wildenberg, Paolieri, Nieuwenhuis, La Heij, & et al., 2008). With research in support of particular theory being more likely to be published, Sanchez-Azanza et al. (2017) emphasise the importance of reporting all findings regardless of the outcome. They recommend that future researchers are cautious with disentangling the factors engaged in the bilingual advantage hypothesis and remain mindful about designing well-controlled and replicable experiments.

Interim Summary

The previous section has demonstrated that there are inconsistent findings regarding bilingual research and that the bilingual executive control advantage has not been consistently found. Recent systematic reviews have highlighted that the bilingual advantage may in fact not exist. The reason why such inconsistencies may be found is publication bias and the idea of the recruitment of heterogeneous samples.

Research investigating the impact of bilingualism on working memory performance specifically is relatively sparse. Findings to date yield largely inconsistent results not allowing for a consistent conclusion to be drawn with studies evidencing the bilingual advantage (Bialystok, Craik, Luk, 2008; Morales, Calvo, & Bialystok, 2013), similar between language-group performance (Martin-Rhee & Bialystok, 2008; Luo, Luk, Bialystok, 2010), as well as bilingual disadvantage (Ratiu & Azuma, 2015).

The above presented inconsistencies within the area of bilingual advantage in executive control and working memory emphasise the need to further investigate the phenomenon to obtain clear and more concise findings.

3.3 Summary

The aim of the current chapter was to introduce various conceptualisations of bilingualism and to present the multidimensional character of this phenomenon. It further described different theoretical frameworks that were proposed in the literature to account for language control in bilingual person's mind. The Chapter also introduced the concept of bilingual advantage and some key hypotheses that aim to explain this advantage with certain executive control underpinnings that contribute to its presence. The crucial part of the Chapter was the critical review of available research regarding bilingual advantage and key findings presented in this section. The aim of this subsection was to identify inconsistencies within up-to-date bilingual advantage research findings along with methodological limitations, publication bias, sample size issues, and heterogeneity of bilingual samples.

The next Chapter explains the relationship between bilingualism and long-term memory mechanisms. It presents a systematic review of literature and research findings on the bilingual advantage in episodic memory.

CHAPTER FOUR

Bilingual Episodic Memory and Paired Associate Learning: A Systematic Review and Meta-analysis

4.1 Introduction

Chapter One, Two, and Three have provided a review of the relevant literature and empirical evidence regarding the mechanisms underlying executive control (inhibition/attentional control), working memory, as well as long-term memory mechanisms. Also, a review of bilingual advantage literature was presented. In the light of the fact that little is known about the long-term memory mechanisms, episodic memory which is of a particular interest to this thesis, a more systematic and rigorous review was conducted. Episodic memory performance was investigated by means of paired associate learning tasks (see Chapter Two, cf. Papagno & Vallar, 1995) and measured by recall and recognition methods which were introduced in more detail in Chapter Two. Paired associate learning paradigm enables examining both short-term memory and long-term memory storages as it requires accessing and recalling lexical items over a period of time.

Owing to globalisation, an increased number of people are now faced with the necessity to learn a foreign language for a variety of reasons, such as education, work, or travel. Understanding the way people learn new languages and how it impacts cognition has been a focal point of interest for many researchers examining bilinguals and speakers of many languages. In fact, being able to communicate in a number of languages has been considered to yield executive control advantages over monolingual speakers (Bialystok, 1999; Bialystok, Craik, Klein, & Viswanathan, 2004; Colzato et al., 2008).

Bilingual experience has been in fact found to influence not only linguistic but also executive control mechanisms, some of which are closely linked to information encoding as well as information retrieval (Bialystok, Craik, Green, & Gollan, 2009). While the plethora of evidence examining this phenomenon has mainly focused on executive control (i.e., inhibition/attentional control) and working memory, very little research was conducted to examine the impact of bilingualism on long-term memory processes (i.e., episodic memory). [Episodic memory has been discussed in more detail in Chapter Two]

The lack of research examining bilingualism and its impact on episodic memory is surprising in light of the fact that neuropsychological evidence has shown that bilingualism delays the onset of clinical symptoms of dementia as it is a form of executive control stimulation (Bialystok et al., 2007; Perani & Abutalebi, 2015). There is also evidence that the quality and quantity of bilingual experience is associated with changes in brain structure (Garcia-Penton, Garcia, Costello, Dunabeitia, & Carrieras, 2016). From that perspective, it can be proposed that bilingualism plays an important role in the organisation of the brain structure and thus may have a direct impact on long-term memory storage.

It has been proposed that the bilingual advantage may be reflected in better performance of episodic memory. Supporting evidence for this claim has been provided by the picture recall task conducted by Schroeder and Marian (2012) with monolingual and bilingual participants. Results revealed that bilinguals recalled more pictures than their monolingual counterparts with factors such as early acquisition and more extensive bilingual experience associated with better episodic memory performance (Schroeder & Marian, 2012). Thus, bilinguals seem to demonstrate enhanced episodic memory performance when non-verbal stimulus is involved as well as decreased age-related cognitive decline.

The bilingual advantage in executive control is still a highly contentious topic with mixed evidence available (see Chapter Three). There is very limited research examining episodic memory via paired associate learning tasks and the available research has provided mixed results. Taking into account the fact that research suggest that bilingualism delays the onset of dementia (Bialystok, Craik, & Freedman, 2007), it is in fact surprising that there is no more behavioural research with younger bilinguals to investigate whether the bilingual advantage may be found in episodic memory. Thus, there is a need to look at this phenomenon in more detail and investigate whether there is a bilingual advantage in episodic memory as measured by paired associate learning tasks. This systematic review will thus be of an exploratory nature in the field.

The aim of this systematic review is to provide cumulative evidence and inform the current state of knowledge regarding the bilingual advantage in episodic memory (i.e., as measured by paired associate learning). Although many studies examining the bilingual advantage have focused on executive control, there is limited understanding of how bilingualism modulates episodic memory (long-term memory mechanisms).

Thus, taking this into account, the following research questions are stated:

1) Is there a bilingual advantage in episodic memory as measured by paired associate learning tasks?
2) Is the bilingual advantage in episodic memory task-dependent (i.e., how paired associate learning is assessed: recall vs. recognition measures)?

3) What is the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance, as measured by paired associate learning tasks?
4) What is the role of bilinguals' vocabulary knowledge (in both languages) in bilingual episodic memory performance, as measured by paired associate learning tasks?
5) Is there a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control)?

4.2 Materials and Methods

Systematic review is one of the common methods that enables the review of relevant literature on a particular topic in a systematic manner. It is commonly referred to as:

> "a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review" (Wright, Brand, Dunn, & Spindler, 2007, p. 24).

Indeed, systematic reviews can reduce bias and resolve controversy between conflicting research findings. Moreover, due to the rigorous procedures, systematic reviews are also replicable.

Meta-analysis is a statistical analysis that enables calculating an overall effect for a body of research based on available empirical evidence that addresses the same research question (Haidich, 2010). Meta-analysis aims to systematically synthesise findings of previous research by combining results of various, often conflicting, studies and analyses. Indeed, meta-analysis provides a more precise estimate of the effect than single studies. The utmost importance of meta-analysis can be defined by its aim to establish statistical significance across various research studies that might yield conflicting results when investigated separately. This, in fact, increases the reliability of information.

This systematic review followed PRISMA guidance (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, 2020) in addition to a meta-analytic approach in accordance with Field and Gillett's protocol (2010).

4.2.1 Search Strategies

4.2.1.1 Inclusion and Exclusion Criteria

4.2.1.1.1 Types of Participants

This systematic review only included samples that comprised certain characteristics. Specifically, participants were deemed eligible to be included in the analysis if they were typically developing monolingual or bilingual individuals aged more than 37 weeks. Research examining infants born prematurely and atypical developing individuals were ineligible and hence not considered in this review. Additionally, samples consisting of participants with developmental/acquired disorders, hearing impairments/deaf, bi-modal bilinguals were excluded.

The review adopted a broad definition of bilingualism (i.e., understood as the ability to communicate effectively in two languages: Bloomfield, 1956), including studies that comprised samples of individuals with the knowledge of more than one language as presented by the authors of eligible studies. Thus, studies with both simultaneous (i.e., started to learn both languages at approximately the same early age) and sequential (i.e., defined as bilinguals who started learning the second language once they have acquired their first language or who began second language acquisition later in life) bilingual speakers were included (Kohnert, 2008). To identify the bilingual status of the samples (whether sequential or simultaneous), the papers were screened in terms of the information regarding participants' proficiency or self-reported proficiency in both languages. When studies reported proficiency (i.e., how well a person has mastered a language) separately for each four skills (listening, speaking, writing, reading), the decision was made by two independent reviewers as to which type of bilingualism is more probable. This was also applied where only the exposure to each language or language measures were provided. In case of the missing proficiency information, the samples were marked as no information provided. Research with second language learners was excluded from the review.

Monolingual participants were defined as individuals who were not proficient in more than one language and who have never learned another language to a functional proficiency.

4.2.1.1.2 Types of Studies

The review only considered studies that examined the bilingual advantage in episodic memory (i.e., as measured by paired associate learning tasks) in eligible groups of monolinguals and bilinguals. Eligibility criteria comprised published and unpublished papers to enable wider access to valuable work that might be of interest to this review and to avoid the risk of biased retrieval (Rosenthal, 1995) of publications characterised by only large effect sizes and small p-values. Thus, conference proceedings, posters, talks, theses (Masters and PhDs), and reports on funder websites were also screened for suitability.

The types of studies that were eligible to be included in the review comprised experimental, quasi-experimental, longitudinal, concurrent, and brain-imaging studies (event-related potential, and functional magnetic resonance imaging). Case studies were not eligible.

In the context of this systematic review, paired associate learning was defined as learning an unfamiliar word (a novel word) for an unrelated and unfamiliar referent (a novel object). Additionally, working memory (i.e., verbal and visuospatial memory) and inhibition/attentional control performance were of primary interest.

4.2.1.1.3 Outcome/Dependent Variables

The outcome measures were the performance on working memory, executive control (i.e., inhibition/attentional control), and paired associate learning tasks (i.e., recognition and recall).

4.2.1.1.4. Moderator Variables

A number of moderators of interest to this meta-analysis comprised participants' age and the socioeconomic status which have been found to impact paired associate learning performance (Buac, Gross, and Kaushanskaya, 2016). For bilinguals, the age of language acquisition, language proficiency and language usage were of interest.

The review also investigated the language environment of participants, namely, whether they operate in a monolingual dominant context (i.e., where one language is used predominantly) or a bilingual dominant context (i.e., where two languages are used on a daily basis).

4.2.1.1.5 Limits

Only articles published in English were included in the review. No date limitation was imposed on the studies with the final search conducted on the 11th of March 2021. Eligible studies were published journal articles, unpublished manuscripts, conference proceedings, posters, talks, theses (Masters and PhDs), and reports on funder websites.

4.2.1.2 Literature Search and Study Selection

The databases searched include PsychINFO (OvidSP), MEDLINE, Web of Knowledge, ProQuest Dissertations and Theses, PubMed, and ERIC (EBSCOhost). PsycINFO was the main search engine used in the review and it was periodically searched from the 26th of March 2018 until the 11th of March 2021 to ensure that all the eligible studies are included in the review. The search strategy is presented in Appendix A. Further manual examination of reference sections of the eligible studies was employed to seek additional studies. Where studies could have not been obtained printed or on-line by the reviewers, the authors were contacted via email and asked to provide the papers for the purpose of the systematic review.

The results yielded by both automatic and manual searches were downloaded and formatted that enabled further analyses with the use of Covidence (Covidence systematic review software available at www.covidence.org). The identified studies were uploaded onto Covidence and screened for duplicates and irrelevant papers. The summary of this process along with the results yielded is presented in PRISMA diagram in Figure 4.1.

Both titles and abstracts of the eligible studies were screened by two researchers using Covidence software. Where a decision could not be made based on title and abstract alone, the full study was obtained and screened for the final decision. With conflicting decisions made by two reviewers, a third reviewer was consulted to resolve the conflict.

Duplicates and studies that did not meet the inclusion criteria were removed. Full-text screening included 226 eligible studies which were further assessed in terms of their eligibility by two reviewers: each reviewer read the studies independently to determine their eligibility for inclusion. Forty-one studies of 226 studies selected for full-text screening were not accessible online and therefore authors (where contact details were available) were contacted and asked to provide the full-text studies for the purpose of the systematic review. Nineteen authors were contacted via Researchgate which resulted in three additional studies being included for full-text screening.

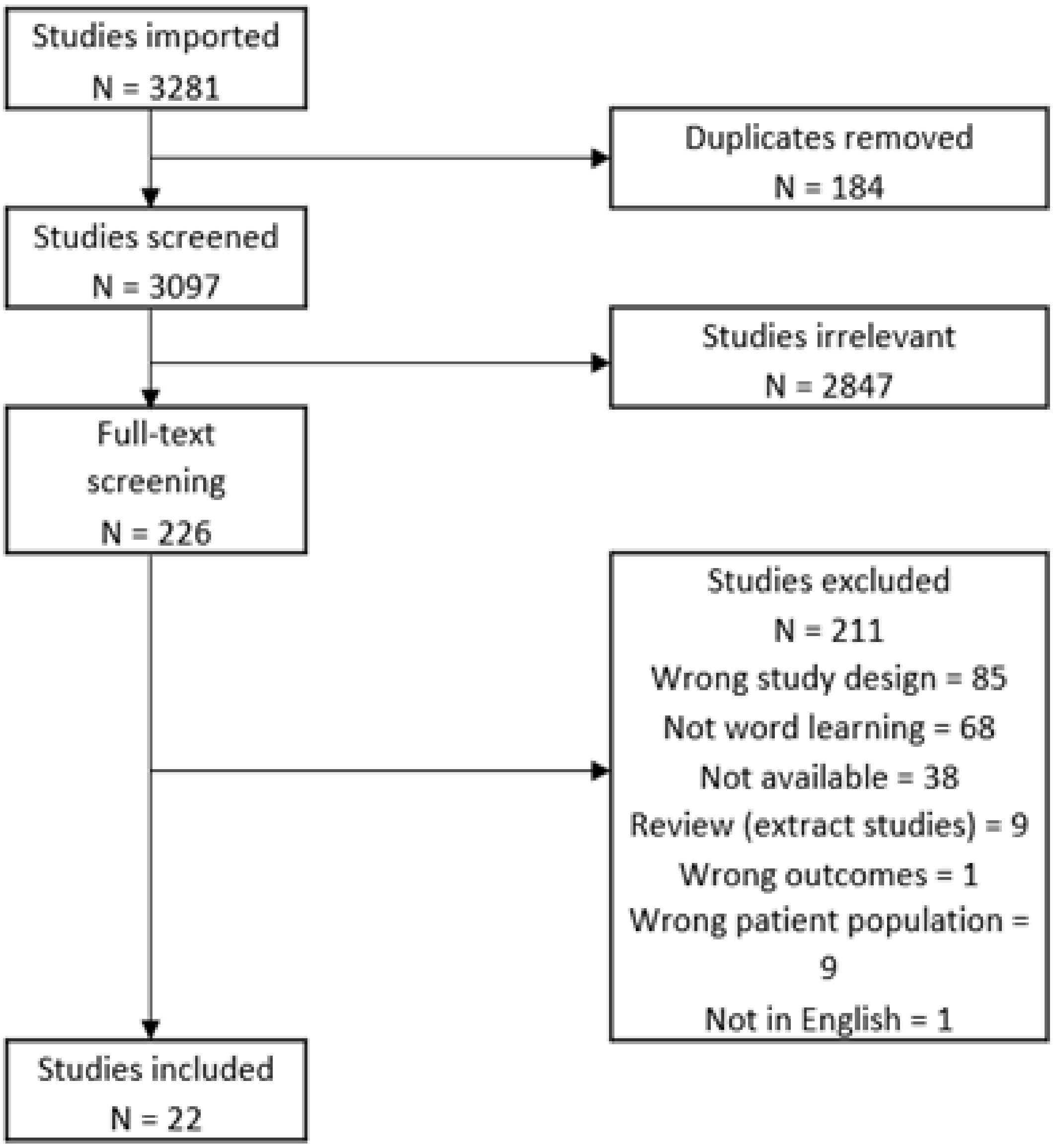


Figure 4. 1 PRISMA flow diagram of search for the systematic review

Of the initial 3281 studies, 184 studies were removed as they were duplicates and 2847 were eliminated as they were irrelevant to the topic of interest. Of 226 studies eligible for full-text screening, 20 studies met the criteria for inclusion. This resulted in 22 studies (2 studies comprising 2 separate eligible experiments) from which the data was extracted and included in further analyses.

4.2.2. Quality Assessment

Once the final set of studies were identified for full-text screening, the quality of the evidence was assessed. Using Gough's (2007) Weight of Evidence (WoE) criteria, two reviewers independently assessed the quality of each study (see Table 4.1).

Two reviewers carried out the quality assessment independently. As this process lends itself to some subjectivity, following Murphy and Unthiah (2015) when there were discrepancies between reviewers the lowest assessment was accepted (for example if reviewer 1 assessed WoE A as high and reviewer 2 as medium the final agreement was medium). Table 4.2 presents the final ratings for these four domains for each study that was included in the systematic review.

Table 4. 1 Quality assessment criteria based on Weight of Evidence (Gough, 2007)

Judgement	WoE A	WoE B	WoE C	WoE D
High, Medium or Low	Taking account of all quality assessment issues, can the individual study findings be trusted?	What is the appropriateness of the research design and analysis for addressing the aims of the individual study?	What is the relevance of the particular focus of the individual study for addressing its aims?	Taking into account the quality of execution, appropriateness of the design and relevance of focus, what is the overall weight of evidence this individual study provides to answer its research questions?

Note: WoE A concerns overall coherence and integrity of the individual study; WoE B refers to the appropriateness of the given form of evidence to answer a question; WoE C concerns the relevance of the evidence available to answer a question; WoE D refers to the overall judgement of the evidence based on the previous questions.

Table 4. 2 Quality assessment based on Weight of Evidence (Gough, 2007)

Author(s)	WoE A	WoE B	WoE C	WoE D
Buac, Gross, & Kaushanskaya (2016)	high	medium	high	high
Cheung (1996)	high	medium	medium	medium
Chlapana & Tafa (2014)	high	medium	medium	medium
Hamada & Koda (2008)	medium	low	low	low
Hamada & Koda (2011)	medium	medium	medium	medium
Jubenville, Senechal, & Malette (2014)	high	low	low	low
Kan & Kohnert (2012)	high	medium	medium	high
Kaushanskaya (2012)	high	high	high	high
Kaushanskaya & Marian (2009) Exp 1/Exp 2	high/high	high/low	high/medium	high/medium
Kaushanskaya, Yoo, & Van Hecke (2013)	high	medium	medium	medium
Majerus, Poncelet, Van der Linden, & Weekes (2008)	high	high	high	high
Menjivar & Akhtar (2017)	high	high	high	high
Morini (2014)	medium	medium	medium	medium
Nair, Biedermann, & Nickels (2016)	medium	medium	medium	medium
Nakai, Lindsay & Ota (2015)	high	medium	low	low
Sasisekaran & Weisberg (2013)	medium	medium	medium	medium
Van der Hoeven & de Bot (2017)	medium	medium	low	low
Warmington, Kandru-Pothineni, & Hitch (2019)	high	high	high	high
Yoshida, Tran, Benitez, & Kuwabara (2011)	high	high	high	high

The quality assessment conducted by two independent reviewers revealed that the majority of the studies identified for this systematic review were of high or medium quality. Hamada and Koda (2008), and Jubenville, Senechal, and Malette (2014) were two exceptions that scored relatively low on three out of four criteria. This suggests that the majority of studies were conducted reliably, with appropriately matched research designs for addressing the research aims stated at the beginning, as well as taking into consideration the quality of study execution.

In addition to applying Gough's WoE, following Murphy and Unthiah (2015), an additional assessment protocol was developed to evaluate the strength of the evidence for each study, specifically to examine features of each study related to the study rationale, level of detail provided in the study, the clarity of the research questions and the variables, methodology, experimental group assignment, sample size, and the appropriateness of the procedure (Table 4.3). As before each study was assessed against each criterion as either High, Medium or Low by two reviewers and where there were discrepancies the lowest rating was agreed as the final rating. The strength of evidence assessment resulted in a rather low robustness of sample size and moderately strong robustness of bilingual sample characteristics – the scores were indeed lower than in the quality assessment conducted initially (Table 4.4).

Table 4. 3 Quality assessment criteria based on Strength of Evidence (Murphy & Unthiah, 2015)

Strength of Evidence - Methodological features (research design)	Explanation	Robustness of sample size	Language Characteristics of bilingual sample	Bilingual Definition (basis on which bilingual sample is selected)	First Language measures	Second Language measures	Executive Control measures	Paired Associate Learning measures
High	Findings are highly secure and makes a substantial contribution to the existing evidence	Sample size is justified (e.g., power analyses)	Characteristics of the bilingual sample are reported*	Clear definition of bilingualism within the context of the study	Robust and valid receptive or productive measures reported which are either standardised or widely acceptable	Robust and valid receptive or productive measures reported which are either standardised or widely acceptable	Robust and valid executive control measures reported which are either standardised or widely acceptable	Robust and valid Novel Word Learning measures reported which are either standardised or widely acceptable
Medium	Findings are moderately secure and makes a contribution to the existing evidence	Sample size is justified by not through the use of conventional means (e.g., power analyses)	Some but not all characteristics are reported	Vague definition provided within the context of the study	Concerns about the robustness and validity of receptive or productive measures	Concerns about the robustness and validity of receptive or productive measures	Concerns about the robustness and validity of executive control measures	Concerns about the robustness and validity of Novel Word Learning measures
Low	Findings are insecure and add little to the existing evidence	Sample size is not justified	None of the characteristics are reported	No definition provided	None reported	None reported	None reported	None reported

*Note: Characteristics of the bilingual sample are reported (i.e., Age of acquisition of first language and second language, languages spoken, frequency of language usage of first and second language, second language context, language proficiency, balance ratio).

Table 4. 4 Quality assessment based on Strength of Evidence (Murphy & Unthiah, 2015)

Strength of Evidence - Methodological Features (Research Design)	Robustness of sample size	Language characteristics of bilingual sample	Bilingual definition	First language measures	Second language measures	Executive control measures	Paired associate learning measures
Buac, Gross, & Kaushanskaya (2016)	low	medium	low	high	high	high	high
Cheung, 1996	low	medium	low	low	high	high	medium
Chlapana, & Tafa (2014)	low	medium	low	low	high	NA	high
Hamada & Koda (2011)	low	medium	medium	low	low	medium	medium
Hamada, & Koda (2008)	low	medium	low	low	high	NA	medium
Jubenville, Senechal, & Malette (2014)	low	medium	low	medium	low	NA	medium
Kan & Kohnert (2012)	low	medium	low	high	high	NA	high
Kaushanskaya & Marian (2009)	low	high	low	high	medium	NA	high
Kaushanskaya, Yoo, & Van Hecke (2013)	low	medium	low	high	high	NA	high
Kaushanskaya & Marian (2009)	low	medium	low	high	medium	NA	high
Kaushanskaya, 2012	low	medium	low	high	medium	high	high
Majerus, Poncelet, Van der Linden, & Weekes (2008)	low	medium	low	high	high	high	high
Menjivar, & Akhtar (2017)	low	medium	low	medium	medium	NA	high
Morini (2014)	low	medium	low	low	low	high	medium
Nair, Biedermann, & Nickels (2016)	low	high	medium	low	low	NA	medium
Nakai, Lindsay & Ota (2015)	low	medium	low	low	low	NA	high
Sasisekaran & Weisberg (2013)	low	medium	medium	high	medium	high	medium
van der Hoeven & de Bot (2017)	low	medium	low	medium	low	high	medium
Warmington, Kandru-Pothineni, & Hitch (2019)	high	high	high	high	high	high	high
Yoshida, Tran, Benitez, & Kuwabara (2011)	low	medium	high	high	medium	high	high

4.2.3 Data Extraction

The data extraction protocol was piloted and the data from the eligible papers extracted for the purpose of the systematic review. A comprehensive record of data was coded and input into an electronic spreadsheet. The outcome measures were assigned to the relevant domains, such as recall and recognition rates in paired associate learning tasks, as well as executive control measures (working memory and inhibition/attentional control). Apart from the main bibliographic data (i.e., author(s), publication title, journal name, year published, publication format and publication status), the following information was also recorded for each study.

Participants' characteristics (i.e., age, gender, and ethnicity) as well as sample sizes (i.e., number of participants, attrition, and exclusion rates) were extracted and investigated in more detail. For some studies, there were multiple groups of language participants (i.e., the first experiment by Kaushanskaya (2012) compared performance of two monolingual groups, whereas the study by van der Hoeven and de Bot (2017) recruited three groups of bilingual participants). A more detailed description of characteristics of participants is shown in Table 4.6.

Another important information extracted for further analyses was also the socioeconomic status which has been found to have an impact on paired associate learning performance (Buac, Gross, & Kaushanskaya, 2016). Additionally, participants' language status was noted (whether monolingual or bilingual) along with the additional details required for bilingual samples. Here, the age of language acquisition and country of birth or residence (with the length of residence in the country if applicable) as well as first language and second language spoken by participants were extracted. Language proficiency for both languages was also noted. In terms of the linguistic details, typology and distance between two languages, if applicable, was noted.

Another set of data extracted for the meta-analysis concerned study design aspects, such as study types (i.e., concurrent, longitudinal, experimental, quasi-experimental) and the number of assessment points. Measurement data of key interest included the number of novel words recognised and recalled in paired associate learning tasks, number of errors, as well as working memory and inhibition/attentional control performance. Additionally, details regarding experimental manipulation (i.e., method of delivery, type of stimuli, and time between training sessions) was also included in data extraction.

To enable meta-analytic calculations, the results data was also extracted in the form of effect sizes (where available). An effect size refers to a number that examines the strength of the relationship between two variables and is a standardised measure of the difference between two groups divided by the pooled standard deviation of the two groups of interest (Wilkinson, 1999). Where effect sizes were not reported, correlational values, means, standard deviations (*SD*), *t*-values, and *F*-values were extracted to enable further data transformations. The moderators included in additional analyses comprised participants' age, their socioeconomic status and nonverbal IQ (general intelligence without the confound of one's language ability), age of language acquisition, as well as publication status.

Once the data extraction had been completed, the quality assessment of included studies was conducted by two independent reviewers.

4.2.4 Calculation

4.2.4.1 Calculation of Effect Sizes

Effect sizes are the key outcomes or research studies and they are considered important for three main reasons Firstly, effect sizes indicate the magnitude of results in a metric that is easily understood by other researchers and regardless of the scale that was employed to investigate the dependent variables. Secondly, effect sizes enable conducting meta-analyses to draw overall cumulative conclusions of the phenomenon across all relevant studies. Finally, they are also useful in planning new studies to determine the sample size to observe statistically significant results.

This meta-analysis calculated Hedges g which is also a measure of effect size (Hedges & Olkin, 1985). Hedges *g*, also known as *the corrected effect size*, was computed for each study in the meta-analysis to indicate the magnitude of the effect size as it has been claimed to outperform Cohen's *d* when sample sizes are relatively small (< 20 ppts). Following Cohen's scale (1988), an effect size of .2 was identified as a small effect, .5 was interpreted as a moderate effect, and .8 was referred to as a large effect.

For the meta-analytic purposes, the effect sizes, reported as Cohen's *d*, were extracted from the studies and where these were not available, they were calculated based on the other values provided (means, standard deviations, correlations, *F*- and *t*-values).

The following formula was implemented to calculate Hedges' *g* based on group means (i.e., M_1 – mean for group 1 (monolinguals), M_2 – mean for group 2 (bilinguals)) and pooled and weighted standard deviations.

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD^*_{pooled}}$$

To calculate pooled and weighted standard deviations, the following formula was used where SD_1 refers to standard deviation of group 1 (monolinguals) and SD_2 for group 2 (bilinguals).

$$SD_{pooled} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$$

Results data including F- and t-values was input into Psychometrica converter (Lenhard & Lenhard, 2016) to calculate effect sizes in a form of Hedges *g*.

Studies with insufficient data to calculate effect sizes were not included in the analysis. If a study reported paired associate learning performance and working memory or inhibition/attentional control performance, these results were reported as separate comparisons. Where studies comprised more than two groups compared against the third group (i.e., two bilingual groups and a monolingual group), the two groups of similar characteristics (such as age of acquisition, scores on language measures) were averaged together. The rationale behind this procedure was not to artificially inflate the number of participants in the calculations of Hedges *g*.

Some studies comprised more than one experiment, and where this was the case and both experiments were of interest to the systematic review, the experiments were extracted and treated as separate studies. Kaushanskaya (2012) examined executive control mechanisms underpinning monolingual and bilingual paired associate learning and the role of phonological memory in this process. In the first experiment, she employed phonologically familiar novel words which were created on the basis of participants' first language phonemes, English. Experiment 2 comprised phonologically unfamiliar novel words which were constructed using non-English phonemes as well. As both experiments were deemed eligible but comprised varied stimuli and separate groups of monolingual and bilingual participants, they were recorded as two separate studies for the purpose of the systematic review. Similarly, Hamada and Koda (2008) recruited the same participants for two experiments, however, the experiments differed in the number of items used in each.

Episodic memory (as measured by paired associate learning rates: recall and recognition), working memory and executive control (i.e., inhibition/attentional control) performance rates were extracted or calculated for each paper and included in a between-group (two or more groups) comparison analyses. The results are presented in forest plots.

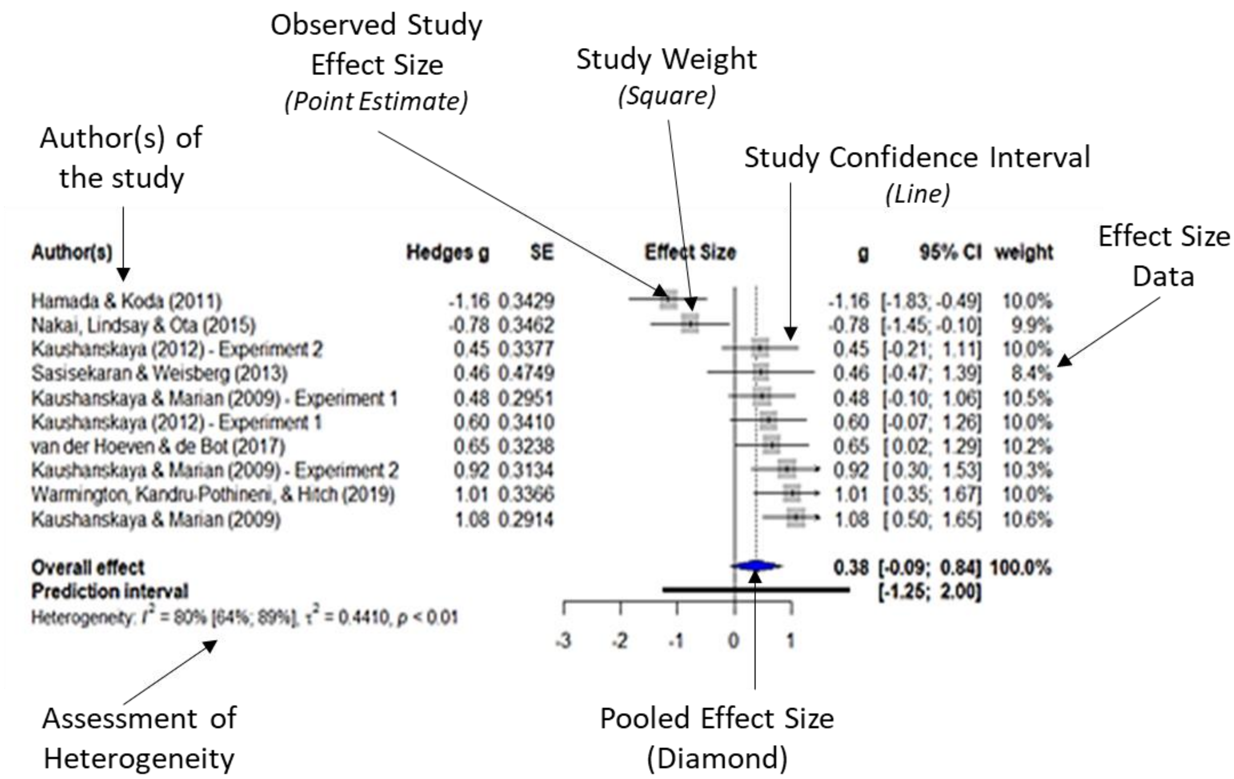


Figure 4. 2 Key elements of a forest plot (adapted from Harrer, Cuijpers, Furukawa, & Ebert, 2021)

A forest plot is a common way to provide a visual presentation of a meta-analysis. They offer a visual display of observed study effect size, as well as study weight (i.e., proportion of contribution to the combined effect), and 95% confidence intervals. Additionally, they present an overall effect size of the studies in the form of a diamond (Figure 4.2 provides a visual guide to interpreting forest plots). On the left, the forest plot lists author(s) and studies included in the meta-analysis along with Hedges *g* reported for each study. For each of these studies, there is also a visual representation of the effect size of the study. The point estimate of a study is usually shown on the x-axis and is further supplemented by a horizontal line that represents the confidence interval range for that study. The longer the horizontal lines are (i.e., the wider the 95% confidence interval), the less reliable the study results. The gray square surrounding the point estimate represents the study weight (i.e., each study's contribution to the overall effect size in terms of the amount of information it contains) and its size depends on the value: the square is smaller for studies with lower study weight and larger for studies with larger weight.

The bottom part of the forest plot presents the results of the meta-analysis. Here, a diamond-shaped indicator refers to the average cumulative effect size across the studies. This is indeed a single and more precise value that is obtained from combining all the measures from the eligible studies. The length of the diamond on x-axis indicates the 95% confidence interval range of the pooled result. Also, a vertical reference line represents a line of no effect (i.e., position at which there is no clear difference between the groups). If the diamond touches the line of no effect, the overall result is not statistically significant. Additionally, a thin black horizontal line under the diamond represents the prediction interval around the pooled effect size. Depending on the outcome variable, a *positive* effect is represented by the square and 96% confidence interval being located on the left or on the right of the line of no effect.

In this thesis, meta-analyses were utilised to examine (1) the role of bilinguals' executive control processes in episodic memory and (2) role of bilinguals' vocabulary knowledge (in both languages) in their performance on episodic memory tasks, as measured by paired associate learning.

Effect sizes and standard errors for each eligible study were extracted and inputted into the software. Where these were not provided, individual effect sizes for each study were computed as well as their weight, and an aggregate effect size, its confidence interval, and p-value calculated. Studies with larger sample sizes (fewer than 20 participants) were given more weight than the ones with smaller samples (Kline, 2004) and the standard error of Hedges' *g* unbiased estimate of the mean effect size (overall effect) was calculated. The output comprised 95% confidence intervals displayed for each weighted mean effect size indicating whether there is a statistically significant difference between the groups of participants. 95% confidence intervals that were above zero suggested the advantage towards the bilingual group whereas confidence intervals below zero indicated that the result was in favour of monolinguals.

4.2.4.2 Assessment of Heterogeneity

The term *between-study heterogeneity* refers to the extent to which individual effect sizes differ from each other within a meta-analysis (Harrer, Cuijpers, Furukawa, & Ebert, 2021). High heterogeneity may in fact be caused by a number of factors, such as different treatment intensity or multiple subgroups within a study. It is of a great importance that such variations in the eligible studies are accounted for in a meta-analytic analysis. In extreme cases, it may be that very high between-study heterogeneity indicates that studies have nothing in common and thus the overall effect size should not be interpreted at all.

Measures of between-study heterogeneity were also employed in this meta-analysis, and the homogeneity of variance was tested by the means of Cochran’s *Q* statistic and I^2 (Higgins & Thompson, 2002). Cochran's *Q*, often employed to distinguish sampling error from actual between-study heterogeneity, was computed to examine the variance among all the effect sizes included in the studies of interest (Borenstein, Hedges, Higgins, & Rothstein, 2009). A significant *Q*-test is indicative of substantial differences between the true effect size among the included studies. As Q-statistic is dependent upon the statistical power and the size of a meta-analysis, an additional heterogeneity measure of I^2 (Higgins & Thompson, 2002) was also calculated and reported to complement the interpretation of Cochran’s *Q* result. I^2 quantifies the between-study amount of variance (variance that is not caused by sampling error) and can be interpreted as follows: I^2 of 25% as low heterogeneity, 50% as moderate, and 75% as substantial heterogeneity. While *Q*-statistic is sensitive to a number of studies included, the I^2 statistic remains unaffected (Harrer, Cuijpers, Furukawa, & Ebert, 2021).

Depending on the results of the heterogeneity measures, meta-analyses were computed using a random- or a fixed-effects model; where $I^2 \leq 50\%$, a fixed effect model of meta-analysis was employed as studies are then considered homogeneous. With $I^2 > 50\%$, random effect model for meta-analysis due to studies being considered heterogeneous.

4.2.4.3 Assessment of Bias

When conducting a meta-analysis, it is crucial to identify and extract data that is comprehensive and representative of the question under examination. According to Borenstein et al. (2009) as studies with statistically significant results and the ones comprising large effect sizes are more likely to be published than studies with nonsignificant findings and thus the examination of publication bias is of great importance. The term *bias* refers to the inclination or prejudice for or against something and it is generally perceived to be unfair (American Psychological Association, 2022). To minimise the risk of publication bias a number of steps were undertaken: (a) the review included studies published in England and in other countries; (b) the literature search attempted to access grey literature, such as conference proceedings, reports, presentations, and posters from scientific meetings, as well as unpublished papers and manuscripts; (c) funnel plots were computed to examine whether the publication bias was present in the review and their asymmetry assessed by Egger's test of the intercept (Egger, Smith, Schneider, & Minder, 1997).

Additionally, Duval and Tweedie's (2000) trim-and-fill procedure was employed to estimate the magnitude of a true effect size by imputing unpublished studies to create an unbiased symmetrical funnel plot. It also provided an effect size adjusted for the publication bias which was then compared to the effect size computed in the initial analyses. According to Borenstein et al. (2009), if these two effect sizes are largely unchanged, the publication bias does not have a significant impact. However, if the initial effect size and an adjusted effect size vary significantly, there is likely an underestimation or an underestimation present in the initial findings reflecting a substantial publication bias.

4.2.4.4 Sensitivity Analysis

The literature search and screening were conducted by two reviewers with the engagement of the third reviewer when conflicting decisions arose. A sensitivity analysis to identify outlier effect sizes was conducted and the data was analysed again with outliers removed. Indeed, heterogeneity of studies may arise due to the presence of a few studies whose results conflict with the remaining studies. Thus, it has been suggested to conduct analyses with and without the outliers to compare yielded results and ensure they are robust (Deeks, Higgins, & Altman, 2019).

All possible studies were included, both published and unpublished work to ensure a thorough retrieval was conducted. Additionally, publication bias was explored by means of the *p*-curve method.

4.2.4.5 Data Synthesis

Due to numerous methodological differences across the studies included, the effect sizes arising from them may differ. Thus, the meta-analyses were computed based on the random-effects model, unless heterogeneity measures proposed otherwise (Harrer, Cuijpers, Furukawa, & Ebert, 2021). The random-effects model makes certain assumptions about the underlying variability between the studies as it assumes that studies vary in terms of target populations and experimental setups (i.e., treatment length and intensity) among other aspects. It addresses the concern regarding various possible differences in true effect sizes of studies. Moreover, this model allows to account for more substantial variance than when studies are drawn from a homogeneous population (Hedges & Vevea, 1996). In other words, the random-effects model assumes that the variance between individual eligible studies stem from not only sampling error but that there is also another source of variance. Indeed, in this model, "each study is seen as an independent draw of a universe of populations" (Harrer, Cuijpers, Furukawa, & Ebert, 2021, sec. 4.1.2). Contrary to the random-effects model, the fixed-effect model assumes that all studies included for meta-analysis come from a homogeneous population. Indeed, this is why an assessment of heterogeneity is so important in decision making process regarding an appropriate model selection for further meta-analytic considerations.

The significance level was set at .05 for all analyses.

4.3 Results

4.3.1 Background of Studies

The description of included studies with the publication status, type of study (whether the study is behavioural or imaging), sample size and breakdown for each group is presented in Table 4.5. Additionally, Table 4.5 reports on the population of the study and presents whether the study comprised monolingual and bilingual samples, monolingual samples only, or bilingual samples only.

Table 4. 5 Background information of studies included in the systematic review

Author(s)	Publication Status	Type of Study	Population	Power Analysis
Buac, Gross, & Kaushanskaya (2016)	Published	Behavioural	Monolingual (*N* = 36); Bilingual (*N* = 46)	No
Cheung (1996)	Published	Behavioural	Bilingual (*N* = 84)	No
Chlapana & Tafa (2014)	Published	Behavioural	Bilingual (*N* = 87)	No
Hamada & Koda (2008)	Published	Behavioural	Bilingual (*N* = 35)	No
Hamada & Koda (2008)	Published	Behavioural	Bilingual (*N* = 35)	No
Hamada & Koda (2011)	Published	Behavioural	Monolingual (*N* = 20); Bilingual (*N* = 20)	No
Jubenville, Senechal, & Malette (2014)	Published	Behavioural	Bilingual (*N* = 64)	No
Kan & Kohnert (2012)	Published	Behavioural	Bilingual (*N* = 32)	No
Kaushanskaya & Marian (2009)	Published	Behavioural	Monolingual (*N* = 24); Bilingual (*N* = 24)	No
Kaushanskaya & Marian (2009)	Published	Behavioural	Monolingual (*N* = 20); Bilingual (*N* = 40)	No
Kaushanskaya (2012)	Published	Behavioural	Monolingual (*N* = 36); Bilingual (*N* = 18)	No
Kaushanskaya (2012)	Published	Behavioural	Monolingual (*N* = 36); Bilingual (*N* = 18)	No
Kaushanskaya, Yoo, & Van Hecke (2013)	Published	Behavioural	Bilingual (*N* = 81)	No
Majerus, Poncelet, Van der Linden, & Weekes (2008)	Published	Behavioural	Bilingual (*N* = 52)	No
Menjivar & Akhtar (2017)	Published	Behavioural	Monolingual (*N* = 16); Bilingual (*N* = 32)	No
Morini (2014)	Unpublished	Behavioural	Monolingual (*N* = 32); Bilingual (*N* = 32)	No
Nair, Biedermann, & Nickels (2016)	Published	Behavioural	Monolingual (*N* = 20); Bilingual (*N* = 40)	No
Nakai, Lindsay & Ota (2015)	Published	Behavioural	Monolingual (*N* = 18); Bilingual (*N* = 36)	No
Sasisekaran & Weisberg (2013)	Published	Behavioural	Monolingual (*N* = 13); Bilingual (*N* = 7)	No
van der Hoeven & de Bot (2017)	Published	Behavioural	Bilingual (*N* = 45)	No
Warmington, Kandru-Pothineni, & Hitch (2019)	Published	Behavioural	Monolingual (*N* = 20); Bilingual (*N* =20)	Yes
Yoshida, Tran, Benitez, & Kuwabara (2011)	Published	Behavioural	Monolingual (*N* = 20); Bilingual (*N* = 20)	No

4.3.1.1 Participants' Characteristics

The number of participants across the studies totaled 1,185 with 311 monolingual and 730 bilingual participants. The summary of the characteristics of the participants with the mean age, gender distribution, and languages spoken are presented in Table 4.6.

In some studies, multiple groups of participants were recruited. For each subgroup of language groups mean age and standard deviation was calculated. Also, Table 4.6 comprises the information regarding participants' first and second languages, gender distribution for each study and a language environment, as a monolingual versus bilingual context, of the samples.

Table 4. 6 Participants' characteristics

Author(s)	Language Group					Language			Gender	Country of Residence	Language Environment
	Monolingual		Bilingual			Monolingual	Bilingual		(Females/ Males)		
	Gr. 1	Gr. 2	Gr. 1	Gr. 2	Gr. 3	First	First	Second			
Buac, Gross, & Kaushanskaya (2016)	6.34 (.84)	NA	6.24 (.76)	NA	NA	English	Spanish	English	44/38	US	monolingual
Cheung (1966)	NA	NA	12.2 (.43)	NA	NA	NA	Cantonese Chinese	English	40/44	China	monolingual
Chlapana & Tafa (2014)	NA	NA	5.14	NA	NA	NA	Mixed (9 languages)	Greek	45/42	Greece	monolingual
Hamada & Koda (2008)	NA	NA	NR	NR	NR	NA	Mixed (2 languages)	English	NR	US	monolingual
Hamada & Koda (2008)	NA	NA	NR	NR	NR	NA	Mixed (2 languages)	English	NR	US	monolingual
Hamada & Koda (2011)	21.1 (2.21)	NA	21.2 (3.97)	NA	NA	English	Chinese	English	26/14	US	monolingual
Jubenville, Senechal, & Malette (2014)	NA	NA	9.25	NA	NA	NA	French	Mixed (2 languages)	30/34	Canada	bilingual
Kan & Kohnert (2012)	NA	NA	4.5 (.50)	NA	NA	NA	White Hmong	English	16/16	US	monolingual
Kaushanskaya (2012) Experiment 1	21.77 (.99)	20.17 (.95)	21.77 (.74)	NA	NA	English	English	Spanish	31/23	US	monolingual
Kaushanskaya (2012) Experiment 2	20.49 (.74)	20.92 (.86)	21.67 (.95)	NA	NA	English	English	Spanish	30/24	US	monolingual

Kaushanskaya, & Marian (2009)	21.57 (.57)	NA	20.83 (.63)	NA	NA	English	English	Spanish	NR	US	monolingual
Kaushanskaya, & Marian (2009)	21.64 (.62)	NA	20.83 (.63)	21.10 (.63)	NA	English	English	Mixed (2 languages)	NR	US	monolingual
Kaushanskaya, Yoo, & Van Hecke (2013)	NA	NA	22.23 (3.92)	22.29 (5.46)	NA	NA	English	Spanish	46/35	US	monolingual
Majerus, Poncelet, Van der Linden, Weekes, & Brendan (2008)	NA	NA	21	NA	NA	NA	English	French	39/13	UK	monolingual
Menjivar & Akhtar (2017)	4.5 (3.90)	NA	4.5 (3.10)	4.4 (3.30)	NA	English	Mixed (9 languages)	English	Y	US	monolingual
Morini (2014)	20.4 (2.30)	NA	20.4 (2.30)	NA	NA	English	Mixed (18 languages)	English	NR	US	monolingual
Nair, Biedermann, & Nickels (2016)	22.7 (1.45)	NA	22.1 (2.02)	21.6 (1.68)	NA	English	Tamil	English	25/35	US	monolingual
Nakai, Lindsay, & Ota (2015)	27	NA	26	27	NA	English	Mixed (2 languages)	English	37/17	NR	bilingual
Sasisekaran, & Weisberg, (2013)	21.2 (4.40)	NA	22.7 (2.56)	NA	NA	English	English	Mixed (4 languages)	NR	US	monolingual
van der Hoeven, & de Bot (2017)	NA	NA	22.4 (2.20)	50.3 (2.60)	76.0 (4.4)	NA	Dutch	French	28/17	NR	NA
Warmington, Kandru-Pothineni, & Hitch (2018) Experiment 2	21.58	NA	23.42	NA	NA	English	Hindi	English	16/24	UK	monolingual
Yoshida, Tran, Benitez, & Kuwabara (2011)	3.06	NA	3.24	NA	NA	English	English	Mixed (6 languages)	NR	US	monolingual

Note: Group column has been further divided into 1 and 2 columns for monolingual participants and 1, 2, and 3 for bilingual participants.

4.3.1.2 Language Background, Executive Control, and Working Memory Measures

Language background of participants in the studies of interest was assessed by means of standardised tests and questionnaires (reported in Table 4.7). Some studies comprised language background information in a form of age of acquisition for the first and the second language, languages spoken by a participant, frequency of language usage for first and second language, second language learning context, language proficiency for the languages known, and a balance ratio.

In studies where executive control performance was of interest, and thus executive control tasks employed, the measures examining working memory and inhibition/attentional control were of interest.

Table 4. 7 Language background and executive control measures

Author(s)	Language Background	Executive Control
Buac, Gross, & Kaushanskaya (2016)	Receptive Vocabulary: *Peabody Picture Vocabulary Test III* (Dunn & Dunn, 1997); T*est de Vocabulario en Imágenes Peabody* (TVIP: Dunn, Padilla, Lugo, & Dunn, 1986) Expressive Vocabulary: *Woodcock–Johnson III Tests of Achievement* (Woodcock, McGrew, & Mather, 2001); *Batería III Woodcock-Muñoz Pruebas de aprovechamiento* (Muñoz-Sandoval, Woodcock, McGrew, & Mathers, 2005)	Working Memory: *Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 1999):* Digit Recall, Nonword Repetition, Backward Digit Recall; Word Recall (bespoke)
Cheung (1996)	Receptive Vocabulary: *Crichton Vocabulary Scale* (Raven, Court, & Raven, 1983) Reading: Bespoke reading comprehension test	Working Memory: Word Recall, Nonword Recall (bespoke)
Chlapana & Tafa (2014)	Receptive Vocabulary: *Peabody Picture Vocabulary Test-R* (Dunn & Dunn, 1981) standardised into Greek (Simos, Sideridis, Protopapas, & Mouzaki, 2011)	NA
Hamada & Koda (2008)	Test of English as a Foreign Language (TOEFL)	NA
Hamada & Koda (2011)	Not Reported	NA
Jubenville, Senechal, & Malette (2014)	NR	NA
Kan & Kohnert (2012)	Picture identification and picture naming (Kan & Kohnert, 2005)	NA

Kaushanskaya, Yoo, & Van Hecke (2013)	Receptive Vocabulary: *Peabody Picture Vocabulary Test III* (Dunn & Dunn, 1997) Expressive Vocabulary: Expressive Vocabulary Test (Williams, 1997) Reading: *Woodcock–Johnson III Tests of Achievement* (Woodcock, McGrew, & Mather, 2001) Self Report: *Language Experience and Proficiency Questionnaire* (Marian, Blumenfeld, & Kaushanskaya, 2007)	Working Memory: *Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 1999):* Digit Recall, Nonword Repetition; Woodcock-Johnson Test of Cognitive Abilities (Woodcock et al., 2001): Backward Digit Recall
Kaushanskaya (2012)	Receptive Vocabulary: *Peabody Picture Vocabulary Test III* (Dunn & Dunn, 1997) Self Report: *Language Experience and Proficiency Questionnaire* (Marian, Blumenfeld, & Kaushanskaya, 2007)	Working Memory: *Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 1999):* Digit Recall
Kaushanskaya & Marian (2009)	Receptive Vocabulary: *Peabody Picture Vocabulary Test III* (Dunn & Dunn, 1997) Expressive Vocabulary: *Expressive Vocabulary Test* (Williams, 1997) Reading: *Woodcock–Johnson III Tests of Achievement* (Woodcock, McGrew, & Mather, 2001) Self Report: *Language Experience and Proficiency Questionnaire* (Marian, Blumenfeld, & Kaushanskaya, 2007)	Working Memory: *Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 1999):* Digit Recall
Majerus, Poncelet, Van der Linden, & Weekes (2008)	Receptive Vocabulary: *British Picture Vocabulary Scales* (Dunn, Dunn, Whetton, & Pintilie, 1982); *Echelles de vocabulaire en images de Peabody* (EVIP; Dunn, Thériault-Whalen,& Dunn, 1993) Self Report: Bespoke Questionnaire	NA

Menjivar & Akhtar (2017)	Parental Report: Bespoke Questionnaire	NA
Morini (2014)	NR	Working Memory: Digit Recall, Backward Digit Recall (bespoke) Attention: Attention Network Task (Weaver, Bédard, & McAuliffe, 2013)
Nair, Biedermann, Nickels (2016)	Self Report: Bespoke Questionnaire	Working Memory: Nonword Repetition (bespoke)
Nakai, Lindsay & Ota (2015)	NR	NA
Sasisekaran & Weisberg (2013)	Expressive Vocabulary: Expressive Vocabulary Test (Williams, 1997) Self Report: Bespoke Questionnaire	Working Memory: Nonword Repetition (Dollaghan & Campbell, 1998); *Clinical Evaluation of Language Fundamentals test-IV* (CELF-IV; Semel Wiig, & secord, 1995): Digit Recall, Sentence Recall
van der Hoeven & de Bot (2017)	Self Report: Bespoke Questionnaire	Working Memory: Backward Digit Recall (bespoke)
Warmington, Kandru-Pothineni, & Hitch (2019)	Expressive Vocabulary: *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) Self Report: Bespoke Questionnaire	Working Memory: *Automated Working Memory Assessment* (AWMA; Alloway, 2007): Digit Recall, Nonword Recall, Dot Matrix, Block Recall, Listening Recall, Backward Digit Recall, Odd One Out, Spatial Recall Attention: Stop, Signal Reaction Time, Flanker
Yoshida, Tran, Benitez, & Kuwabara (2011)	Expressive Vocabulary: *MacArthur–Bates Communicative Development Inventories* (MCDI; Fenson et al., 1993)	Attention: Attention Network Task (bespoke)

4.3.1.3 Paired Associate Learning Tasks

A summary of different types of stimuli used in paired associate learning tasks with details regarding the number of stimuli learned can be found in Table 4.8. Additionally, the number of assessment points, in other ways time between training and assessment of learning, are also reported.

Table 4. 8 Types of paired associate stimuli

Author(s)	Number of Stimuli	Type of Stimuli	Number of Learning Assessment Points	Learning Assessment Mode
Buac, Gross, & Kaushanskaya (2016)	16	novel words & real and novel objects	1	Recognition
Cheung (1996)	3	real words & translations	2	Trials to criterion*
Chlapana & Tafa (2014)	56	real words	2	Recognition
Hamada & Koda (2008) - Experiment 1	40	novel words & real objects	1	Recall
Hamada & Koda (2008) - Experiment 2	16	novel words & real objects	1	Recognition
Hamada & Koda (2011)	12	novel words & real objects	1	Both
Jubenville, Senechal, & Malette (2014)	12	novel words & novel objects	1	Recall
Kan & Kohnert (2012)	16	novel objects & novel words	4	both
Kaushanskaya & Marian (2009)	48	novel words & translations	2	Both
Kaushanskaya & Marian (2009)	48	novel words & translations	2	Both
Kaushanskaya (2012) - Experiment 1	48	novel words & translations	2	Both
Kaushanskaya (2012) - Experiment 2	48	novel words & translations	2	Both
Kaushanskaya, Yoo, & Van Hecke (2013)	48	novel words & real and novel objects	2	Recognition
Majerus, Poncelet, Van der Linden, & Weekes (2008)	8	real words & novel words	2	Recall
Menjivar & Akhtar (2017)	6	novel words & real and novel objects	1	Recognition
Morini (2014)	78	novel words & novel objects	1	Recall
Nair, Biedermann, & Nickels (2016)	10	novel words & real objects	1	Both

Nakai, Lindsay & Ota (2015)	36	novel words	5	Recall
Sasisekaran & Weisberg (2013)	7	novel words	1	Recall
van der Hoeven & de Bot (2017)	116	real & novel words & translations	2	Recall
Warmington, Kandru-Pothineni, & Hitch (2019)	40	novel words & novel objects	2	Both
Yoshida, Tran, Benitez, & Kuwabara (2011)	4	novel words & real objects	1	Recognition

Note: *Trials to Criterion refers to the number of responses necessary to achieve a specific performance goal.

4.3.2 Descriptive Statistics

The effect sizes and standard errors for each study were extracted and analysed using *meta* package (Schwarzer, 2007) in R software. The procedures outlined by Harrer, Cuijpers, Furukawa, and Ebert (2019) were followed. Forest plots were screened visually for any potential outliers and two heterogeneity measures (i.e., Cochran's Q and I^2) employed and reported for each meta-analysis. Additional analysis for detecting outliers and influential cases was utilised where appropriate to examine the differences in the overall effect size when removing extreme effect sizes. Moreover, *p*-curve analysis to assess the publication bias and obtain an estimation of a true effect size in the data collected.

4.3.2.1 Is there a Bilingual Advantage in Episodic Memory as measured by Paired Associate Learning Tasks? Is the Bilingual Advantage in Episodic Memory Task-Dependent (i.e., how Paired Associate Learning is assessed: Recall versus Recognition)?

4.3.2.1.1 Recall

Ten studies were included in the comparison of recall performance between monolingual and bilingual samples (Figure 4.3). Other studies were excluded from this analysis due to the fact that they were composed of bilingual samples exclusively or the lack of information regarding the number or novel words recalled by participants during the testing/assessment of learning phase.

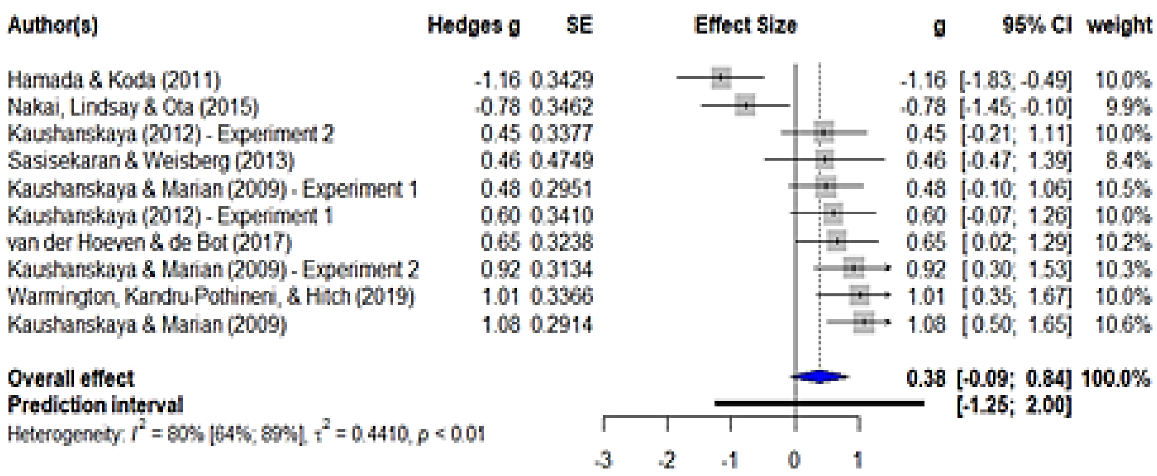


Figure 4. 3 Forest plot with individual and summary effect size estimates for recall with monolinguals on the left and bilinguals on the right

Two measures to assess the between-study heterogeneity were employed: Cochran's *Q* statistic and I^2 test. Cochran's *Q* statistic revealed a statistically significant difference between the true effect size among the included studies, $Q = 44.72$, $df = 9$, $p < .001$. I^2 test resulted in a substantial heterogeneity (79.9%) and thus the random effect model was chosen for this meta-analysis. There was a not statistically significant effect size observed between the monolingual and bilingual in terms of their recall of the novel words, $g = .38$, 95% 95% CI [-.09; .84], $p = .109$. This was further corroborated by the fact that the diamond touched the line of no effect as presented on a forest plot (Figure 4.3).

Further analysis for detecting outliers and influential cases identified two studies with extreme effect sizes: Nakai, Lindsay and Ota (2015) and Hamada and Koda (2011). After exclusion of these, the overall effect size estimate changed, $g = .73$, 95% CI [.50; .96], $p < .001$, suggesting a statistically significant large difference in recall between bilinguals and monolinguals. Assessments of between-study heterogeneity also yielded different results: Q-statistic became statistically nonsignificant, $Q = 4.40$, $df = 7$, $p = .733$ indicating no difference between the true effect size among the studies. This was further corroborated by $I^2 = 0$.

The output provided by Egger's test suggested presence of the publication bias due to the significant funnel asymmetry (Intercept = -5.08, 95% CI [-17.23; 7.08], $t = -.082$, $p = .438$). No studies were imputed in Duval & Tweedie's trim-and-fill procedure (2000).

Additional *p* -curve analysis revealed that 6 studies were included in the analysis and 5 of them had the *p*-value lower than .025. Power estimate of the analysis equaled 59% with 95% CI [16.7% - 88.6%]. The evidential value was present which means that there is indeed a *true* effect size behind the findings for recall, and these are not resulting only from the publication bias.

4.3.2.1.2 Recognition

Eleven studies were included in the comparison of recognition performance between monolingual and bilingual samples (Figure 4.4). Other studies were excluded from this analysis due to comprising bilingual samples exclusively or the lack of information regarding the number or novel words recognised by participants during the testing/assessment of learning.

Two measures to assess the between-study heterogeneity were employed: Cochran's *Q* statistic and I^2 test. Cochran's *Q* statistic revealed a statistically significant difference between the true effect size among the included studies, $Q = 46.82$, $df = 10$, $p = .001$. I^2 test resulted in a substantial heterogeneity (78.6%) and thus the random effect model was chosen for this meta-analysis. There was a statistically significant effect size observed between the monolingual and bilingual in terms of their Recognition of the novel objects, *g* = .44, 95% CI [.05; .83], *p* = .027 indicating a bilingual advantage (Figure 4.4).

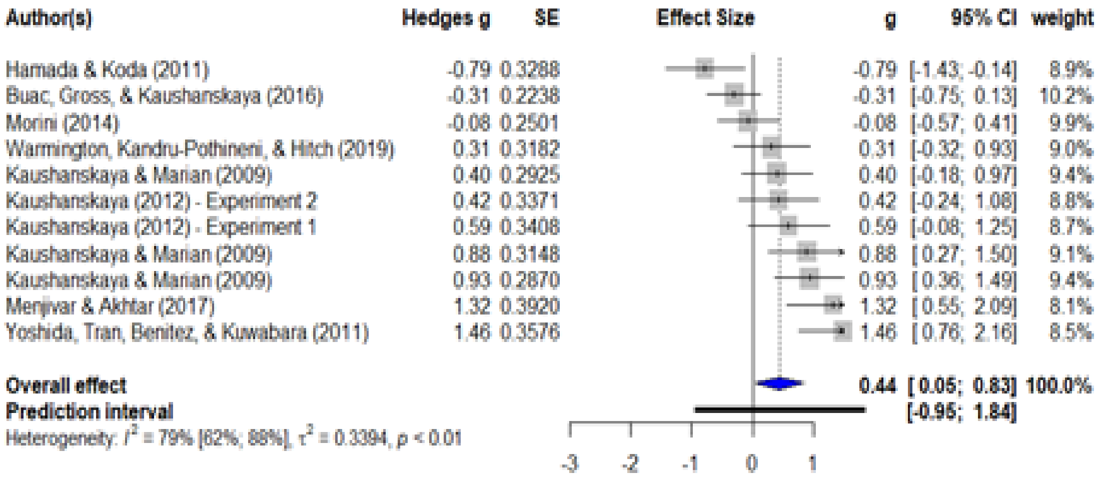


Figure 4. 4 Forest plot with individual and summary effect size estimates for recognition with monolinguals on the left and bilinguals on the right

Further analysis for detecting outliers and influential cases identified one study with an extreme effect size: Hamada & Koda (2011). After its exclusion, the overall effect size estimate changed, *g* = .56, 95% CI [.19; .92], *p* = .003, suggesting a statistically significant difference in recognition between bilinguals and monolinguals. Assessments of between-study heterogeneity also yielded different results: Q-statistic remains statistically significant, *Q* = 33.90, *df* = 9, *p* = .001 indicating a statistically significant difference between the true effect size among the studies. I^2 decreased slightly to 73.5%.

The output provided by Egger's test suggested potential presence of the publication bias due to the marginally significant funnel asymmetry (Intercept = 7.90, 95% CI [1.23; 14.56], *t* = 2.31, *p* = .047). Duval and Tweedie's (2000) trim-and-fill procedure identified and trimmed 3 studies. The overall effect that was estimated by this procedure equaled *g* = .24. When compared with the initial pooled effect size of *g* = .44 and *g* = .56 after the removal of one outlier study, the initial results were possibly overestimated due to publication bias identified by trim-and-fill procedure, and the true effect size is small with *g* = .24 when controlling for selective publication.

Additional *p*-curve analysis revealed that 6 studies were included in the analysis and all had the *p*-value lower than .025. Power of the analysis was 86% with 95% CI [56.4% - 96.8%].

The evidential value was present which means that there is indeed a *true* effect size behind the findings for recognition, and these are not resulting only from the publication bias.

4.3.2.1.3 Summary

Effect sizes were calculated to determine if there was a bilingual advantage in episodic memory as measured by recall and recognition in paired associate learning tasks. Only studies which recruited both monolinguals and bilingual samples were included in the analyses. Recall between monolingual and bilingual samples was compared in ten eligible studies. This limited number of publications was due to single samples of monolingual speakers or bilingual speakers as well as research where the number of words recalled was not reported.

The analysis of recall revealed substantial heterogeneity and a statistically significant large effect size was observed between the language groups after removing two outliers (g = .73). Similarly, for the between-group comparisons of recognition, eleven studies were included. Again, substantial heterogeneity among the publications was observed and a statistically significant small effect size was noted between the monolingual and bilingual Recognition performance (g = .24) after a removal of one study with an extreme effect size. This is indicative of bilingual participants performing better on both recall and recognition compared to monolingual speakers which is in line with previous research (i.e., Bialystok et al., 2003; Kaushanskaya & Marian, 2009).

However, the results of these meta-analyses should be treated with caution due to the very limited number of eligible studies and high heterogeneity between studies included in the analysis.

4.3.2.2 What is the Role of Working Memory and Executive Control (i.e., Inhibition/Attentional Control) in Bilingual Episodic Memory Performance as measured by Paired Associate Learning Tasks?

4.3.2.2.1 Working Memory

Six studies with 15 working memory measures were included in the investigation of the role of working memory in bilingual episodic memory performance (Figure 4.5).

Two measures to assess the between-study heterogeneity were employed: Cochran's Q statistic and I^2 test. Cochran's Q statistic revealed a statistically significant difference between the true effect size among the included studies, $Q = 32.19$, $df = 14$, $p = .004$. I^2 test resulted in a moderate heterogeneity (56.5%) and thus the random effect model was chosen for this meta-analysis. The results revealed a significant medium effect size, *g* = .53, 95% CI [.31; .74], $p < .001$ suggesting that working memory plays a significant role in bilingual episodic memory performance as measured by paired associate learning tasks.

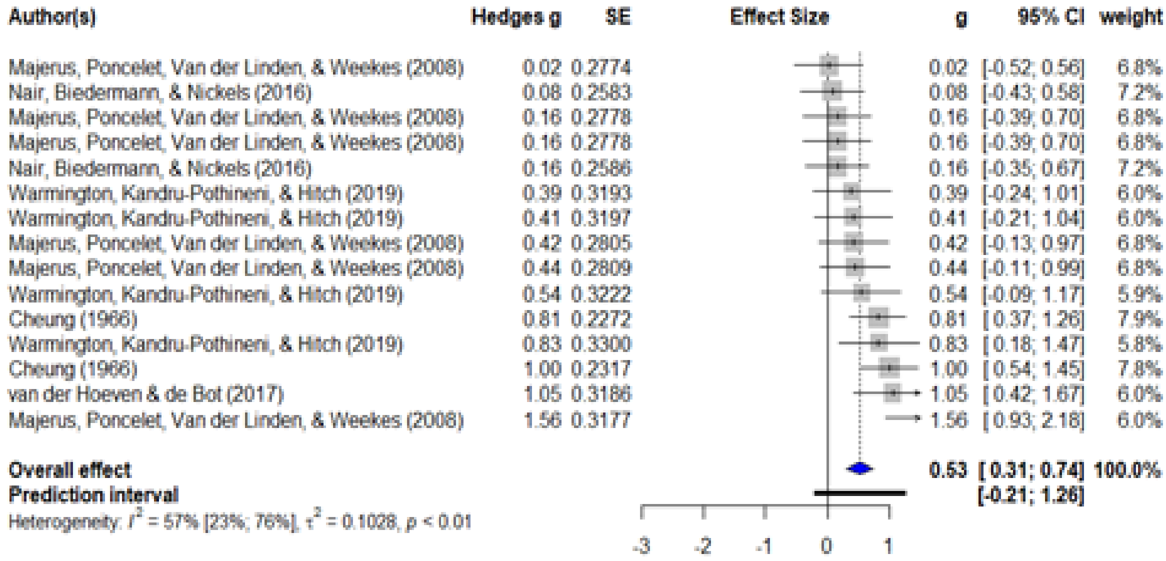


Figure 4. 5 Forest plot with individual and summary effect size estimates for the role of bilinguals' working memory in episodic memory with monolinguals on the left and bilinguals on the right

4.3.2.2.2 Executive Control (i.e., Inhibition/Attentional Control)

Two studies with 3 inhibition/attentional control measures were included in the investigation of the role of executive control in bilingual episodic memory performance (Figure 4.6).

Two measures to assess the between-study heterogeneity were employed: Cochran's *Q* statistic and I^2 test. Cochran's *Q* statistic revealed a nonsignificant difference between the true effect size among the included studies, $Q = 2.17$, $df = 2$, $p = .339$. I^2 test resulted in a very low heterogeneity (7.6%) and thus the fixed effect model was chosen for this meta-analysis. The results revealed a statistically significant large effect size, *g* = .71, 95% CI [.34, 1.09], $p = .001$ suggesting that there is evidence that bilingual inhibition/attentional control plays a role in bilingual episodic memory performance, however, due to the very limited number of studies the evidence is not conclusive. Moreover, due to the very few studies included, Egger's test of funnel plot asymmetry may lack the statistical power to detect bias.

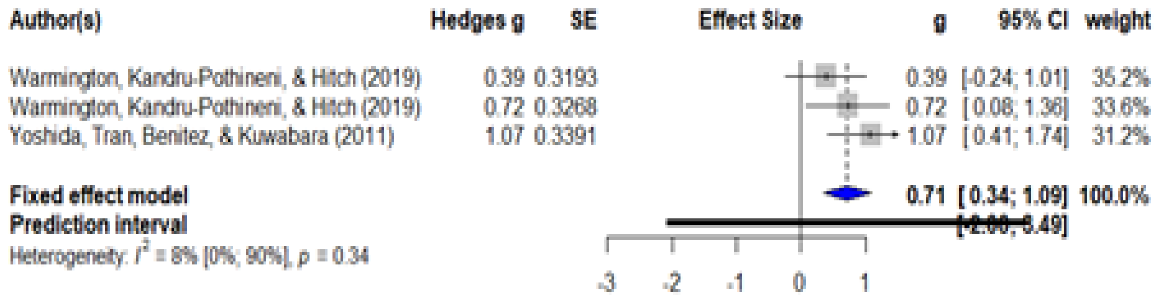


Figure 4. 6 Forest plot with individual and summary effect size estimates for the role of bilinguals' executive control (i.e., inhibition/attentional control) in episodic memory with monolinguals on the left and bilinguals on the right

4.3.2.2.3 Summary

This meta-analysis investigated the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance. A significant medium effect size within six eligible papers that comprised 15 working memory tasks was observed suggesting that working memory plays a significant role in bilingual episodic memory performance as measured by paired associate learning tasks. For executive control meta-analysis, a statistically significant large effect size was observed again indicating that executive control plays a significant role in bilingual episodic memory. The results should however be treated cautiously due to a very limited number of eligible studies included in the analysis.

4.3.2.3 What is the Role of Bilinguals' Vocabulary Knowledge (in English) in Bilingual Episodic Memory Performance as measured by Paired Associate Learning Tasks?

Five studies were included in the investigation of the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory performance (Figure 4.7).

Two measures to assess the between-study heterogeneity were employed: Cochran's Q statistic and I^2 test. Cochran's Q statistic revealed a statistically significant difference between the true effect size among the included studies, $Q = 16.51$, $df = 4$, $p = .002$. I^2 test resulted in a substantial heterogeneity (75.8%) and thus the random effect model was chosen for this meta-analysis. The results revealed a significant large effect size, $g = .76$, 95% CI [.29; 1.22], $p = .001$ suggesting that bilinguals' vocabulary knowledge (in English) plays a significant role in bilingual episodic memory performance.

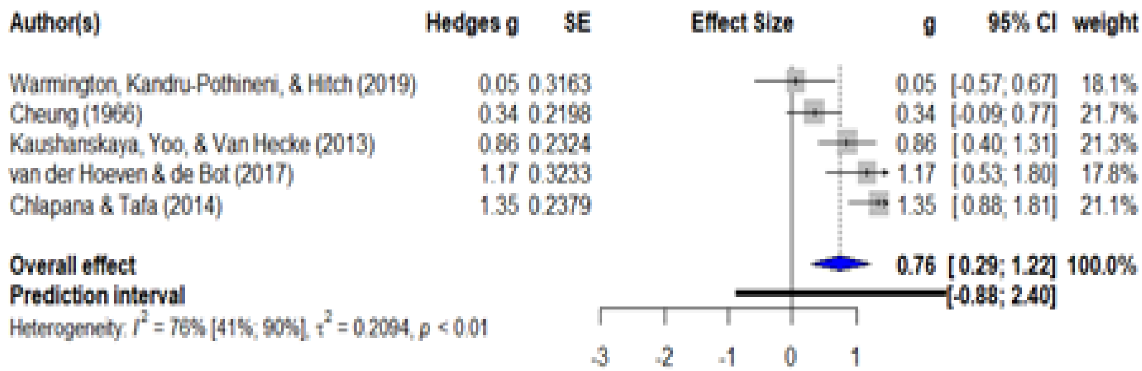


Figure 4. 7 Forest plot with individual and summary effect size estimates for the role of bilinguals' vocabulary knowledge (in English) in their performance on episodic memory tasks, as measured by novel word learning tasks with monolinguals on the left and bilinguals on the right

4.3.2.3.1 Summary

The results revealed a statistically significant large effect size thus indicating that vocabulary plays a significant role in bilingual episodic memory performance. However, only five papers were analysed and the results revealed substantial heterogeneity between the included studies.

4.3.2.4 Is there a Bilingual Advantage in Working Memory and Executive Control (i.e., Inhibition/Attentional Control)?

4.3.2.4.1 Working Memory

Eight studies with twenty-three measures were included in the investigation of the role of bilingual advantage in working memory. Other studies were excluded from this analysis due to homogenous sample comparisons or the lack of information regarding the number or novel words produced by participants during the testing/assessment of learning.

Two measures to assess the between-study heterogeneity were employed: Cochran's Q statistic and I^2 test. Cochran's Q statistic revealed a statistically significant difference between the true effect size among the included studies, Q = 128.68, df = 22, p = .001. I^2 test resulted in a substantial heterogeneity (82.9%) and thus the random effect model was chosen for this meta-analysis. There was a statistically significant effect size observed between the monolingual and bilingual in terms of their working memory performance, g = .36, 95% CI [.06; .66], p = .018 (Figure 4.8).

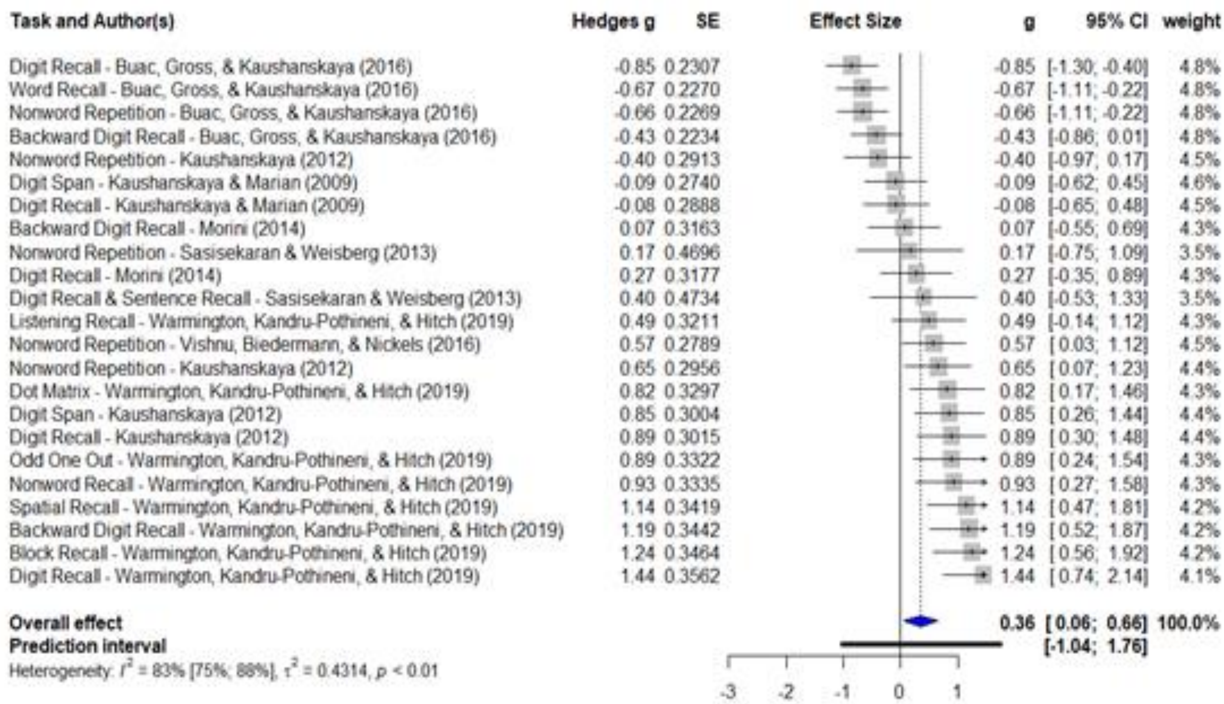


Figure 4. 8 Forest plot with individual and summary effect size estimates for working memory with monolinguals on the left and bilinguals on the right

Further analysis for detecting outliers and influential cases identified one study with extreme effect sizes: Buac, Gross, & Kaushanskaya (2016). After its exclusion, the overall effect size estimate changed, g = .55, 95% CI [.32; .78], p = .001, suggesting a statistically significant difference in working memory between bilinguals and monolinguals. Assessments of between-study heterogeneity also yielded different results: Q-statistic remains statistically significant, Q = 40.40, df = 17, p = .001 indicating a significant difference between the true effect size among the studies. I^2 decreased to 57.9%.

The output provided by Egger's test suggested presence of the publication bias due to the significant funnel asymmetry (Intercept = 9.21, 95% CI [5.48; 12.93], t = 4.79, p < .001). Duval and Tweedie's (2000) trim-and-fill procedure identified and trimmed 9 studies. The overall effect that was estimated by this procedure was g = .12. When compared with the initial pooled effect size of g = .36 and g = .55 after the removal of one outlier study, the initial results were possibly overestimated due to publication bias identified by trim-and-fill procedure, and the true effect size is small with g = .12 when controlling for selective publication.

Additional p-curve analysis revealed that 12 studies had the p-value lower than .025. Power estimate of the analysis equaled 73% with 95% CI [48.6%-88.6%]. The evidential value was present which means that there is indeed a *true* effect size behind the findings for working memory and these are not resulting only from the publication bias.

4.3.2.4.2 Executive Control (i.e., Inhibition/Attentional Control)

Three studies with 7 measures were included in the investigation of the role of bilingual advantage in executive control (Figure 4.9). Other studies were excluded from this analysis due to homogenous sample comparisons or the lack of information regarding the number or novel words produced by participants during the testing/assessment of learning.

Two measures to assess the between-study heterogeneity were employed: Cochran's Q statistic and I^2 test. Cochran's Q statistic revealed a statistically significant difference between the true effect size among the included studies, $Q = 20.50$, $df = 6$, $p = .002$. I^2 test resulted in a substantial heterogeneity (70.7%) and thus the random effect model was chosen for this meta-analysis. There was a marginally statistically nonsignificant effect size observed between the monolingual and bilingual in terms of their executive control performance, $g = .44$, 95%, 95% CI [-.01; .88], $p = .055$. This was further corroborated by the fact that the diamond touched the line of no effect as presented on a forest plot (Figure 4.9).

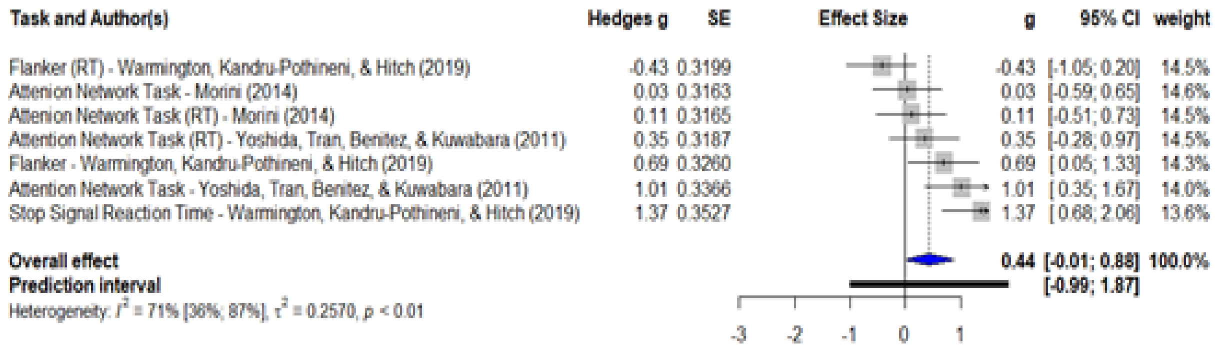


Figure 4. 9 Forest plot with individual and summary effect size estimates for executive control (i.e., inhibition/attentional control) with monolinguals on the left and bilinguals on the right

Further analysis for detecting outliers and influential cases identified no studies with extreme effect sizes.

The output provided by Egger's test suggested presence of the publication bias due to the significant funnel asymmetry (Intercept = 40.35, 95% CI [19.97; 60.74], $t = 3.88$, $p = .017$). No studies were imputed in Duval and Tweedie's (2000) trim-and-fill procedure.

Additional p-curve analysis revealed that 2 studies that were included in the analysis had the p-value lower than .025. Power estimate of the analysis equaled 70% with 95% CI [14%-96.1%]. The evidential value was present which means that there is indeed a *true* effect size behind the findings for executive control (i.e., inhibition/attentional control) and these are not resulting only from the publication bias.

4.3.2.4.3 Summary

Effect sizes were calculated to determine if there was a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control). Only studies which recruited both monolinguals and bilingual samples were included in the analyses. This limited number of publications was due to single samples of monolingual speakers or bilingual speakers as well as research where the results for working memory and executive control (i.e., inhibition/attentional control) were not reported.

Working memory between monolingual and bilingual samples was compared in eight eligible studies. The analysis of working memory revealed moderate heterogeneity and a statistically significant small effect size was observed between the language groups after removing one outlier (g = .12). Similarly, for the between-group comparisons of executive control (i.e., inhibition/attentional control), three studies were included. Again, substantial heterogeneity among the publications was observed and a statistically nonsignificant effect size was noted between the monolingual and bilingual groups. This is indicative of bilingual participants performing better on working memory but not on executive control (i.e., inhibition/attentional control) compared to monolingual speakers.

However, the results of these meta-analyses should be treated with caution due to the very limited number of eligible studies and high heterogeneity between studies included in the analysis.

4.3.2.5 Moderator Analysis

The aim of this meta-analysis was also to examine whether the effect sizes were influenced by methodological features of the research. The following moderators were of interest: socioeconomic status and age. For bilinguals, the age of language acquisition, language proficiency, language usage, and language environment (monolingual/bilingual dominant learning context) were also of interest. The moderator analyses could not be conducted reliably due to either the limited information included in the studies or relevant information not being reported at all. Also, the age as a moderator was not also investigated due to the numerous experimental groups where measures of central tendency could not have been calculated.

4.4 Discussion

The aim of this meta-analytic investigation was to provide cumulative evidence and inform the current state of knowledge regarding the bilingual advantage in episodic memory as well as in working memory and executive control (i.e., inhibition/attentional control). Thus, the following research questions were examined:

1) Is there a bilingual advantage in episodic memory as measured by paired associate learning tasks?
2) Is the bilingual advantage in episodic memory task-dependent (i.e., how paired associate learning is assessed: recall versus recognition measures)?
3) What is the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance, as measured by paired associate learning tasks?
4) What is the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory performance, as measured by paired associate learning tasks?
5) Is there a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control)?

Research Questions 1 and 2 were answered by investigation of recall and recognition in paired associate learning tasks. The analysis of recall revealed a statistically significant large effect size between the language groups. Similarly, for recognition, a statistically significant small effect size was noted between the monolinguals and bilinguals. These results suggest a bilingual advantage in episodic memory as measured by paired associate learning tasks and indicate that it is task-dependent. This is in line with previous research (i.e., Bialystok et al., 2003; Kaushanskaya & Marian, 2009).

Research Question 3 was examined by looking at the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance. A significant medium effect size was observed for working memory and a significant large result was observed in terms of executive control (i.e., inhibition/attentional control). This suggests that both working memory and executive control (i.e., inhibition/attentional control) play a significant role in bilingual episodic memory as measured by paired associate learning tasks. This is again in line with previous research findings (Papagno and Vallar, 1995; Bialystok et al., 2004; Colzato et al., 2008; Costa et al., 2008).

Research Question 4 concerned the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory. The results revealed a statistically significant large effect size thus indicating that vocabulary plays a significant role in bilingual episodic memory performance.

Research Question 5 investigated the bilingual advantage in working memory and executive control (i.e., inhibition/attentional control) between monolingual and bilingual groups. The analysis of working memory revealed a statistically significant small effect size. For the between-group comparisons of executive control (i.e., inhibition/attentional control), a statistically nonsignificant effect size was noted. This suggests that there is a bilingual advantage in working memory but not executive control (i.e., inhibition/attentional control).

Although statistically significant effect sizes were found in this meta-analysis, the high between-study heterogeneity impacts some of them and as such no conclusive findings regarding understanding of how bilingualism modifies long-term memory mechanisms can be drawn. A series of additional analyses was performed to minimise the risk of significant heterogeneity and bias, and these were discussed in more detail. In line with the Cochrane Handbook, in order to obtain reliable results from a random-effects meta-analysis, at least ten studies should be included for it not to be underpowered (Sterne, Egger, & Moher, 2011). Indeed, most analyses run in this systematic review were based on the fewer number of studies and thus may have been underpowered.

The results obtained from the analyses conducted in this systematic review are also aligned with the strength of evidence table presented as a part of our quality assessment process (Table 4.4). Low scores given for the robustness of sample size meant that sample size was not justified and sample characteristics were often not reported. There was also a lack of clarity in terms of the characteristics of the bilingual sample which was further reflected in the analyses by the high heterogeneity between the studies. Taken together, these aspects indicate that some meta-analytic findings were in fact inconclusive.

Apart from the meta-analyses conducted, the review aimed to investigate whether the effect sizes were influenced by methodological features of the research. Moderator analyses could not be conducted reliably due to either the limited information included in the studies or relevant information not being reported at all. Also, the age as a moderator was not also investigated due to the numerous experimental groups where measures of central tendency could not have been calculated.

As this review is limited by substantial heterogeneity observed across studies, the results from conducted meta-analyses should be treated with caution and may be found difficult to interpret. Egger's test computed for each forest plot separately implied the presence of a significant publication bias in almost all analyses conducted.

One of the possible explanations to the present results might be methodological features of the studies included in the analyses. As already mentioned, the results obtained are in line with the low scores that studies were given in the strength of evidence assessment. Low scores given for the robustness of sample size as well as medium evidence for the robustness of the characteristics of the bilingual samples were in fact reflected in the effect sizes obtained. Conducting power-analyses to assess the required sample prior to recruitment would ensure the sample size is justified and leads to reliable and trustworthy results.

Not only are further experiments needed, there also needs to be more consistency in terms of measures being used in such studies to enable between-group comparisons in terms of their episodic memory and executive control performance. In fact, the implementation of broader executive control measures and more rigid experimental setups would enable the exploration of the wider spectrum of bilingual and monolingual abilities more specifically.

One of the crucial recommendations that stem from this systematic review is the fact that more research is needed to enable the exploration of the role of bilinguals' executive control in episodic memory performance (as measured by paired associate learning tasks) and further investigate the role of working memory, inhibition/attentional control, and linguistic indices in this process.

CHAPTER FIVE

Thesis Aims and Overview

5.1 Rationale

As discussed in Chapter One, executive control consists of a broad set of higher order cognitive skills which show some unity and diversity. In other words, these skills share some commonality, but they are also unrelated to each other at the same time. Although executive control and working memory are separable processes, the latter is implicated in executive control processes (Engel, Kane, & Tuholski, 1999). Chapter Three presented research examining bilingual advantage in working memory and executive control and emphasised inconsistencies in results (for a review of bilingual advantage see Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012: however, see De Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2015). As such, there is a lack of consensus whether there is a bilingual advantage in executive control and working memory.

Executive control processes are not only linked to short-term memory, but they are also implicated in long-term memory mechanisms particularly from the point of view of how they interact with episodic memory processes (i.e., the ability to learn associations, integrate them into long-term memory, and retrieve them from long-term memory over time). Research to date has focused on bilingual influence on executive control and working memory mechanisms, however there is still very limited understanding of how bilingualism impacts long-term memory mechanisms. This evidence is further corroborated by the results of the systematic review of the existing literature which revealed a very limited number of studies that investigate episodic memory in bilinguals as measured by paired associate learning tasks (i.e., a paradigm that involves the learning of novel names for novel objects, see Chapter Two). Results of the meta-analyses (see Chapter Four) revealed a bilingual advantage in episodic memory in terms of recall, recognition, and working memory, but not executive control, thus indicating that bilingual advantage is indeed task-dependent. Also, working memory, executive control (i.e., inhibition/attentional control), and vocabulary knowledge (in English) were found to play a significant role in bilingual episodic memory performance as measured by paired associate learning tasks. However, due to the high levels of between-study heterogeneity, the findings of meta-analyses were inconclusive in terms of answering the research questions stated in Chapter Four.

Long-term memory can be measured not only by paired associate learning but also by verbal fluency (see Chapter Two) where participants are asked to retrieve as many words as possible from episodic memory within one minute. As verbal fluency incorporates long-term memory access and executive control, it provides more insight into the mechanisms that underpin bilingual word recall and allows investigation into whether bilingualism impacts episodic memory.

Thus, the aim of the thesis was to examine whether there is a bilingual advantage in working memory, executive control (i.e., inhibition/attentional control), and episodic memory as measured by the paired associate learning task and verbal fluency. Additionally, the thesis aimed to examine the role of working memory, executive control (i.e., inhibition/attentional control), and English vocabulary knowledge in bilingual episodic memory performance.

This thesis addressed the following research questions:

1) Is there a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control)?
2) Is there a bilingual advantage in episodic memory as measured by the paired associate learning task?
3) Is the bilingual advantage in episodic memory task-dependent (i.e., how the paired associate learning is assessed: recall versus recognition measures)?
4) What is the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance as measured by the paired associate learning task?
5) Is there a bilingual advantage in episodic memory as measured by verbal fluency?
6) What is the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory performance?

The following hypotheses are thus stated:

Research hypothesis 1: There is a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control).

Research hypothesis 2: There is a bilingual advantage in episodic memory as measured by the paired associate learning task.

Research hypothesis 3: The bilingual advantage in episodic memory is task-dependent (i.e., recall versus recognition).

Research hypothesis 4: Working memory and executive control (i.e., inhibition/attentional control) play significant role in bilingual episodic memory performance as measured by the paired associate learning task.

Research hypothesis 5: There is a bilingual advantage in episodic memory as measured by verbal fluency.

Research hypothesis 6: Bilinguals' vocabulary knowledge (in English) plays a significant role in bilingual episodic memory performance.

In order to minimise variability associated with cultural diversity, this study investigated a single bilingual population of Polish-English bilingual adults. Indeed, Polish language is important in its own right, as it is the main language other than English that is widely spoken in England (spoken by more than 500,000 people). Moreover, it is an under-represented language in bilingualism and cognition literature and thus research with Polish-English bilinguals is of great importance.

5.2 Overview of Experimental Chapters

Taking the research questions into account, Chapter Six will outline the methodological design of the study. To assess and answer Research Question 2, a reliable paired associate learning task had to be developed. Thus, Chapter Six will present five phases of stimuli development and piloting of the paired associate learning task. The task is a fundamental measure employed in Experiment 2 discussed in Chapter Eight that investigates one’s ability to learn novel paired associates and thus the ability to access and retrieve information from episodic memory.

Chapter Seven is the first experimental chapter and will address Research Question 1 concerning a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control). In this Chapter, the differences in monolingual and bilingual accuracy and error performance on working memory and executive control measures will be presented.

Chapter Eight is the second experimental chapter and will address Research Questions 2, 3, and 4. Thus, this Chapter will investigate a bilingual advantage in episodic memory as measured by the paired associate learning task and examine whether a bilingual advantage in episodic memory is task-dependent (i.e., as assessed by recall and recognition). Additionally, of interest to this Chapter is investigating the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance.

Chapter Nine is the final experimental chapter and will address Research Questions 5 and 6 concerning a bilingual advantage in episodic memory as measured by verbal fluency. It will also examine the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory performance.

The thesis is concluded by Chapter Ten which will present the key findings of this study and discuss the strengths and limitations of the experiments conducted. This Chapter will close with a discussion of theoretical and practical implications of the study and will propose future directions in the field of bilingual advantage in episodic memory research.

5.3 Ethical Considerations

The study obtained ethical approval from the Ethics Panel administered by the Department of Human Communication Sciences at the University of Sheffield (see Appendix B). Information sheets for interested parties included details regarding legal basis for data protection and processing in compliance with General Data Processing Regulation (GDPR) from 25 May 2018. Participants were informed that collected data would be anonymised and therefore each participant would be assigned a random ID number to ensure confidentiality. They were also informed about the voluntary character of participation and their right to withdrawal from the study at any point without any consequences.

The same set of participants took part in all the experiments reported in the three experimental chapters (i.e., Chapter Seven, Eight, and Nine). However, different participants were recruited for the development of the paired associate learning task as reported in Chapter Six. Participants did not receive any credits or monetary compensation for taking part in the study. They could however opt in to participate in a £50 Amazon voucher draw at the end of the study.

5.4 Covid Statement

The original study aimed to recruit two larger groups of participants. An a priori power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2007) to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 50% power for detecting a medium effect, at a significance criterion of $\alpha = .05$, was $N = 51$ for group 1 (bilinguals) and $N = 51$ for group 2 (monolinguals) to test the study hypotheses."

Unfortunately, due to the COVID-19 pandemic, consequent closure of the campus, and national and international lockdowns, it was impossible to conduct additional face-to-face assessments. Additionally, the pandemic also coincided with my pregnancy and maternity leave which had an impact on my ability to recruit and test more adult participants. I also collected data from children in Poland and England but the data did not allow for any meaningful between-group comparisons.

CHAPTER SIX

Methodological Considerations - the Development and Piloting of the Paired Associate Learning Task

6.1. Introduction

As outlined in Chapter Two, paired associate learning is a paradigm that involves the learning of novel pairings of two unrelated stimuli, therefore, creating novel associations. In the context of this thesis, a novel object is paired with a novel word and participants are asked to recall these associations on three different days. The task aims to assess how well participants have learned these paired associates based on their ability to retrieve the name for a novel object from their long-term memory. As such, paired associate learning is a measure of the long-term memory mechanisms, episodic memory more specifically, as it involves the ability to access and recall the novel associations over time.

One consistent finding emerging from studies with novel word learning, is that new words seem to be processed differently after time. Therefore, it has been argued that novel lexical items undergo consolidation and hence the role of sleep in this process has been further investigated (Gaskell & Dumay, 2007). Indeed, studies with sleep and no-sleep condition after a training on novel lexical items have revealed that the former one contributes to the engagement of newly acquired items in lexical competition with already existing lexicon whereas this has not been observed for items in no-sleep condition. The crucial role of sleep in memory consolidation have been evidenced in a number of studies (Brawn, Fenn, Nusbaum, & Margoliash, 2008; Cross, Kohler, Schlesewsky, Gaskell, & Bornkessel-Schlesewsky, 2018) and specifically in the field of language learning (Gomez, Bootzin, & Nadel, 2006; St. Clair & Monaghan, 2008; Ellenbogen, Hu, Payne, Titione, & Walker, 2007).

The evidence accumulated by research to date, suggests that the extent of a novel word's impact on lexical processing is time dependent and some aspects of lexical behaviour exhibit themselves sooner while others are restricted by overnight or over week consolidation. As explained by Gaskell and Dumay (2003), the acquisition of a new phonological representation, along with establishing a new lexical representation for the new word and further integration of the two into long-term memory requires time. Given that of interest to this project is looking at episodic memory, in other words the integration of the target items fully into long-term memory mechanisms, testing participants at different timepoints was important. By testing a day later (Day 2), overnight consolidation and short-term memory mechanisms could be investigated whereas Day 7 enabled testing of long-term memory mechanisms. As already mentioned, consolidation effects seem not to arise until a week later (Gaskell & Dumay, 2007) and thus it is reasonable to ascertain whether or not they are to be found a week later in this study.

The purpose of this Chapter is to detail the development and piloting of the Paired Associate Learning task (i.e., the learning of novel names for novel objects) that is reported in Experiment 2 (see Chapter Eight). The development of the task is divided into three phases.

The aim of Phase 1 is to identify the novel word stimuli for the task and assess their familiarity in Polish. Phase 2 aims to identify novel objects and investigate their similarity. In Phase 3, the main task is to assess the integrity of novel word-object pair associates. The Chapter is concluded by Phase 4 and 5, which test the overall efficacy of the paired associate learning task.

6.2. Phase 1

The aim of Phase 1 was to identify the novel words for the task. The novel word stimuli were selected so that they were not familiar (or known) in either Polish or English. The aim of the task was to establish novel word-object pairs that had never been encountered before. To reduce the likelihood of participants' utilising their existing lexical and semantic knowledge, the decision was made to select novel words that did not have a verbal or visual label. If the task employed a 'neutral' or 'artificial' language created for the purpose of the experiment, the study would have been investigating second language learning in monolinguals and third language learning in bilingual participants.

Indeed, novel English-based words have also been used in previous studies on language learning (Warmington, Kandru-Pothineni, & Hitch, 2018; Kaushanskaya & Marian, 2009). In this study, the novel words were the words that conformed to the phonotactic structure of real English words (i.e. were matched on syllable, phoneme, stress pattern, bigram count and the target sound presence). Phonotactic similarity of novel words to English words enabled the examination of the real process of word learning (as a second language for bilingual participants who spoke English [as a second language] as well as monolinguals for whom the novel words conformed to their first language rules). Novel words taken from Bretherton-Furness, Ward, and Saddy (2016) were selected. Each novel word was recorded by a native English speaker. Polish speakers assessed the familiarity of the spoken novel words to the Polish language. As the novel words were based on English real words, it was not necessary to test their familiarity to English by English speaking participants. The novel words which were the least familiar to Polish words were selected.

6.2.1 Participants

Ten native Polish speakers living in Poland (6 females, 4 males) with a *mean* age of 26 years (*SD* = 5.72, *range* = 19 - 40 years) with no reported hearing problems, developmental or acquired disorders took part in Phase 1. Participants were recruited via Facebook advertisement and followed by snowball sampling.

6.2.2 Materials and Design

A list of 226 novel words taken from Bretherton-Furness et al. (2016) were selected for this phase. Bretherton-Furness et al. developed their stimuli based on real nameable English words and as such were matched on syllable, phoneme, stress pattern, bigram count and the target sound presence (for example, the real word /'dɒŋ.kiː/ derived the novel word /'mɒn.veɪ/). For Phase 1 each novel word was recorded by a native English speaker.

6.2.3 Procedure

Participants rated the familiarity of each novel word to Polish on a 5-point scale (where 1 was *not familiar at all* and 5 was *very familiar*).

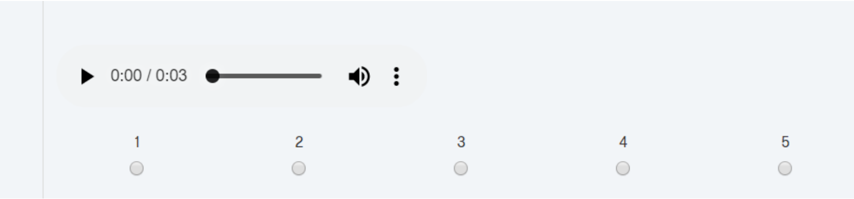


Figure 6. 1 The display of the novel word in the online questionnaire in Phase 1

Each novel word was presented one at a time to participants and participants were instructed to assess the familiarity of novel words to real Polish words via a button choice (see Figure 6.1). There was no time limit set on completing the task and participants could listen to each novel word as many times as they wanted before providing a rating. Participants completed this task online via the Qualtrics platform (https://www.qualtrics.com/uk/, 2017). They could complete the task using a laptop/PC/tablet or phone. The task lasted approximately 10 minutes.

6.2.4. Results and Summary

An item-analysis was conducted to identify novel words that were suitable for the paired associate learning task. Means, medians, and standard deviations for familiarity rating were calculated for the 226 items with a *mean* rating of 2.01 (*SD* = .38) and *median* rating of 1.73 (*SD* = .98). Novel words that had a *mean* rating higher than 1.5 were excluded, and on this basis 40 novel items were selected (overall *mean* rating of 1.23, *SD* = .25; *median* rating = 1.25). These items were then divided into four sets of 10 items (see Table 6.1). The purpose of dividing the 40 items across four sets was to ensure that during the paired associate learning task novel stimuli were counterbalanced across participants. A one-way Analysis of Variance (ANOVA) (with item set as the variable of interest) revealed that there were no statistically significant differences in the familiarity ratings between item sets, $F(2, 39) = 1.71$, $p = .959$.

Following Phase 1, 40 novel words were identified for the paired associate learning task.

Table 6. 1 Mean and median (and standard deviation) familiarity score for novel words

Set 1	*Mean*	*Median*	*SD*	Set 2	*Mean*	*Median*	*SD*
'enlə.diːv	1	1	0	'ɒ.trɪpt	1	1	0
'kɒtə.mægən	1	1	0	'wæn.daʊ	1	1	0
'kaɪnæ.fəl	1.1	1	.32	'feəp	1.1	1	.32
'wɪl.mɪkt	1.1	1	.32	'skwɪ.rɪt	1.2	1	.42
'keə.hɒs	1.2	1	.42	'klaʊp	1.2	1	.42
'rɜːl	1.2	1	.63	'ə.fɑːdiən	1.3	1	.95
'brɪ.pə	1.3	1	.67	'streɪ.betʃi	1.3	1	.48
'tʃrɪb.draɪvə	1.3	1	.95	'bræb	1.3	1	.48
'stɒd	1.4	1	.97	'mɒn.veɪ	1.4	1	.70
'fluː.əl	1.4	1	.70	'krəʊt	1.4	1	.70
Mean	1.2	1		Mean	1.22	1	
SD	.15	0		*SD*	.15	0	

Set 3	*Mean*	*Median*	*SD*	Set 4	*Mean*	*Median*	*SD*
'kraɪ.pɒkəbɑː	1	1	0	'lemə.feɪn	1	1	0
'emə.fens	1.1	1	.32	'keng.suːn	1.1	1	.32
'kæŋ.breʃ	1.1	1	.32	'blem.pɪt	1.1	1	.32
'fæn.və	1.2	1	.42	'kla.ɪəs	1.2	1	.63
'felə.suːsən	1.2	1	.63	'trɔː	1.2	1	.63
'aɪrəʊ.treɪt	1.3	1	.48	'kætə.bɜːgə	1.3	1	.67
'keɪnt.grʌʃ	1.3	1	.67	'pɪ.keɪtə	1.3	1	.95
'preɪn	1.3	1	.67	'ʌs.frɒlə	1.3	1	.48
'treɪdʒ	1.4	1	.84	'hemɪ.teltə	1.4	1	.70
'bæs.kəl	1.4	1	.52	'drəʊks	1.4	1	.70
Mean	1.23	1		Mean	1.23	1	
SD	.13	0		*SD*	.13	0	

Note: using the scale: 1 = novel word is not familiar to Polish at all, 5 = novel word is very familiar to Polish

6.3 Phase 2

Phase 2 focused on selection of the novel objects for the paired associate learning task. As such, the novel object stimuli were taken from Horst and Hout (2015). Additionally, each novel object was paired with another novel object that would share some common features (i.e., colour, shape, and size) but be equally distinguishable at the same time. The aim of such a manipulation was to create a target novel object as well as a foil novel object that would be equally probable to be assigned a novel label by participants when they were completing the paired associate learning task.

Therefore, in Phase 2, a set of existing novel objects was used with some adaptations of a few items and a new group of participants recruited to assess the similarity between the chosen novel object pairings by answering two questions: (1) how similar these two objects are based on your first impression and (2) how similar these objects are based on colour/shape/size separately. The two rating scales enabled the elimination of novel object pairings that were rated as too similar and not similar at all.

6.3.1 Participants

A new set of participants was recruited for Phase 2. Fifteen Polish English bilinguals (11 females, 4 males) with a *mean* age of 27 years (*SD* = 7.23) with no reported hearing problems, developmental or acquired disorders took part in Phase 2. They were recruited by means of social media and snowball sampling.

6.3.2 Materials and Design

A total of 80 novel objects (64 taken from Horst and Hout (2015) and additional 16 novel objects developed for the purpose of the study, see Appendix C) were paired together to create 44 novel object pairings. The aim of this was to match the objects based on their colour, shape, and size.

6.3.3 Procedure

Participants rated the similarity of each novel object pairing on a 5-point scale (where 1 was *not similar at all* and 5 was *very similar*) based on the following criteria: (1) ***overall first impression*** of the similarity between the items in each pair, and (2) ***feature*** (i.e., colour, shape, and size) similarity between the items in each pair (see Figure 6.2). Participants completed this task online via the Qualtrics platform (https://www.qualtrics.com/uk/, 2017). They could complete the task using a laptop/PC/tablet or phone. The task lasted approximately 10 minutes.

Figure 6. 2 An example of novel object pairing and experimental setup in Phase 2

6.3.4 Results and Summary

Data were scored as the mean ratings for overall first impressions and features (i.e., averaged across *c*olour, shape, and size) for each pair and are presented in Table 6.2.

Table 6. 2 Mean (and standard deviation) overall first impression and feature score

Object Pair	Feature *mean*	Feature *SD*	Overall First Impression *mean*	Overall First Impression *SD*	*Mean* Difference
1	2.80	.63	2.80	1.01	.00
2	2.82	.65	2.80	.94	.02
3	4.64	.53	4.67	.62	.02
4	3.62	.50	3.67	.82	.04
5	3.89	.50	3.93	.46	.04
6	2.40	.69	2.33	1.29	.07
7	3.40	.55	3.47	1.13	.07
8	2.62	.92	2.53	1.19	.09
9	3.96	.80	4.07	.70	.11
10	2.58	.96	2.47	1.36	.11
11	2.58	.70	2.47	1.25	.11
12	2.71	.80	2.60	1.30	.11
13	3.11	.73	3.00	1.13	.11
14	4.84	.25	4.73	.59	.11
15	2.53	.61	2.67	1.11	.13
16	2.73	.80	2.87	1.06	.13
17	2.20	.96	2.07	.96	.13
18	3.82	.73	3.67	1.11	.16
19	2.33	.70	2.53	.83	.20
20	3.82	.79	3.60	1.12	.22
21	2.53	.85	2.27	1.10	.27
22	4.20	.84	4.47	.83	.27
23	2.29	.56	2.00	1.00	.29
24	4.16	.76	4.47	.74	.31
25	2.38	.78	2.00	1.07	.38

26	2.58	.90	2.20	1.08	.38
27	2.73	.59	3.13	.74	.40
28	2.38	.74	2.80	1.15	.42
29	3.36	.82	2.93	1.03	.42
30	3.84	.93	3.33	1.05	.51
31	3.07	.81	3.60	1.12	.53
32	3.47	.75	4.00	.65	.53
33	3.00	.85	2.47	1.25	.53
34	2.29	.72	1.73	.80	.56
35	3.44	.78	4.00	.85	.56
36	2.42	.96	3.00	1.31	.56
37	3.76	.90	4.33	.82	.58
38	2.20	1.01	1.60	1.06	.60
39	2.64	.75	2.00	1.07	.64
40	2.98	1.02	2.20	1.26	.78
41	**3.16**	**.74**	**4.20**	**1.01**	**1.04**
42	**3.89**	**.39**	**4.93**	**.26**	**1.04**
43	**2.93**	**.90**	**1.80**	**.86**	**1.13**
44	**3.58**	**.56**	**4.73**	**1.03**	**1.16**

Note: Mean feature (i.e., averaged across colour, shape, and size). Pairings in bold (41, 42, 43, and 44) were eliminated based on high mean differences (> 1.00)

Item pairs that demonstrated discrepancies between their *overall first impression* and *feature* similarity rating scores (i.e., mean differences greater than 1) were excluded. The mean difference above 1.00 indicated that there was a difference greater than 1 score between *overall first impression* and *feature* similarity rating scales (see Table 6.2). As such, four item pairs were excluded. For the remaining 40 object pairs, there were no statistically significant differences between their *overall first impression* (*mean* = 3.30, *SD* = .91) and their *features* (*mean* = 3.19, *SD* = .67), $t(39) = 1.50$, $p = .140$, $\eta_p^2 = .406$.

6.4 Phase 3

Phase 3 examined the plausibility of novel word-novel object pairings, that is, the quality of whether proposed pairings seem probable. Forty novel words were randomly matched with 40 novel object pairs from Phase 2.

6.4.1 Participants

A new group of participants was recruited for Phase 3. Six Polish participants (4 females, 2 males) with a *mean* age of 32 years (*SD* = 3.76) with no reported hearing problems, developmental or acquired disorders took part in Phase 3. Participants were recruited by means of snowball sampling and via social media advertising.

6.4.2 Materials and Design

Forty novel words selected in Phase 1 were randomly paired with 40 novel object pairings selected in Phase 2 (see Appendix D).

6.4.3 Procedure

Novel words and novel objects in pairs were uploaded on Qualtrics. Participants could complete the task using laptop/PC/tablet or phone. The task lasted approximately 10 minutes. Participants were asked to choose an object from the pair of novel objects that they thought matched the novel word they heard. They made their choice by selecting the correct button (either item A on the left of the screen or item B on the right) which were situated below the novel objects respectively (Figure 6.3). Their responses were recorded by Qualtrics.

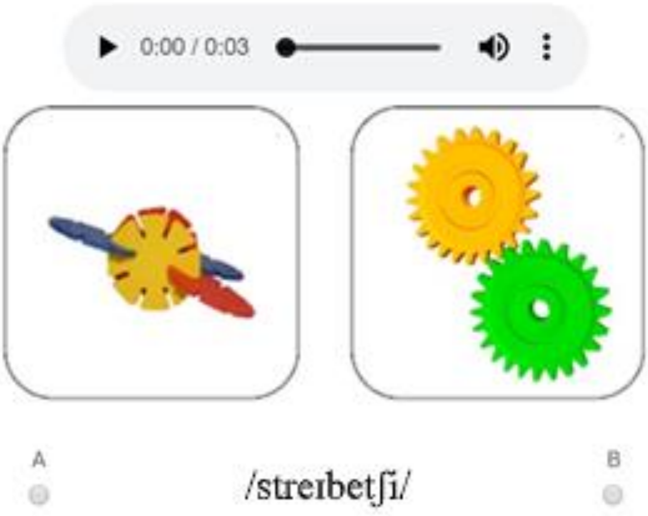


Figure 6. 3 An example of novel object pairing and a novel word (Item A and Item B pairing) in Phase 3

6.4.4 Results and Summary

Novel word-object pairings in which all six participants selected the same response were eliminated and reconfigured into a novel word-object pairing. For example, in Figure 6.3 if all participants selected Item A as their response, this overall pair was eliminated. This resulted in the removal of five pairings from the set of items and the creation of new word-object pairing arrangements.

In the final set of 40 pairings, there were no statistically significant differences in terms of whether or not participants selected Item A or Item B for their response (Item A: *mean* = 54.17, *SD* = 15.45 vs Item B: *mean* = 45.83, *SD* = 15.45), $t(39) = 1.71$, $p = .096$, $\eta_p^2 = .521$).

6.4.5 Interim Summary

Phase 1, Phase 2, and Phase 3 were carried out in order to select the novel words, the novel objects, and to test the integrity of novel word-object pairings, respectively. The outcome of Phase 1 was selection of 40 novel words. The outcome of Phase 2 was 40 novel object pairings where one object acted as a target and the other one as a foil. Phase 3 combined the novel words with the novel object pairings to test the plausibility of one of the items matching onto the word. This resulted in a final set of items which can be found in Appendix E. The aim of the next phase, Phase 4, was to test the efficacy of novel word-object pairings within the paired associate learning task.

6.5 Phase 4

The previous section outlined the development of the stimuli for the paired associate learning task. Having decided upon the most appropriate verbal and visual stimuli, the next step involved piloting the task with a new group of participants to examine its efficacy as a paired associate learning task.

6.5.1 Participants

Five Polish-English bilinguals (3 females and 2 males with *mean* chronological age 28 years and 3 months, *SD* = 5.03) who were residing in the UK participated in Phase 4. They had no known hearing or language impairments. Participants were asked to complete a Language Background Questionnaire (Appendix G) in which they rated their Polish and English proficiency on a 7-point rating scale (where 1 means *very poor* and 7 means *native-like*) as well as their usage of Polish and English in the home/work/school environment on a 1-5 rating scale (where 1 means *never* and 5 means *all the time*).

Table 6. 3 Mean (and standard deviation) self-reported language proficiency and language usage in Polish-English bilinguals

	Mean	*SD*
Polish (first language) proficiency	7.00	0
English (second language) proficiency	6.00	0
Polish usage at home	4.80	.45
Polish usage with friends	4.60	.89
Polish usage at work/study	3.40	1.82
Mean Polish usage	4.27	1.28
English usage at home	3.20	1.48
English usage with friends	4.60	.55
English usage at work/study	4.80	.45
Mean English usage	4.20	1.15
Age of exposure to English at home (in years)	n/a	n/a
Age of exposure to English at school (in years)	7.40	2.88

On average, participants started to learn English (i.e., in either formal educational settings, private tutoring or reading English books) during their early childhood, with one participant reported starting to learn at age 3, two participants at the age of 7 and other two participants at the age of 10.

On average, participants reported using both languages, Polish and English, with the same level of frequency; $t(8) = .091$, $p = .929$, $\eta_p^2 = .336$ (see Table 6.3). They reported using Polish only slightly more often than English. However, they reported that they used Polish more often at home than English, but English more often at work or at the university.

6.5.2 Materials and Design

The 40 items were randomly divided into two sets of items which enabled for counterbalancing across the participants (see Appendix F). Participants were randomly allocated to learning either set A or set B.

Table 6. 4 An example of list A in set A

Negative mapping:	Positive mapping:	Negative mapping:	Positive mapping:
enlədiːv	wɪlmɪkt	ɒtrɪpt	skwɪrɪt
kɒtəmægən	keəhɒs	wændaʊ	klaʊp
kaɪnæfəl	rɜːl	feəp	əfɑːdiən
stɒd	brɪpə	mɒnveɪ	streɪbetʃi
fluːəl	tʃrɪbdraɪvə	krəʊt	bræb

Note: List B comprised the same novel items but with the reverse mapping.

The paired associate learning task consisted of three blocks: Familiarisation, Training, and Test block.

In the Familiarisation block, participants heard a novel word presented over headphones and were required to repeat it. The purpose of this was to ensure that participants are able to pronounce the novel words prior to learning the novel word – novel object pairing. Participants were presented with each novel word only once.

In the Training block, participants were presented with a novel word over headphones while they were simultaneously presented with two novel objects on the screen. They were then required to select which novel object matched the novel word. They were provided with corrective feedback which was visually presented in a form of the words CORRECT or WRONG.

The Training block consisted of four training trials in which 20 object pairings were displayed (20 target objects and 20 foils). For each, item pairs were randomised, and the location of a target item was counterbalanced on the screen. Participants saw each target object once in each sub-block, four times in the entire block altogether. Each target item was displayed twice on the left-hand side of a computer screen and twice on the right in the training phase. To ensure that the items were counterbalanced, sets were further divided into two lists (see Appendix F) where pairings were either positively or negatively mapped, hence half received positive feedback, and the other half negative feedback (see Table 6.4).

In the first part of the Test block, recognition, participants were instructed to match novel words that they had heard in previous blocks to novel objects they had learned in Training. There were two novel objects presented on the screen: the target and the foil. As this was a testing block, no feedback was given. Participants heard each novel word once and were asked to respond by pressing the correct button, either left or right, as quickly as possible. Accuracy and reaction times were recorded but not displayed on the screen.

In the second part of the Test block, recall, participants received no feedback. They were shown the 20 novel objects one at a time that they had encountered in Training and asked to explicitly name each target object. Accuracy and error types were recorded manually by the researcher.

6.5.3 Procedure

The experiment was run across three separate days: Day 1, Day 2, and Day 7. Day 1 and Day 2 were separated by one day, with Day 7 occurring one week later.

During Day 1, participants completed the Familiarisation and the Training block followed by the Test block. Similarly, on Day 2 and Day 7, participants completed the Test block (see Figure 6.4). The assessment across multiple days was conducted to evaluate participants' long-term memory retention of the learned material. The experiment was employed using DMDX software (Forster & Forster, 2003) and was administered between January and August 2018.

Participants were tested individually in a quiet room.

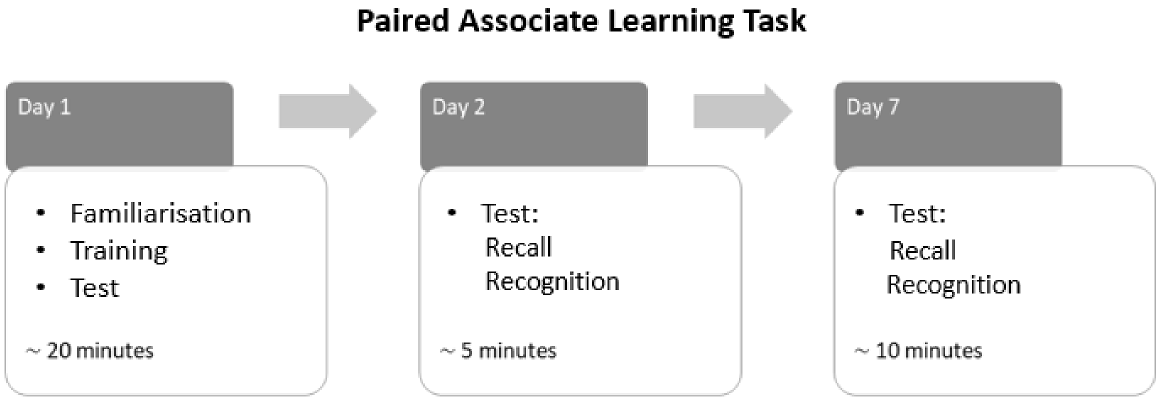


Figure 6. 4 Initial paired associate learning task design

6.5.4 Results

6.5.4.1 Familiarisation

On average, in this block, participants were able to repeat the novel words accurately 70 % of the time (*SD* = .84). To further examine participants' performance on this task, an error analysis was conducted to understand the differences in responses that participants were making. The errors were classified as follows: *omissions* (i.e., participants did not give a response when the target item was presented), *real word intrusions* (RWI) (i.e., participants gave a real word response instead of the target novel word), and *phonological errors* (i.e., participants made an error that deviated from the phonological structure of the novel word).

Table 6.5 presents the pattern of responses during Familiarisation. The majority of errors that the participants were making were phonological in nature with only a small proportion of those being real word intrusion errors. A paired-samples *t*-test revealed a statistically significant difference between error types with significantly more phonological errors being made than RWI, $t(4) = -3.28$, $p = .030$, $\eta_p^2 = .435$.

Table 6. 5 Response patterns in Familiarisation

Correct	RWI	Phonological
.700	.030	.270

6.5.4.2 Training

A one-way ANOVA showed that participants' performance improved across four training trials, $F(3,16) = 4.52$, $p = .024$, $\eta_p^2 = .665$ (see Figure 6.5).

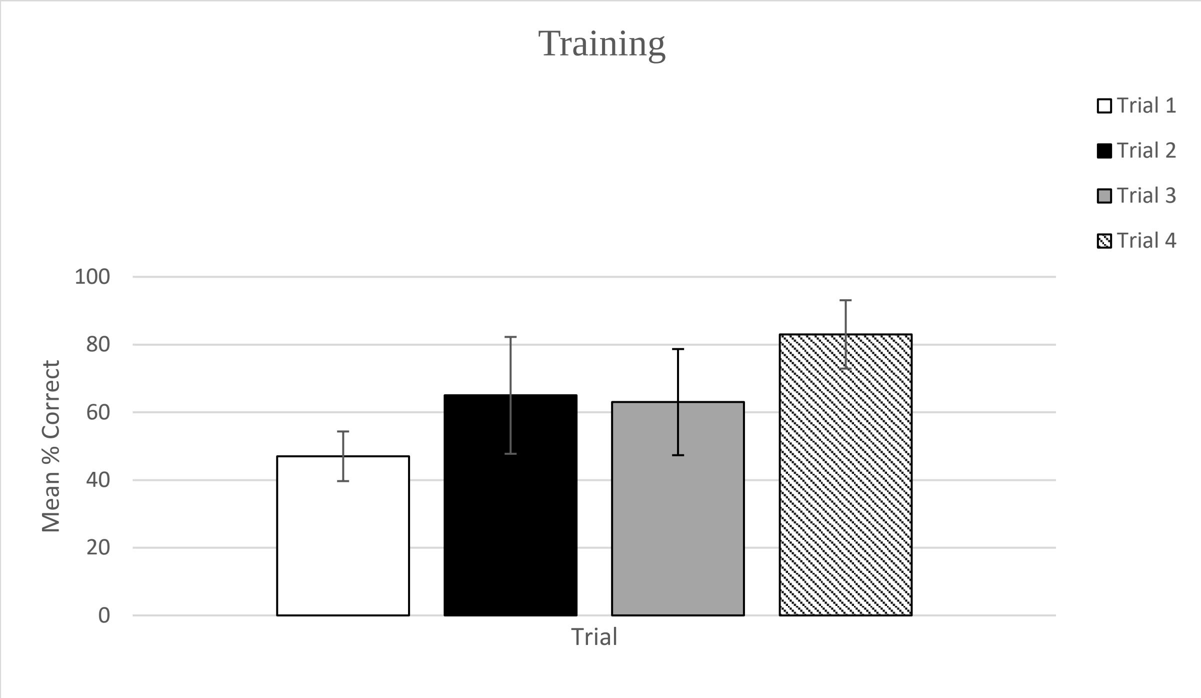


Figure 6. 5 Mean number correct responses for four trials in Training (with 95% confidence intervals as error bars)

6.5.4.3 Test

6.5.4.3.1 Recall

Accuracy data were entered into a one-way ANOVA with Day as the variable of interest (see Table 6.6). The results revealed that there was not a statistically significant main effect of Day, $F(2, 12) = 1.41$, $p = .283$, $\eta_p^2 = .829$.

To further analyse the types of responses that participants were making, an error analysis was conducted on participants' responses. Errors were classified as follows: *omissions* (i.e., participants did not give a response when the target item was presented), *real word intrusions* (RWI) (i.e., participants gave a real word response instead of the target novel word), *phonological errors* (i.e., participants made an error that deviated from the phonological structure of the novel word), *intra-experimental intrusions* (IEI) (recalling an item from the learning list that had been paired with a different picture object), (4) *extra-experimental intrusions* (EEI) (novel word responses that were phonologically dissimilar to the target).

A multivariate analysis of variance (MANOVA) with Day as a fixed factor was conducted. As dependent variables were uncorrelated, separate analyses of variance (ANOVAs) were employed to investigate the relationships between dependent variables (i.e., error types). The results revealed no statistically significant differences between days in terms of errors, $F(2, 84) = .68$, $p = .507$, $\eta_p^2 = .743$. Additionally, there were no statistically significant differences between different error types, $F(2, 84) = .42$, $p = .932$, $\eta_p^2 = .874$.

Table 6. 6 Response patterns in recall by day

	Correct	Errors				
		Omission	RWI	Phonological	IEI	EEI
Day 1	.180	.570	.030	.160	.030	.030
Day 2	.290	.430	.050	.121	.060	.030
Day 7	.290	.440	.060	.101	.060	.030

6.5.4.3.2 Recognition

A one-way ANOVA with Day as the variable of interest showed that there were no statistically significant differences across the three days, $F(2, 12) = .51$, $p = .616$, $\eta_p^2 = .362$.

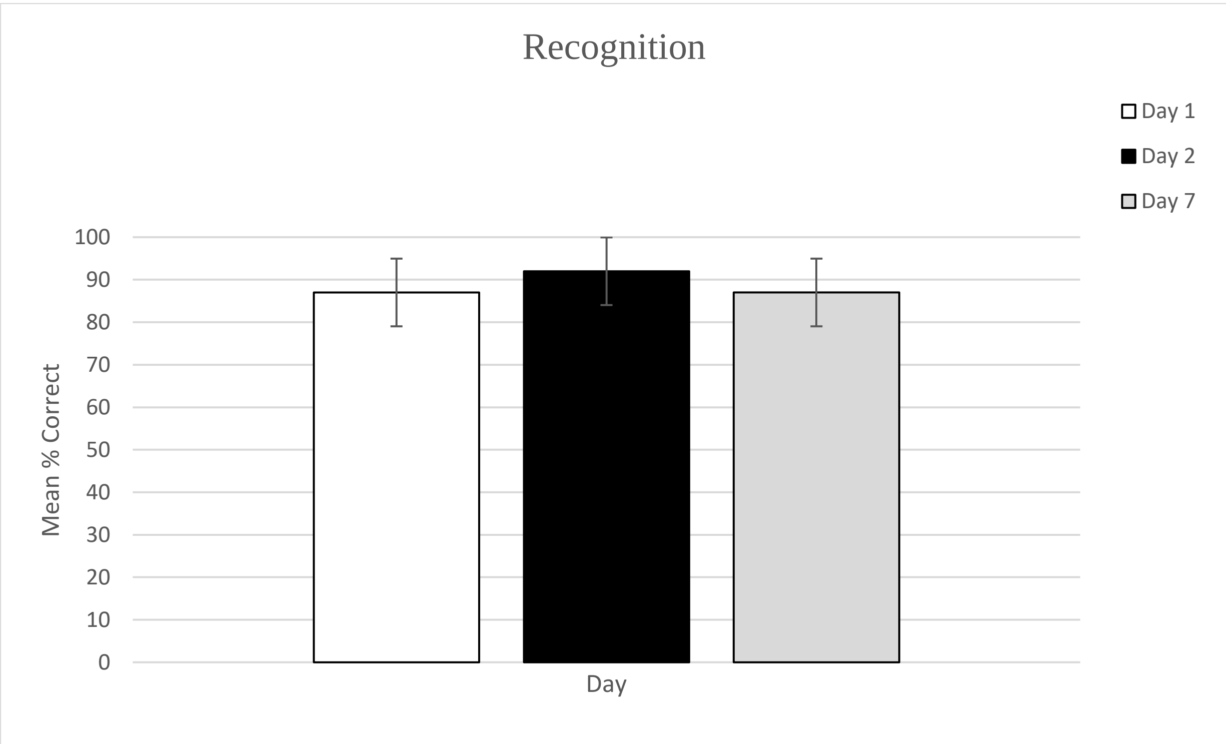


Figure 6. 6 Mean number correct responses for three days in recognition (with 95% confidence intervals as error bars)

This suggests that participants were able to form good associations between the verbal label (i.e., novel word) and a visual referent (i.e., novel object). The overall accuracy across all three days was quite high at 89% accuracy (see Figure 6.6).

6.5.5 Phase 4 Discussion

Overall, participants showed a training effect and thus an improvement across four training trials was observed. There were no statistically significant differences between the three days in terms of how well participants were able to match the verbal label to a visual object. This demonstrates that participants were able to learn the associations between the two stimuli and the response. Therefore, there were no difficulties in terms of establishing the integrity of the associations that were formed. This was also the same for recall rates over the three days.

Table 6. 7 Final novel words and mean number of correct responses in Familiarisation

	Set A		Set B
Item	Familiarisation Accuracy (%)	Item	Familiarisation Accuracy (%)
feəp	100	hemɪteltə	100
əfɑːdiən	100	mɒnveɪ	100
tʃrɪbdraɪvə	100	skwɪrɪt	100
aɪrəʊtreɪt	100	krəʊt	100
drəʊks	100	ɒtrɪpt	100
streɪbetʃi	100	wændaʊ	100
kætəbɜːgə	100	klaɪəs	100
trɔː	100	kraɪpɒkəbɑː	100
feləsuːsən	100	kɒtəmægən	100
keəhɒs	100	pɪkeɪtə	100
klaʊp	66.67	kengsuːn	100
kaɪnæfəl	66.67	enlədiːv	66.67
wɪlmɪkt	66.67	ʌsfrɒlə	50
treɪdʒ	50	eməfens	50
	mean = 89.29		*mean* = 90.48

Note: Mean Familiarisation accuracy for the revised set A is 89% (*SD* = 12.22) and 90% (*SD* = 12.95) for set B.

Based on the error analysis during Familiarisation, novel words which participants had difficulties repeating were removed. The aim was to eliminate the words that were consistently incorrectly produced and exhibited the worst performance on within the initial set of 40 novel words. Thus, the criteria for removing a novel word were those that had less than 40% accuracy across all the participants for both item sets. This resulted in the elimination of 12 items altogether, six from set A and six from set B, to ensure that participants could correctly repeat the novel words in the later stages of the experiment.

Thus, the final set of items comprised 28 novel words, 14 in each set (see Table 6.7). As a result of the revisions that were made to the task, Phase 5 was conducted to test the efficacy of the final set of items included in the paired associate learning task.

6.6 Phase 5

The previous section outlined Phase 4 which involved piloting of the paired associate learning task with a new group of participants to examine its efficacy. Phase 4 revealed that participants were able to learn the novel word-object associations and thus the task was deemed efficacious. However, the error analysis during the Familiarisation block revealed 12 novel words which participants had difficulties repeating and therefore these novel words were removed (see Section 6.5.5).

Having reviewed the verbal stimuli in the previous section, the following step involved piloting a revised final set of items with a new group of participants to examine the efficacy of the revised version of the paired associate learning task.

6.6.1 Participants

A new group of ten bilingual Polish-English participants (7 females and 3 males with *mean* chronological age 32 years and 5 months, *SD* = 6.90) who were residing in the UK participated in Phase 5. They had no known hearing or language impairments. Participants were asked to complete a Language Background Questionnaire in which they rated their Polish and English proficiency on a 7-point rating scale (where 1 means *very poor* and 7 means *native-like*) as well as their usage of Polish and English in the home/work/school environment on a 1-5 rating scale (where 1 means *never* and 5 means *all the time*).

Table 6. 8 Mean (and standard deviation) self-reported language proficiency and language usage in Polish-English bilinguals

	Mean	*SD*
Polish (first language) proficiency	7.00	0
English (second language) proficiency	4.80	.79
Polish usage at home	4.90	.32
Polish usage with friends	4.90	.32
Polish usage at work/study	4.90	.32
Mean Polish usage		
English usage at home	2.10	.99
English usage with friends	2.30	.82
English usage at work/study	3.00	.82
Mean English usage		
Age of exposure to English at home (in years)	n/a	n/a
Age of exposure to English at school (in years)	10.00	3.57

On average, participants started to learn English at school at the age of 10 and they did not have a significant exposure to English at home. Participants reported using Polish significantly more often than English, $t(9) = 2.03$, $p = .011$, $\eta_p^2 = .324$ (see Table 6.8).

6.6.2 Materials and Design

Based on the results obtained in Phase 4, several changes were implemented to the initial phase of the study. The total number of items was reduced from 40 to 28 (14 in each set) due to the difficulties encountered by participants in successfully producing some of the novel words in Familiarisation. In this experiment, 14 novel words, which were correctly recalled by adults in Phase 4 were selected as target stimuli to be learned by participants.

The number of Familiarisation trials was increased from one to two to ensure that participants received extended exposure to a spoken version of each novel word. Thus, in Phase 5 participants heard each target novel word twice. The Training block was increased from four to six trials. Test for Phase 5 was the same as in Phase 4 with the exception that participants completed recognition only on Day 7 (see Figure 6.7).

6.6.3 Procedure

The experiment was run across three separate days: Day 1, Day 2, and Day 7. Day 1 and Day 2 were separated by one day, with Day 7 occurring one week later.

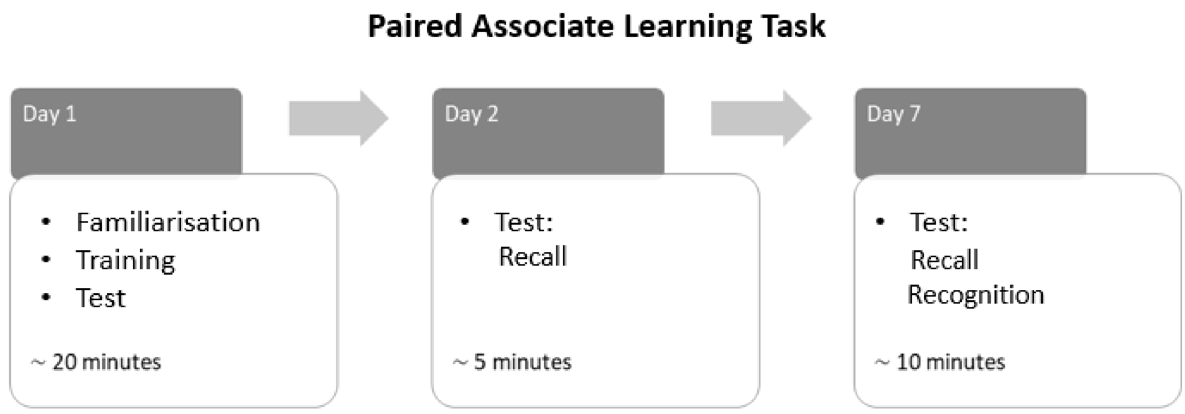


Figure 6. 7 Revised Paired Associate Learning task design

During Day 1, participants completed the Familiarisation block and the Training block followed by Test (Figure 6.7). Similarly, on Day 2 participants completed the Test block (i.e., recall). On Day 7, participants completed the Test block (i.e., recall and recognition). The experiment was employed using DMDX software (Forster & Forster, 2003) and was administered between January and August 2018.

Participants were tested individually in a quiet room.

6.6.4 Results

6.6.4.1 Familiarisation

On average, participants were able to repeat the novel words accurately 86 % of the time (*SD* = .11) in the first trial and 87% of the time (*SD* = .12) in the second trial. Overall, accuracy was high (i.e., 86.5%) and there was no statistically significant difference between trials, $t(9) = .47$, $p = .638$, $\eta_p^2 = .453$.

To further examine participants' performance on this task, an error analysis was conducted to understand the different types of responses that participants were making. The errors were classified as follows: *omissions* (i.e., participants did not give a response when the target item was presented), *real word intrusions* (RWI) (i.e., participants gave a real word response instead of the target novel word), and *phonological errors* (i.e., participants made an error that deviated from the phonological structure of the novel word).

Table 6. 9 Response patterns in Familiarisation

	Correct	RWI	Phonological
Trial 1	.859	.026	.098
Trial 2	.871	.019	.098

Table 6.9 presents the pattern of responses during Familiarisation. The majority of errors that participants were making were phonological in nature with only a smaller proportion of those being real word intrusion errors. A 2 (error type: RWI, phonological) x 2 (trial: Trial 1, Trial 2) repeated measures ANOVA revealed a statistically significant main effect of error type, $F(1, 9) = 19.24$, $p = .001$, with participants making more phonological errors than RWI. However, there was a nonsignificant main effect of trial, $F(1, 8) = .326$, $p = .573$, $\eta_p^2 = .635$; and a not statistically significant interaction between error type and trial, $F(1, 8) = .204$, $p = .655$, $\eta_p^2 = .324$.

6.6.4.2 Training

A one-way ANOVA showed that participants' performance improved across six training trials (see Figure 6.8), $F(5, 58) = 72.35$, $p = .021$, $\eta_p^2 = .438$.

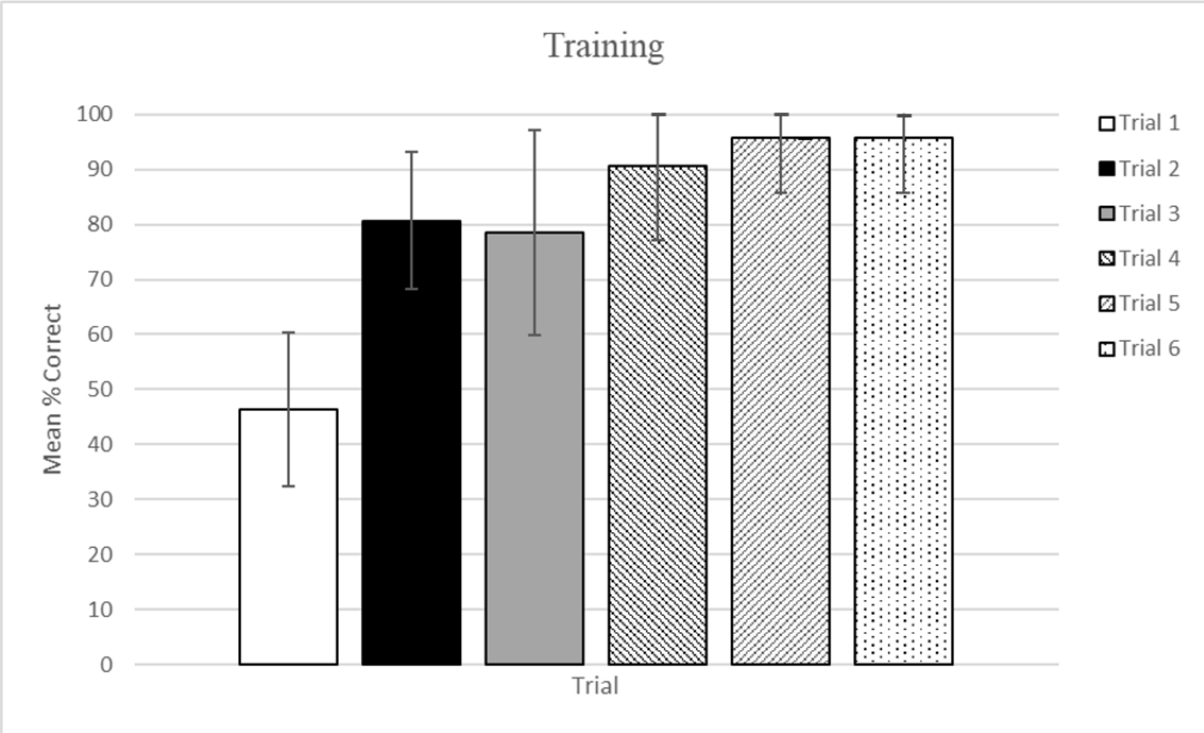


Figure 6. 8 Mean number correct responses for six trials in Training (with 95% confidence intervals as error bars)

6.6.4.3 Test

6.6.4.3.1 Recall

Accuracy data were entered into a one-way ANOVA with Day as the variable of interest (see Table 6.10). Results revealed no statistically significant main effect of Day, $F(2, 27) = .19$, $p = .828$, $\eta_p^2 = .567$.

To further analyse the types of responses that participants were making, an error analysis was conducted on participants' responses. Errors were classified as in Phase 4: omissions, RWI, phonological errors, IEI, and EEI.

Table 6. 10 Response patterns in recall by day

	Correct	Errors				
		Omission	RWI	Phonological	IEI	EEI
Day 1	.238	.460	.105	.093	.014	.041
Day 2	.212	.522	.128	.086	.007	.044
Day 7	.187	.527	.110	.086	.071	.032

A one-way multivariate MANOVA with Day as the variable of interest was conducted. The results revealed no statistically significant multivariate effect for errors, $F(1, 13) = .56$, $p =$

.466, η_p^2 = .766. Additionally, there were no differences between different error types, $F(2, 84)$ = .43, p = .932, η_p^2 = .856.

6.6.4.3.2 Recognition

Recognition was conducted only on Day 7. On average, in this block, participants were able to recognise the novel words accurately 91% of the time (*SD* = .63).

6.6.5 Phase 5 Discussion

Participants' performance on the Familiarisation and the Test block (i.e., recognition) was rather high at 86 % and 91% accuracy respectively. Overall, participants showed a training effect and thus an improvement across the six training trials was observed. There was a statistically significant difference between Day 1 and Day 2 for recall in terms of how well participants were able to produce the verbal label for the visual object. Thus, Phase 5 provided supporting evidence in terms of the efficacy of the revised paired associate learning task.

6.7 Discussion

As discussed in Chapter Two, paired associate learning is a paradigm that investigates one's ability to access and recall novel associations from episodic memory. The aim of this Chapter was to present the process of stimuli development and testing the efficacy of the paired associate learning task in which participants were asked to learn novel words for novel objects.

Phase 1 identified 40 novel word stimuli and assessed their familiarity to Polish. Similarly, Phase 2 resulted in a final selection of 80 novel objects which were further combined into 40 novel object pairings. Based on the results of Phase 3, 40 plausible novel word-novel object pairings were selected. Altogether, the three phases discussed above enabled the selection of suitable and valid stimuli for the paired associate learning task.

Phase 4 assessed the efficiency of the task. Overall, in Training, participants demonstrated that they were able to learn the associations between the two stimuli and the response and as such the integrity of the associations were formed. Error analysis revealed that participants had difficulties repeating certain novel words in Familiarisation and thus 12 novel words altogether were removed. This resulted in the final 28 novel words (two sets of 14 novel words) to be learned in the main experiment. The aim of Phase 5 was to assess the efficacy of the modified version of the paired associate learning task. In Phase 5, accuracy on

Familiarisation was improved thus providing further evidence that supported the efficacy of the paired associate learning task.

Chapter Eight will focus on the main experiment employing the task. Results will be discussed in the light of the current state of literature concerning paired associate learning in various linguistic groups. Firstly, Chapter Seven will discuss Experiment 1 where a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control) is investigated.

CHAPTER SEVEN

Experiment 1 – Is there a Bilingual Advantage in Working Memory and Executive Control?

7.1 Introduction

It has been proposed that bilingualism offers a wide range of benefits far beyond the obvious ones such as the ability to communicate in more than one language. As presented in Chapter Three, executive control (i.e., inhibition/attentional control) has been reported to be superior in bilingual populations (for a review see Bialystok, Craik, Klein, & Viswanathan, 2004) due to the bilingual management of two languages in terms of activation of the target language and inhibiting of the non-target language (Hilchey & Klein, 2011). However, the bilingual advantage in executive control has not been consistently found (De Bruin, Treccani & Della Sala, 2015) and thus bilingual research yields inconsistent findings (see Chapter Three). This is further corroborated by meta-analytic results in Chapter Four which indicate no bilingual advantage in executive control (i.e., inhibition/attentional control).

Based on the evidence discussed in Chapter One, working memory has been found to be implicated in executive control processes (Engel, Kane, & Tuholski, 1999). As such, it is of interest to this thesis to examine the relationship between executive control, working memory, and bilingualism. Research on the impact of bilingualism on working memory specifically is relatively sparse and inconclusive with studies evidencing the bilingual advantage (Bialystok, Craik, Luk, 2008; Morales, Calvo, & Bialystok, 2013), similar between language-group performance (Martin-Rhee & Bialystok, 2008; Luo, Luk, Bialystok, 2010), as well as bilingual disadvantage (Ratiu & Azuma, 2015) (see Chapter Three). The results of meta-analysis in Chapter Four supported the notion of the bilingual advantage in working memory and pointed to its significant role in episodic memory as measured by paired associate learning tasks. However, as revealed by results obtained in the systematic review of the relevant literature (see Chapter Four), there is a very limited number of studies that investigate the role of working memory in bilinguals.

In the context of this thesis, executive control is referred to as inhibition/attentional control and thus the Flanker task and the Simon task (see Chapter One for task characteristics) were employed to investigate the bilingual advantage in executive control. The congruency effect in the Simon and the Flanker tasks provides further information with regards to one's ability to efficiently inhibit a motor response (by refraining themselves from choosing the incorrect feature) and ignore irrelevant information (by non-attendance to distractors). Bilinguals have been reported to exhibit faster responses as well as smaller congruency effects when compared to monolinguals in both, the Flanker (Costa et al., 2008; Luk, De Sa & Bialystok, 2011) and the Simon tasks (Bialystok et al., 2004; Bialystok, Craik, & Ryan, 2006; Bialystok, 2017). As such, they have been found to resolve conflict more efficiently in comparison to their monolingual counterparts and this bilingual advantage has been conceptualised within the Bilingual Inhibitory Control Advantage (see Chapter Three for more details). The Bilingual Inhibitory Control Advantage assumes that bilinguals have more efficient inhibitory abilities compared to monolingual speakers as they need to manage language interference (Hilchey & Klein, 2011). Conflicting Monitoring Advantage is yet another proposal of the bilingual executive control advantage in tasks that require resolution of conflict (see Chapter Three for further discussion). As bilinguals are in constant need to monitor and resolve conflict within and between their languages, they have been claimed to have superior conflict monitoring skills which apparently seem to be also transferred to non-conflict situations (Von Bastian et al., 2016).

However, Miyake and Friedman (2012) shed some light onto how to interpret individual differences in terms of the executive control performance. Namely, they propose that once a subject has demonstrated superior interference control in one task, this should be replicated in other tasks measuring the same ability. This means that in tasks that explore the same executive control component, scores obtained by one participant should be correlated with each other and thus indicate convergent validity. Interestingly, research to date has shown that this is not always the case (Paap & Greenberg, 2013; Paap & Sawi, 2014). For instance, Paap and Greenberg (2013) demonstrated that cross-task correlation between Simon and Flanker was negative and weak ($r = -.01$).

Test-retest reliability of commonly used tasks measuring executive control has been investigated by Hedge, Powell and Sumner (2017). Results revealed that although the tasks are often used in experiments, their reliability is surprisingly low (it ranges from 0 to 0.82). It has been thus proposed that this might be caused by low between-participants variability and, although considered robust by many researchers, some measures fail to consistently capture individual differences within a given population. Hedge et al. (2018) explains that the reliability of a task will decrease with higher measurement error while between-subject variance does not change. In other words, some participants will score higher in a second task, whereas others will score lower. What is more, the task will have lower reliability if it demonstrates more homogeneity. Indeed, tasks that tend to display low reliability should not be used in correlations as it does not allow to distinguish between individual performance.

The investigation into test-retest reliability of executive control measures is still very limited and as such warrants a further investigation into the psychometric properties of these measures and their test-retest reliability. Within the context of this Experiment a multiple task approach was taken and a suite of measures to examine working memory was also employed: Digit and Backward Digit Recall to assess verbal working memory, as well as Dot Matrix and Mister X to assess visuospatial working memory. All measures are further discussed in more detail in this Chapter.

Thus, the aim of this Chapter is to investigate whether there is a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control) in paired associate learning tasks and examine the relationship between working memory and executive control. Additionally, correlational analyses are employed to determine whether this relationship is mediated by language proficiency in bilinguals.

The following main research questions were explored in this Experiment:

1) Is there a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control)?
2) Is there a relationship between working memory, executive control, and vocabulary?

Based on the reviewed literature, the following hypotheses guided this Experiment:

1) There is a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control).
2) There is a relationship between working memory, executive control, and vocabulary.

7.2 Participants

Forty-five participants were recruited for this study (33 females, 12 males) with *mean* chronological age of 30 years 6 months (*SD* = 9.02). Of this sample 17 were English monolingual adults (11 females, 6 males; with *mean* chronological age of 30 years and 6 months, *SD* = 5.21) and 28 Polish-English bilingual adults (22 females, 6 males; with *mean* chronological age of 31 years and 7 months, *SD* = 6.26) with no known language or hearing disorders. Participants were matched on age, $t(44) = -.21$, $p = .846$, $\eta_p^2 = 0.436$ and their socioeconomic status as measured by their education level, $t(44) = -4.57$, $p = .093$, $\eta_p^2 = .527$.

Participants tested in the UK were recruited by means of snowball sampling as well as through recruitment using the University of Sheffield MyAnnounce System where both staff and students were invited to participate via University mailing system. Subjects tested in Poland were recruited largely via snowball sampling. Researcher has established links at the University of Gdansk where they studied TOEFL as well as in a bilingual high school which they attended. This enabled recruitment and testing of participants who reported to be highly proficient in English.

Interested parties were asked to get in touch with the author to arrange a mutually convenient time for testing. Participants were tested by the author of this thesis. Participants completed the tasks individually in a separate room at the University campus.

7.2.1 Bilingual Sample

Bilingual participants were asked to complete the Language Background Questionnaire (Appendix G) in which they rated their Polish and English proficiency on a 7-point rating scale (where 1 means *very poor* and 7 means *native-like*) as well as their usage of Polish and English in the home/work/school environment on a 1-5 rating scale (where 1 means *never* and 5 means *all the time*) (see Table 7.1). Bilinguals reported to have started learning English as a second language at home on average at the age of 8 years, whereas 15 participants did not learn English in home settings at all. A few participants stated that their education of English at home resulted from, or was directly linked to, the beginnings of learning this language at school. The age of exposure to English in educational settings, as well as through private tutoring and reading English books in free time, was reported to start at around 9 years which used to be quite common in Poland for this age group. Twenty-six participants were University graduates (21 graduates from a university in Poland, five graduates from a university in England) and two current undergraduate students at the university in England.

Participants were all native speakers of Polish who were also proficient English speakers (see Table 7.1). They rated their proficiency in both languages and their usage in various contexts, as well as age of acquisition of each language within home and school settings. The average of their self-rating of their English proficiency was 5.46 (*SD* = 1.04). In the recruited sample, there was a significant difference between Polish and English proficiency, $t(27) = -7.99$, $p = .001$ with English being the language for which reported proficiency was lower. A paired sample *t*-test revealed that overall Polish-English bilingual participants reported using Polish significantly more than English, $t(27) = -3.87$, $p = .001$ (see Table 7.1). As they reported using Polish language significantly more often than English in all possible contexts provided ($p < .001$) and none of them was exposed to a substantial amount of English language at home during their childhood, they can be referred to as sequential bilinguals.

Table 7. 1 Mean (and standard deviation) self-reported language proficiency and language usage in Polish-English bilinguals

	Mean	*SD*
Polish (first language) proficiency	7.00	0
English (second language) proficiency	5.46	1.04
Balance Ratio	.78	
Polish usage at home	4.39	1.26
Polish usage with friends	4.61	.88
Polish usage at work/study	4.43	1.10
Mean Polish usage	4.76	
English usage at home	2.54	1.53
English usage with friends	2.82	1.28
English usage at work/study	3.62	1.06
Mean English usage	3.00	
Age of exposure to English at home (in years)	8.23 (13 ppts, 15 ppts - n/a)	4.28
Age of exposure to English at school (in years)	9.19	3.41

Additionally, for bilingual participants a balance ratio was estimated following the method proposed by Warmington, Kandru-Pothineni, and Hitch (2019). This balance ratio is calculated by dividing self-reported proficiency in English (second language) by their self-reported level of proficiency in Polish (first language). The balance ratio for this sample was .78 (*SD* = .15). This differed significantly from a value that might be taken to indicate perfect bilingual balance (i.e., a ratio of 1.0), $t(27) = 28.89$, $p = .001$ (Warmington, Kandru-Pothineni, & Hitch, 2019).

7.3 Materials and Design

7.3.1 Background Measures

7.3.1.1 Vocabulary and Fluid Intelligence

Participants were administered the Vocabulary and Matrix Reasoning subtests from the Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II, Wechsler, 2011) to assess vocabulary and fluid intelligence, respectively. Vocabulary subtest requires participants to provide verbal definitions of words in English. In Matrix Reasoning participants are shown a series of matrices with missing elements and have to identify the missing part that fits the pattern by choosing the correct answer out of five alternatives.

7.3.1.2 Simple Reaction Times

To assess processing speed participants were administered a Simple Reaction Time (SRT) task taken from Warmington, Hitch, and Gathercole (2013). Participants were required to respond to a visual target as quickly as possible via a button press. At the start of each trial the word READY was presented on the screen for approximately 700 milliseconds (ms), and was followed by the target which was presented either after 1, 3, or 5-second delay. There were 30 trials with 10 trials for each delay. Prior to the trials, the participants completed a practice which contained 6 trials, with 2 trials for each delay.

Simple Reaction Times task was used to ensure that the groups were matched on general processing speed. If no statistically significant difference between the groups' performance is found in this task, this suggests that a statistically significant difference on other reaction times tasks (here the Simon task and the Flanker task) arise from the underlying skills that the tasks are assessing.

7.3.2 Experimental Measures

7.3.2.1 Working Memory

Working memory was assessed using tasks adapted from the Automated Working Memory Assessment (AWMA: Alloway, 2007). The following working memory subtests were conducted in English only, the Polish version was not available. This experiment did not seek to make adaptations of the English version as the direct English to Polish translation does not allow a reliable comparison of cross-linguistic differences due to the varying numbers of letters and syllables between target words in both languages. Additionally, due to the time constrains it was not possible to develop and pilot a statistically reliable and valid assessment.

To assess verbal short-term memory and verbal working memory, participants were administered Digit Recall and Backward Digit Recall, respectively. In Digit Recall, participants were exposed to a series of digits being produced by a speaker and asked to recall them in the same order in which the digits were auditorily presented. In Backward Digit Recall, participants listened to a series of numbers and were asked to provide a response by recalling a series of numbers in a reverse order.

To conduct assessments of visuospatial short-term and visuospatial working memory, Dot Matrix and Mister X were employed respectively. In Dot Matrix, participants were required to remember and recall the location of a series of dots and the pattern in which all the dots were presented. Subsequently, participants were asked to point to a block on the grid where the dot had been displayed one by one in an order they had been presented. In Mister X, two cartoon characters holding a target object were presented and participants were asked to (1) decide whether a target object was located in the same or a different hand of one of the cartoon characters when compared to the other cartoon character and to (2) remember the location of the target object and, once it disappeared, point to where it was on a grid provided.

Each task consisted of practice and test. During practice, participants were provided with instructions and an opportunity to perform the task. During test, participants were presented with a set of items which increased in list length, starting from one item and ending with nine items. Each set consisted of six trials, and participants had to get four correct before moving on to the next list length. The test was terminated when participants made three errors.

The data analysis utilised the raw scores.

7.3.2.2 Executive Control

To assess executive control, participants were administered the Simon (Lu & Proctor, 1995) and the Flanker tasks (Eriksen & Eriksen, 1974). In the Simon task, participants were shown the face of a cat which was presented on either the left- or right-hand side of the computer screen. The cat could have either red eyes or blue eyes, and participants were required to identify the colour of the cat's eyes via button press (i.e., press the blue key for blue eyes and press the red key for red eyes). The presentation of the target (i.e., cat's face) was either congruent or incongruent to the location of the response button (i.e., right or left shift keys - left shift key was blue and right shift key was red). In congruent trials, the location of the cat's face was presented in the same location as the response button. In incongruent trials, however, the target was presented in a location on the screen that was inconsistent or opposite to the location of the correct response button thereby creating a conflict. As the target features were displayed on the screen in a random order and on a random side, this task scrutinised the ability to resolve conflict by suppressing initial motor reaction and ignoring interference caused by a distractor. Both accuracy of responses and reaction times for each condition were recorded.

On each trial the cat appeared either on the right or left side of the screen for 650 ms. The task consisted of a practice and a test block. The practice block consisted of eight trials (i.e., four congruent and four incongruent trials) and the test block consisted of 80 trials (i.e., 40 congruent and 40 incongruent trials). Trial presentation was randomised within each block and across participants.

The Flanker task was taken from Warmington, Kandru-Pothineni, and Hitch (2019) and was used to assess the ability to inhibit motor reaction, as participants were asked to identify the direction of a central target while ignoring the congruent or incongruent distracters that surrounded it. Specifically, five arrows, comprising a target (an arrow towards the left or the right) and four distractors (two arrows on both sides of the target), were visually presented on a computer screen in one row for the maximum of 650 ms. The target, always displayed in the centre of the screen, became the focal point of attention for participants. In the congruent trials, the four arrows (distracters) pointed in the same direction as the target arrow, whereas in the incongruent trials the distracters pointed in the opposite direction as the central arrow. Participants were instructed to ignore the orientation of the distractors in both congruent and incongruent trials and respond to the direction of the target as quickly and accurately as possible by a button press. The correct response was a choice of a button that is congruent with the target object placed in the centre of the screen. Similarly to the Simon task, both accuracy and reaction times for both conditions were the principal outcomes. Prior to the test block, the practice block of eight trials was introduced. The test task comprised two blocks with 40 congruent and 40 incongruent trials in each block.

7.4 Procedure

Participants completed vocabulary and fluid intelligence measures first followed by a speed processing task. After a short break, they were instructed to complete the Simon task which was also followed by a short break to ensure participants' comfort. The Flanker task was the final task to be completed during the session. Participants were provided with instructions first and given an opportunity to practise before the main tasks began.

The experiment was run using DMDX software (Forster & Forster, 2003). Participants completed the tasks individually in a separate room at the University campus. Each session lasted approximately 30 minutes.

7.4.1 Treatment of Data

Simple Reaction Times, Simon, and Flanker data were checked for normality. The average of reaction times and the average of accuracy was calculated for each participant. Additionally, erroneous responses and responses that were shorter than 200 ms and longer than 2000 ms, were regarded as outliers, and thus excluded prior to analyses. Scores were then transformed into inverse scores to reduce outliers. The proportion of outliers comprised 2.24% of all responses in Simple Reaction Times, 5.27% of all responses for Simon task and 1.85% for Flanker task.

Correlations between Simon and Flanker tasks were small and nonsignificant: in monolinguals (N = 17); r = .213, p = .427, and in bilinguals (N = 28); r = -.250, p = .200.

7.4.2 Reliability

Reliability was computed using Cronbach's α. Reliability in the Simple Reaction Times, Simon, and Flanker were good, .92., .94., and .95, respectively.

7.5 Results

7.5.1 Performance on Background Measures

An independent sample t-test revealed that participants were matched on fluid intelligence as measured by Matrix Reasoning, $t(44) = -1.05$, $p = .302$, $\eta_p^2 = .625$. However, an independent sample t-test revealed that English monolinguals' vocabulary scores were higher than Polish-English bilinguals, $t(44) = 2.13$, $p = .039$, $\eta_p^2 = .846$ (see Table 7.2). Mean and standard deviations for performance on background measures are presented in Table 7.2.

Another independent t-test revealed no significant difference between monolinguals and bilinguals in terms of reaction times on Simple Reaction Times task performance, t(44) =-1.33, $p = .192$, $\eta_p^2 = .214$. Additionally, no statistically significant difference between language groups was found for Simple Reaction Times accuracy data, $t(44) = -1.38$, $p = .177$, $\eta_p^2 = .312$.

Thus, it can be concluded that participants in the two groups were similar in terms of their processing speed.

Table 7. 2 Mean (and standard deviation) performance on background measures by group

	Monolinguals		Bilinguals	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*
WASI Matrix (Raw)	21.55 (2.33)	14-26	22.97 (2.13)	19-28
WASI Vocab (Raw)	38.41 (6.52)	30-50	34.72 (6.99)	19-45
Processing Speed Accuracy	95.83 (.06)	82-100	97.94 (.04)	82-100
Processing Speed RTs	292.06 (39.06)	244.86 375.31	306.83 (33.40)	249.64 - 366.30

7.5.2 Performance on Experimental Measures

7.5.2.1 Working Memory

The mean number correct, standard deviations, skewness, and kurtosis for both language groups are presented in Table 7.3.

Table 7. 3 Mean (and standard deviation) raw scores by working memory measures by group

	Monolinguals			Bilinguals		
	Mean (SD)	*Skewness (SE)*	*Kurtosis (SE)*	*Mean (SD)*	*Skewness (SE)*	*Kurtosis (SE)*
Digit Recall	37.06 (5.85)	.411 (.536)	-.803 (1.038)	29.46 (3.39)	.264 (.434)	-.443 (.845)
Backward Digit Recall	21.06 (7.17)	.312 (.536)	-.798 (1.038)	17.86 (5.01)	.690 (.434)	.048 (.845)
Dot Recall	30.18 (6.24)	.654 (.550)	1.499 (1.063)	30.00 (4.06)	-.412 (.434)	.005 (.845)
Mister X	22.29 (7.05)	1.233 (.550)	.936 (1.063)	22.25 (5.81)	.138 (.441)	-.316 (.858)

Skewness and kurtosis results indicate that the working memory data set is not normally distributed as the values are not close to zero. Verbal working memory in bilinguals, as measured by Backward Digit Recall, represents the largest skewness in this language group. In monolinguals, the largely skewed distribution was presented by visuospatial working memory, as measured by Mister X.

Further examination of boxplots revealed that the distributions for some tasks were not approximately normally distributed, and outliers were found. The outliers were removed (one outlier comprising 1.9% from bilingual dataset and two outliers comprising 2.7% from monolingual dataset).

Table 7. 4 Shapiro-Wilk test of normality for working memory measures by group

	Monolinguals	Bilinguals
Verbal STM	.926, p = .189	.960, p = .346
Verbal WM	.927, p = .197	.919, p = .033
Visuospatial STM	.957, p = .584	.973, p = .656
Visuospatial WM	.860, p = .015	.972, p = .644

Note: STM = short-term memory, WM = working memory

Test of normality was conducted for both language groups separately and the results are presented in Table 7.4. Thus, medians were also calculated for each working memory task for both target groups to reflect the non-normal distribution of variables. Medians are presented in Table 7.5.

Table 7. 5 Median raw scores by working memory measures by group

	Monolinguals	Bilinguals
Verbal STM	37	29.5
Verbal WM	21	18
Visuospatial STM	30	30.5
Visuospatial WM	19	22

In the working memory dataset, assumptions of normality and homogeneity of variance were violated for Digit Recall, $F(1, 42) = 6.37$, $p = .015$, $\eta_p^2 = .523$, and Backward Digit Recall, $F(1, 42) = 5.06$, $p = .029$, $\eta_p^2 = .424$. Dot Matrix and Mister X met the assumptions of normality and homogeneity of variance. As such, a non-parametric test was used to further analyse the data.

Independent-Samples Mann-Whitney U Test revealed a statistically significant difference between language groups in Digit Recall, U ($N_{monolingual} = 17$, $N_{bilingual} = 28$) = 259.00, $z = .073$, $p < .001$) with monolingual English participants being able to recall significantly more sequences of numbers than their bilingual counterparts (Table 7.5). No statistically significant differences between language groups were found for Backwards Digit Recall, U ($N_{monolingual} = 17$, $N_{bilingual} = 28$) = 196.00, $z = .073$, $p = .154$); Dot Matrix, U ($N_{monolingual} = 17$, $N_{bilingual} = 28$) = 244.00, $z = .008$, $p = .954$), and Mister X, U ($N_{monolingual} = 17$, $N_{bilingual} = 28$) = 259.00, $z = .073$, $p = .622$) and thus the null hypothesis of no difference between language groups was retained for these three tasks.

7.5.2.2 Executive Control

7.5.2.2.1 Simon task

7.5.2.2.1.1 Accuracy

The accuracy data were checked for normality and were approximately normally distributed for both groups (Test of Homogeneity of variances revealed the equal variances distribution; $F(1, 43) = .02$, $p = .893$, $\eta_p^2 = .653$). The descriptive statistics are provided in Table 7.6.

Table 7. 6 Mean (and standard deviation) percentage of correct responses in the Simon task

	Monolinguals	Bilinguals
Congruent	96.48 (.03)	96.79 (.03)
Incongruent	92.19 (.04)	92.68 (.06)

Overall accuracy on this task was excellent (i.e., 95.53%). A 2 (language group: bilinguals and monolinguals) x 2 (congruency: congruent and incongruent) repeated measures ANOVA revealed no statistically significant differences between monolinguals and bilinguals, $F(1, 43) = .13$, $p = .718$, $\eta_p^2 = .003$. However, there was a statistically significant main effect of congruency, $F(1, 43) = 27.39$, $p < .001$, $\eta_p^2 = .395$ as participants were significantly more accurate on congruent trials than on incongruent trials. Importantly, there was no statistically significant interaction between language group and congruency; $F(1, 43) = .01$, $p = .907$, $\eta_p^2 = .000$.

7.5.2.2.1.2 Reaction Times

The reaction times data were checked for normality and were approximately normally distributed for both groups (Test of Homogeneity of variances revealed the equal variances distribution, $F(1, 43) = .09$, $p = .757$). The descriptive statistics are provided in Table 7.7.

Table 7. 7 Mean (and standard deviation) reaction time (ms) in the Simon Task by group

	Monolinguals	Bilinguals
Congruent	415.86 (50.54)	429.07 (42.41)
Incongruent	461.50 (58.38)	457.72 (42.12)
Simon Effect	45.62 (19.71)	28.63 (23.01)

A 2 (language group: bilinguals and monolinguals) x 2 (congruency: congruent and incongruent) repeated measures ANOVA revealed no statistically significant main effect of the language group, $F(1, 43) = .11$, $p = .745$, $\eta_p^2 = .003$. As expected response latencies on congruent trials were significantly faster than on incongruent trials, $F(1, 43) = 117.18$, $p < .001$, $\eta_p^2 = .736$. Additionally, the interaction between language group and congruency was statistically significant, $F(1, 43) = 6.14$, $p = .017$, $\eta_p^2 = .127$.

This statistically significant interaction was further decomposed using a paired sample *t*-test with Bonferroni correction (i.e., $p = .025$). The results revealed that there was a statistically significant difference between the congruent and incongruent trials in monolinguals performance, $t(16) = -9.23$, $p < .001$, $\eta_p^2 = .837$. The same was observed for bilinguals with congruent trials being attended to faster than incongruent ones; $t(28) = -6.58$, $p < .001$, $\eta_p^2 = .679$.

The observed effect size for the monolingual group was larger than in the bilingual group, tentatively suggesting that the size of the congruency (or the Simon) effect (i.e., difference between reaction times on congruent and incongruent trials) was smaller in bilinguals. Thus, a paired sample *t*-test was conducted on the congruency effect to determine whether there were language groups differences, and results showed that the congruency effect was significantly smaller in bilinguals than in monolinguals, $t(44) = -2.48$, $p = .017$, $\eta_p^2 = .722$. Taken together, this suggests that the bilinguals resolved the conflicts more efficiently than the monolinguals.

7.5.2.2.2 Flanker task

7.5.2.2.2.1 Accuracy

The accuracy data were checked for normality (Test of Homogeneity of variances revealed the equal variances distribution, $F(1, 43) = 14.01$, $p = .062$). Descriptive statistics are presented in Table 7.8.

Table 7. 8 Mean (and standard deviation) percentage of correct responses in the Flanker task

	Monolinguals	Bilinguals
Congruent	97.35 (.03)	94.12 (.04)
Incongruent	98.84 (.02)	97.32 (.02)

The overall accuracy for this task was high (98.01%). A 2 (language group: bilinguals and monolinguals) x 2 (congruency: congruent and incongruent) repeated measures ANOVA revealed a statistically significant main effect of language group, $F(1, 43) = 12.26$, $p = .001$, $\eta_p^2 = .222$, with bilinguals being more accurate on the task; and a statistically significant main effect of congruency, $F(1, 43) = 23.59$, $p < .001$, $\eta_p^2 = .354$, with congruent trials being responded to more accurately than incongruent ones. The language group and congruency interaction was not statistically significant, $F(1, 43) = 3.08$, $p = .086$, $\eta_p^2 = .067$.

7.5.2.2.2.2 Reaction Times

The descriptive statistics for reaction times data for the Flanker task are presented in Table 7.9.

Table 7. 9 Mean (and standard deviation) reaction time (ms) in the Flanker Task by group

	Monolinguals	Bilinguals
Congruent	383.17 (79.33)	365.92 (35.96)
Incongruent	395.13 (81.23)	375.46 (34.52)
Flanker Effect	11.96 (13.76)	9.54 (16.40)

A 2 (language group: bilinguals and monolinguals) x 2 (congruency: congruent and incongruent) repeated measures ANOVA revealed a statistically significant main effect of congruency, $F(1, 43) = 20.42$, $p < .001$, $\eta_p^2 = .322$. However, there was no statistically significant main effect of language group, $F(1, 43) = 1.16$, $p = .288$, $\eta_p^2 = .026$; nor was there a statistically significant interaction between language group and congruency, $F(1, 43) = .26$, $p = .614$, $\eta_p^2 = .006$.

Though the Flanker effect (i.e., difference between reaction times on congruent and incongruent trials) was smaller in bilinguals (see Table 7.9), this difference was not statistically significant, $t(44) = 3.64$, $p = .614$.

7.5.2.3 Relationship between Working Memory, Executive Control, and Vocabulary

Since the initial analyses revealed that both verbal working memory scores (Digit Recall and Backwards Digit Recall: r = .503, p = .001 for monolinguals and r = .639, p = .026 for bilinguals) and both visuospatial working memory scores (Dot Matrix and Mister X: r = .508, p = .005 for monolinguals and r = .703, p = .013 for bilinguals) were significantly correlated, the composite scores were calculated. For each participant, working memory measures were divided into two main composite scores: verbal working memory score was computed by adding Digit Recall and Backwards Digit Recall together; similarly, visuospatial working memory score was the sum of Dot Matrix and Mister X. Both subsets comprise the same maximum raw scores, the same number of trials and blocks. As such, the subsets are identical in terms of their design and scoring procedure and can be combined by adding up scores to create a composite score. This approach was also taken by Alloway et al. (2005) and Warmington, Kandru-Pothineni, and Hitch (2018) who investigated working memory in monolingual and bilingual groups.

The rationale for creating the described composites stem from the fact that individual scores often do not provide an easy and accessible overview of participant’s performance. They are commonly used in research to represent concepts that are challenging or impossible to directly measure with a single task. Indeed, composite scores enable to aggregate individual performance measures from a number of tasks into one cumulative summary score. A single score tends to be simpler to analyse and interpret than multiple variables, whereas fewer variables reduce the probability of Type I errors.

Thus, to examine the relationship between working memory, executive control and vocabulary, verbal working memory composite, visuospatial working memory composite, the Flanker effect, the Simon effect, and vocabulary were entered into a correlation. For both groups, vocabulary was measured by the Vocabulary Subtest from the WASI. Additionally, for bilingual participants, the balance ratio and frequency of English usage (i.e., averaged across English usage at home, at work/school, and with friends) were also introduced.

Due to the non-normal distribution of certain variables imputed into correlational analyses, Spearman's rank correlation coefficient was used to investigate relationships between working memory, executive control, and vocabulary.

Table 7. 10 Spearman's correlations between working memory, executive control (i.e., inhibition/attentional control), and vocabulary by group

	1	2	3	4	5	6	7
1. Verbal Working Memory		.473	.443	-.150	-.018		
2. Visuospatial Working Memory	.372		.296	.386	-.103		
3. Flanker Effect	.015	.162		-.006	-.745**		
4. Simon Effect	.116	.155	-.149		.033		
5. Vocabulary (raw score)	.275	.195	-.261	.261			
6. English Usage	.269	.383*	-.093	.475*	.704**		
7. Balance Ratio	.371*	.204	-.387*	.276	.846**	.717**	

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

Note: Correlations for monolingual participants are above the diagonal and bilingual participants below the diagonal.

Table 7.10 presents the outcome of this analysis. For monolinguals, the Flanker effect correlated highly and significantly with vocabulary. No other significant correlation between working memory, executive control, and vocabulary were observed for this language group.

For bilinguals, verbal working memory correlated weakly and significantly with the balance ratio whereas visuospatial memory weakly and significantly with frequency of English usage. The Flanker effect correlated weakly and significantly with the balance ratio whereas the Simon effect moderately and significantly with frequency of English usage. Vocabulary demonstrated strong and highly significant correlation with frequency of English usage and with the balance ratio.

Together, these findings highlight that working memory and executive control (i.e., inhibition/attentional control) are not related in monolinguals and bilinguals. In monolinguals, executive control but not working memory is related to vocabulary, whereas in bilinguals, both working memory and executive control are related to the level of bilingualism (i.e., balance ratio) as well as frequency of English usage.

7.6 Discussion

The aim of this Chapter was to investigate whether there is a bilingual advantage in working memory and executive control (i.e., inhibition/attentional control) and examine the relationship between them. A suite of measures tapping executive control and working memory was employed including the Flanker and the Simon tasks to examine the former, while the latter was investigated by means of Digit Recall, Backward Digit Recall, Dot Matrix, and Mister X.

Research hypothesis 1 was not supported by this experiment as no bilingual advantage in working memory was found. However, analysis for verbal short-term memory, as measured by Digit Recall, showed a monolingual advantage. This is in line with studies indicating similar between language-group performance (Martin-Rhee & Bialystok, 2008; Luo, Luk, Bialystok, 2010; Engel de Abreu, 2011) and bilingual disadvantage (Ratiu & Azuma, 2015) (see Chapter Three). It needs to be noted though that stimuli used to assess verbal short-term and verbal working memory were based on English numbers and thus might explain better performance of the monolingual group.

Results of this Experiment revealed a significant correlation between bilingual performance on verbal working memory measure and their balance ratio score suggesting that highly proficient bilinguals were more successful at Digit Recall and Backward Digit Recall. This further indicates that the bilingual advantage in working memory may have not been found due to the lower English proficiency levels of the bilingual sample. It would be interesting to examine further whether employing first language stimuli to assess these memory components would yield different results. Overall, as revealed by the results obtained in the systematic review (see Chapter Four), there is a very limited number of studies that investigate the role of working memory in bilinguals.

Bilingual participants were found to resolve the conflicts in the Simon task significantly faster than monolinguals which is in line with previous research findings that indicate that bilingual participants are indeed found to demonstrate faster responses as well as smaller congruency effects when compared to monolingual counterparts (Bialystok et al., 2004; Bialystok, Craik, & Ryan, 2006; Bialystok, 2017). Taken together, this suggests that bilinguals resolved the conflicts more efficiently than their monolinguals.

This study did not replicate the bilingual advantage in the Flanker task found in previous studies (Costa et al., 2008; Luk, De Sa & Bialystok, 2011). Although the Flanker effect was smaller in bilinguals, this was not statistically significant. This may come as a result of varying English proficiency levels within the bilingual sample as indicated by highly significant correlations between the Flanker effect and the balance ratio as well as frequency of English usage (*rho* = .704 and *rho* = .846, respectively). Thus, similarly to working memory results discussed above, the bilingual advantage in executive control might not have been detected due to the low English proficiency levels of the bilingual sample. The score obtained for the balance ratio of the bilingual sample (*M* = .78, *SD* = .15), differed significantly from a value that might be taken to indicate perfect bilingual balance. This is in line with previous research findings indicating that the Flanker effect may not be observed in less proficient bilinguals due to the lack of enhanced executive control mechanisms engaged in language control (i.e., efficient managing of two languages in terms of activation of a target language and inhibiting of a non-target language, see Hilchey & Klein, 2011).

Results obtained on the Simon task support the Bilingual Inhibitory Control Advantage (Hilchey & Klein, 2011) proposing that bilinguals are more efficient at inhibitory control than their monolingual counterparts as they have to manage language interference from a non-target language. This leads to enhanced ability to resolve conflict. Bilingual Inhibitory Control Advantage predicts that bilinguals demonstrate faster reaction times in incongruent trials which results in better interference suppression and thus enhanced inhibitory control. This has been replicated in the Simon task with bilinguals performing faster than monolinguals in incongruent trials. There was, however, a nonsignificant difference between the groups.

Simon task and Flanker task both require the resolution of conflict and in both response selection is influenced by task-irrelevant stimuli. Simon task (Lu & Proctor, 1995) is typically used to assess the ability to resolve conflict by inhibiting an irrelevant motor response whereas Flanker task (Eriksen & Eriksen, 1974) assesses the ability to selectively attend to the target while ignoring distracting information (inhibiting irrelevant information). Although they share similarities, there are also certain differences in terms of the features of the irrelevant stimuli in the two tasks. In Flankers, both relevant and irrelevant stimuli are of the same type and spatially separated whereas in the Simon both relevant and irrelevant features are of a different type and are not spatially separated. What is more, in the former task irrelevant features are irrelevant only because of their location which is irrelevant for the task. Since the target location here is defined, the impact of irrelevant features can be constrained by spatial attention. In the latter task, on the other hand, stimulus location is generally irrelevant and interferes due to its correspondence to the required left or right response location (Hübner & Többel, 2019).

Compelling evidence in support of the bilingual advantage in inhibitory control would imply that both target measures, here the Simon task and the Flanker task, correlate with each other if they indeed both tap a common underlying skill associated with inhibitory control. Interestingly, in this study the correlations between them were small and nonsignificant. This is in line with findings by Stins et al. (2005) who found nonsignificant and small correlations between the Flanker, the Simon, and the Stroop task in their study with 12-year-olds. Similarly, Kousaie and Phillips (2012) and Keye, Wilhelm, Oberauer, and van Ravenzwaaij (2009) also obtained such results with no associations between the Simon task and the Flanker task suggesting that both measures may tap into different aspects of executive control (i.e., inhibition/attentional control) and thus support the notion of the unity and diversity of executive control (see Chapter One).

Bilingual research has often proposed that for the bilingual advantage in executive control to be coherently demonstrated and present, it should be found in two different tasks and these tasks should correlate with each other. If the tasks do not correlate with each other, as in this experiment, the bilingual advantage is most likely to be task-specific rather than domain-general. Paap and Greenberg (2013) explains that there is very limited evidence that the measures often employed to investigate individual differences in inhibitory control (i.e. Flankers, Simon) and widely acknowledge as valid constructs tap into the same domain-general ability. Indeed, as indicated in Chapter One, research has revealed that certain tasks that are assumed to indicate a specific executive process in one task, often do not predict this process in a related task (e.g., inhibitory control in Simon and Flanker task).

An alternative explanation as to why a bilingual advantage was found in the Simon task but not in the Flanker task can be found in Bialystok et al.' (2014) suggestion that the bilingual advantage in the former may in fact be linked to the enhanced ability to monitor attention as an answer to fast changing task demands rather than inhibitory control itself:

> "The advantage for bilinguals, therefore, may be not in the enhanced ability to inhibit the misleading spatial cue but in the ability to manage attention to a complex set of rapidly changing task demands'' (Bialystok, 2014, p. 292).

Similarly, Costa et al. (2009) propose that bilinguals may be more efficient in dealing with mixed congruent/incongruent trials in Simon and thus better at moving between conflicting and conflict-free trials, again pointing to monitoring once again. They also propose that monitoring of languages in a bilingual person leads to enhanced conflict monitoring skills in bilinguals:

> ''The bilingual advantage in overall RTs may reveal the better ability of bilinguals to handle tasks that involve mixing trials of different types: bilinguals would be more efficient at going back and forth between trials that require implementing conflict resolution and those that are free of conflict. This better functioning of the monitoring system may come about because of the bilinguals' need to continuously monitor the appropriate language for each communicative interaction. That is, proper communication in bilingual settings involves the monitoring of the language to be used depending on the interlocutor(s) language knowledge''. (Costa et al., 2009, p. 136)

These proposals suggest that the Simon and the Flanker tasks do indeed require different task-specific abilities and thus a bilingual advantage may be found in one but not in the other. Additionally, according to Paap and Greenberg (2013), there is more and more research available revealing that bilingual advantage in monitoring does not cooccur with enhancements in inhibitory control or switching.

One of the key limitations of the study is the relatively small sample size and varying proficiency levels within the bilingual sample which may contribute to the difficulties in capturing the bilingual advantage in executive control. Indeed, Martin-Rhee and Bialystok (2008) explained that highly proficient bilinguals are characterised by the enhanced ability to selectively attend to target cues in situations involving conflict resolution. As already discussed in the previous Chapter, the recruited sample of bilinguals was not highly proficient in their second language but the executive control advantage in the Simon task was still observed. There was no bilingual advantage in the Flanker task which may stem from the small sample size as well as low English proficiency of the bilingual sample. However, it may also be the case that the Simon task and the Flanker task tap different underlying executive control components. In fact, several tasks believed to be tapping the same executive control skill should be employed to examine bilingual advantage in executive control in the future (Paap & Greenberg, 2013).

CHAPTER EIGHT

Experiment 2 – Is there a Bilingual Advantage in Episodic Memory as measured by the Paired Associate Learning Task?

8.1 Introduction

As presented in Chapter Three, the majority of bilingual studies have focused on examining executive control and working memory, yet there still remains little consensus regarding the impact of bilingualism on executive control and working memory and the extent to which these two are organised and interrelated in bilinguals (for a review of bilingual advantage see Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012: however, see De Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2015). There is also very limited research that investigates how bilingualism modifies long-term memory. As revealed by the systematic review of available literature and meta-analyses in Chapter Four, the current state of findings yields inconclusive results.

Episodic memory, a long-term memory system which is of particular interest to this thesis, is typically assessed by paired associate learning tasks (see Chapter Two) as it requires individuals to first learn novel stimuli associations, then access them, and eventually retrieve them from long-term memory at a later date. The testing element of the task is divided into two types of measures tapping recall and recognition. In basic recognition measures, participants are presented with stimuli (i.e., words or pictures) and asked to decide whether they had already seen the items in the previous trials. Recall measures, on the other hand, include producing names of the targets (see Chapter Two for a more detailed description). Indeed, both recall and recognition are long-term memory assessments as they require participants to access and retrieve the targets from their episodic memory.

Papagno and Vallar (1995) were the first to utilise *paired associate learning task to investigate memory* mechanisms. They found that bilinguals demonstrate better performance on paired associate learning tasks which might stem from enhanced phonological working memory or phonological long-term memory (Papagno & Vallar, 1995; but see Kaushanskaya, 2012). In other words, it has been proposed that bilingualism may lead to a more general cognitive enhancement and in turn to a more efficient encoding and retrieval ability. Kaushanskaya and Marian (2009) propose that bilinguals might be more efficient in terms of allocating their selective attention to target information and this advantage leads to more efficient learning in general.

As bilinguals operate in dual language environments daily, they might have developed more efficient ways of encoding new information (i.e., novel words). In experiments where paired associate learning tasks are conducted in their dominant language, bilinguals often outperform monolinguals (Kaushanskaya, 2012; Kaushanskaya & Marian, 2009; Kaushanskaya & Rechtzigel, 2012; Bogulski, Bice, & Kroll, 2018).

There is very little research that investigates the role of bilingualism on long-term memory mechanisms, episodic memory in particular. In a study by Costa et al. (2008), bilingual participants exhibited an advantage in executive control which led to assumptions that their extensive experience of managing two languages on a daily basis results in executive control superiority. Executive control and episodic memory are interlinked, and thus Schroeder and Marian (2013) suggest that superior bilingual executive control may result in superior episodic memory performance in bilinguals. In their study with mono- and bilingual participants, they used a picture recall task which aimed to engage strategies encouraging visual encoding whilst discouraging verbal remembering at the same time. The subjects were asked to remember a series of pictures with complex scenes and they were not informed that they would be later instructed to recall the scenes. Indeed, bilinguals performed better by recalling more pictures than monolinguals. Another interesting explanation of how bilingualism impacts on episodic memory has been proposed by Ullman (2001). They suggest that the superiority in episodic memory performance in bilinguals may be directly linked to the limited number of memories encoded in their second language. In other words, as these memories are largely encoded in their first language, there is little competition between memories during their retrieval (Ullman, 2001).

Based on the research findings, bilinguals tend to outperform monolinguals on tasks including non-verbal stimuli (Costa et al., 2008). Bialystok et al. (2007) also suggest that bilingualism may contribute to memory improvements and delays in cognitive decline as bilinguals diagnosed with Alzheimer's dementia have been found to exhibit impairments much later than monolinguals.

There is limited understanding of the role of bilingualism on episodic and thus there is a precedent to investigate it in more detail. Meta-analyses presented in Chapter Four revealed significant differences in recall and recognition performance between monolinguals and bilinguals in studies included in the systematic review. This indicates a bilingual advantage in episodic memory as measured by paired associate learning tasks and suggests that it may be task-dependent. However, while the evidence presented in Chapter Four showed a bilingual advantage in paired associate learning, this evidence is not robust due to the limited number of studies and substantial between-study heterogeneity.

Additionally, both working memory and executive control (i.e., inhibition/attentional control) were found to play a significant role in bilingual episodic memory as measured by paired associate learning tasks. This is in line with previous research (i.e., Bialystok et al., 2003; Kaushanskaya & Marian, 2009). Again, due to a small number of studies eligible for meta-analytic considerations, no conclusive evidence is available as to whether these indices play a role in episodic memory in bilinguals.

To date, research concerning bilingualism has focused mainly on explicit memory mechanisms as opposed to implicit memory mechanisms. Paired Associate Learning task enables investigation of language learning which is in fact a form of implicit learning. Much less is however known about the processes involved in the integration of these novel associations into long-term memory and how they are retrieved from there. Current state of bilingual literature has broadly investigated the immediate short term memory tasks and experiments focused on a specific task. As such, further research is needed to enable the exploration of the role of bilinguals' executive control in episodic memory performance and further investigate the role of working memory and linguistic indices in this process. Therefore, the aim of this experiment is to extend our understanding of the impact of bilingualism on long-term memory mechanisms by employing the paired associate learning task to examine episodic memory. In this experiment, a suite of additional executive control, working memory, and vocabulary (as reported in Chapter Seven) measures were also employed to capture the interplay between the variables of interest.

Thus, taking this into account, the following research questions are stated:

1) Is there a bilingual advantage in episodic memory as measured by the paired associate learning task?
2) Is the bilingual advantage in episodic memory task-dependent (i.e., how paired associate learning is assessed: recall versus recognition measures)?
3) What is the role of working memory and executive control (i.e., inhibition/attentional control) in bilingual episodic memory performance as measured by the paired associate learning task?
4) What is the role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory performance as measured by the paired associate learning task?

5) Is there a relationship between working memory, executive control, vocabulary, and episodic memory?

Research questions 1 and 2 are answered from the perspective of the paired associate learning task which development and piloting were presented in more detail in Chapter Six. Research questions 3 and 4 comprise correlational analyses. Thus, executive control, working memory, and vocabulary measures are used as correlates to examine the relationships between episodic memory, working memory, and executive control.

Based on the reviewed literature, the following hypotheses guided this Chapter:

Research hypothesis 1: There is a bilingual advantage in episodic memory as measured by the paired associate learning task.

Research hypothesis 2: Bilingual advantage in episodic memory is task-dependent (i.e., recall versus recognition measures).

Research hypothesis 3: Working memory and executive control (i.e., inhibition/attentional control) play a significant role in bilingual episodic memory performance as measured by the paired associate learning task.

Research hypothesis 4: Bilinguals' vocabulary knowledge (in English) plays significant role in bilingual episodic memory performance as measured by the paired associate learning task.

8.2 Participants

Forty-four participants were recruited for this study (33 females, 11 males) with *mean* chronological age of 31 years 8 months (*SD* = 9.17). Of this sample 16 were English monolingual adults (11 females, 5 males; with *mean* chronological age of 28 years and 10 months, *SD* = 7.26) and 28 Polish-English bilingual adults (22 females, 6 males; with *mean* chronological age of 31 years and 7 months, *SD* = 6.26) with no known language or hearing disorders. Bilingual participants who participated in this experiment were the participants from Experiment 1 and their language background characteristics are presented in more detail in Chapter Seven. Additionally, one monolingual participant was not included in this experiment as they did not complete the paired associate learning task.

Participants were matched on age, $t(43) = -.20$, $p = .842$, $\eta_p^2 = .363$ and their socioeconomic status as measured by their education level, $t(43) = -3.35$, $p = .163$, $\eta_p^2 = 0.320$.

They were asked to report their level of education on a scale from 1 to 5 (1 = primary, 2 = secondary, 3 = vocational, 4 = sixth form, 5 = university).

Participants tested in the UK were recruited by means of snowball sampling as well as through recruitment using the University of Sheffield MyAnnounce System where both staff and students were invited to participate via University mailing system. Subjects tested in Poland were recruited largely via snowball sampling. Researcher has established links at the University of Gdansk where they studied TOEFL as well as in a bilingual high school which they attended. This enabled recruitment and testing of participants who reported to be highly proficient in English.

Interested parties were asked to get in touch with the author to arrange a mutually convenient time for testing. Participants were tested by the author of this thesis. Participants completed the tasks individually in a separate room at the University campus.

8.3 Materials and Design

8.3.1 Background Measures

8.3.1.1 Vocabulary and Fluid Intelligence

The same measures to assess vocabulary and fluid intelligence that were used in Experiment 1 (see Chapter Seven) were used in this experiment.

8.3.2 Experimental Measures

8.3.2.1 Executive Control and Working Memory

The same measures that were used to assess executive control and working memory in Experiment 1 (see Chapter Seven) were used in this experiment.

8.3.2.2 Paired Associate Learning Task

The paired associate learning task that was used in the Experiment was the final task which was outlined in Phase 5 of Chapter Six. In this task, participants learned the novel names for novel objects. The stimuli and the design of the task are the same as the ones used in Phase 5.

8.4 Procedure

The administration of the paired associate learning task was the same as in Phase 5 (see Chapter Six). Executive control and working memory measures were administered on a separate day to the paired associate learning task.

8.4.1 Treatment of Data

The paired associate learning task data were checked for normality and were approximately normally distributed for both groups for each block: Familiarisation, Training, and Test. The average of reaction times and the average of accuracy was calculated for each participant for each block: Familiarisation, Training, and Test. Additionally, for reaction times-based blocks, responses that were shorter than 200 ms and longer than 2000 ms were regarded as outliers, and thus excluded prior to analyses. The proportion of outliers across both groups comprised 3.21% of all responses in Training and 1.62% of all responses in Test (i.e., recognition).

8.4.2 Reliability

Reliability was computed using Cronbach's α. Reliability of the Familiarisation block was .72 and of the Training block was .76. The reliability of the Test block: recall and recognition was good at .82 and .86, respectively.

8.5 Results

8.5.1 Performance on Background Measures

An independent sample *t*-test revealed that participants were matched on fluid intelligence as measured by WASI Matrix Reasoning, $t(43) = -1.12$, $p = .270$. However, an independent sample *t*-test revealed that English monolinguals' vocabulary scores were higher than Polish-English bilinguals, $t(43) = 2.37$, $p = .022$ (see Table 8.1).

Table 8. 1 Mean (and standard deviation) performance on background measures by group

	Monolinguals		Bilinguals	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*
WASI Matrix (Raw)	22.15 (2.99)	14-26	22.97 (2.13)	19-28
WASI Vocab (Raw)	39.4 (6.45)	30-50	34.72 (6.99)	19-45

8.5.2 Performance on Experimental Measures

8.5.2.1 Executive Control and Working Memory

All the same participants (with the exception of one monolingual participant) from Experiment 1 did the tasks in Experiment 2. Even with one fewer participant, the pattern of results was the same as reported in Chapter Seven. Therefore, these will not be reported in detail here.

8.5.2.2 Paired Associate Learning Task

8.5.2.2.1 Familiarisation

Bilingual participants were able to repeat the novel words accurately 86 % of the time (*SD* = .10) and monolinguals 98% of the time (*SD* = .03). An independent-samples *t*-test revealed that monolingual participants were significantly more accurate in novel word repetition than bilinguals, $t(43) = 4.57$, $p = .001$, and they almost reached the ceiling threshold.

To further examine participants' performance on this task, an error analysis was conducted to understand the differences in responses that participants were making. The errors were classified as follows: omissions (i.e., participants did not give a response when the target item was presented), real word intrusions (RWI) (i.e., participants gave a real word response instead of the target novel word, i.e. ˈɛːrəpleɪn [aeroplane] instead of ˈaɪrəʊ.treɪt), phonological errors (i.e., participants made an error that deviated from the phonological structure of the novel word by a phoneme, i.e. ˈaɪrəʊ.traɪt instead of ˈaɪrəʊ.treɪt).

A 2 (language group: monolingual and bilingual) x 3 (error type: omission, RWI, phonological) ANOVA revealed a statistically significant main effect of error type, $F(2, 86) = 5.48$, $p = .006$, $\eta_p^2 = .314$, a statistically significant main effect of language group, $F(1, 43) = 20.85$, $p < .001$, $\eta_p^2 = .498$; and a statistically significant language group and error type interaction, $F(2, 96) = 7.83$, $p = .001$, $\eta_p^2 = .788$. Bilingual participants made significantly more RWI and phonological errors.

Response patterns for both groups in the Familiarisation block are presented in Table 8.2.

Table 8. 2 Response patterns in Familiarisation by group

	Correct	Errors

		Omission	Real Word	Phonological Error
Monolingual	.989	.010	0	0
Bilingual	.865	.012	.023	.099

8.5.2.2.2 Training

A 2 (language group: monolingual and bilingual) x 6 (training trials) ANOVA revealed a significant difference in accuracy performance across six Training trials, $F(5, 43) = 88.15$, $p < .001$, $\eta_p^2 = .726$, which suggests that participants' performance improved over time and they were able to form good associations between the novel words and the novel objects. The results also revealed a statistically significant main effect of language group, $F(1, 43) = 7.02$, $p = .011$, $\eta_p^2 = .624$, with bilingual participants being on average more accurate than monolinguals (80% ($SD = .10$) and 69% ($SD = .16$), respectively) (see Figure 8.1).

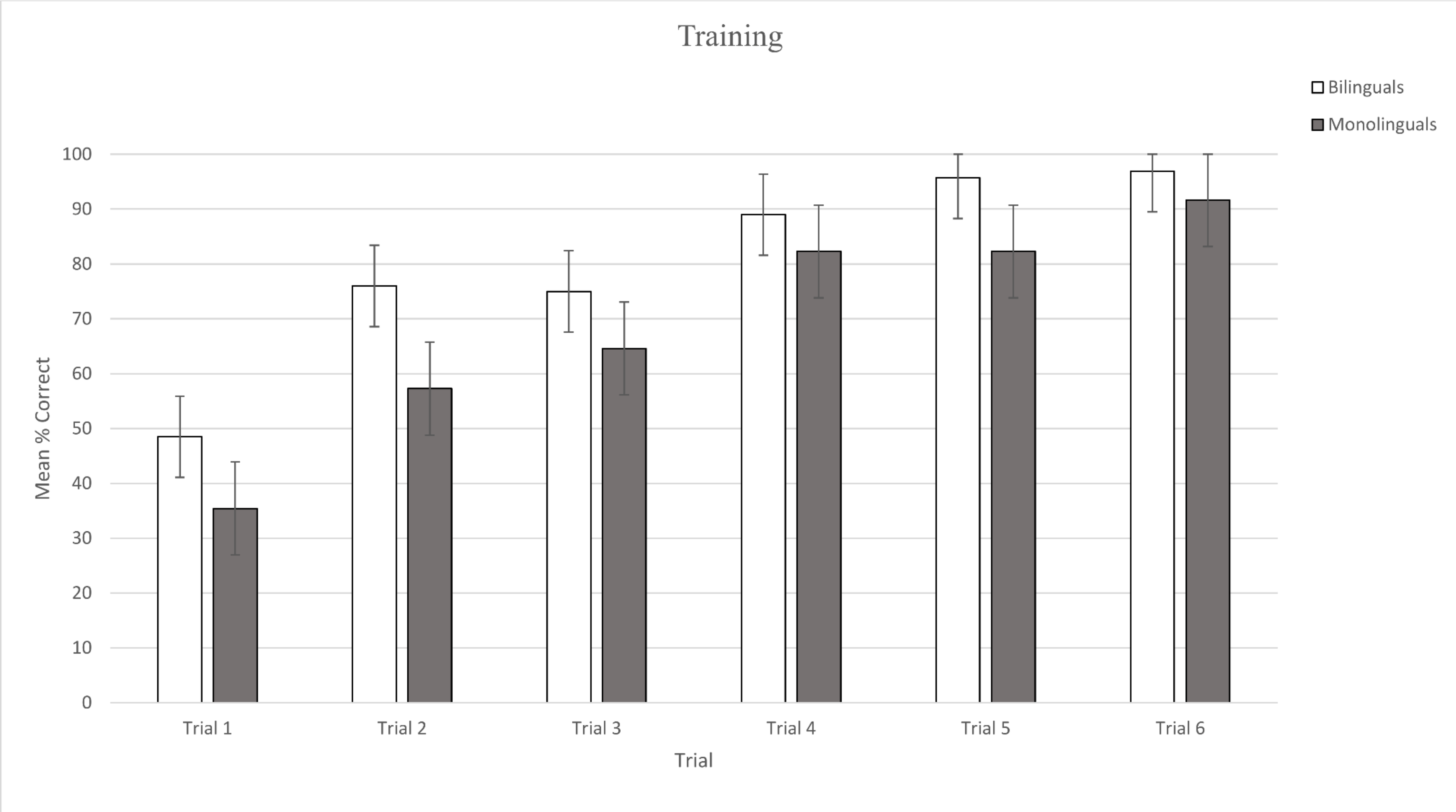


Figure 8. 1 Mean number correct responses for six trials in the Training block for bilinguals and monolinguals (with 95% confidence intervals as error bars)

8.5.2.2.3 Test

8.5.2.2.3.1 Recall

Accuracy data were entered into a 2 (language group: monolingual and bilingual) x 3 (day: Day 1, Day 2, Day 7) repeated measures ANOVA (see. Table 8.3). The results revealed no statistically significant differences between language groups, $F(1, 129) = 1.31$, $p = .255$, $\eta_p^2 = .301$. For both language groups, their accuracy on recall across three testing days was also not significantly different, $F(2, 129) = .43$, $p = .654$, $\eta_p^2 = .372$. Additionally, there was no significant interaction between language group and day, $F(2, 129) = 1.23$, $p = .287$, $\eta_p^2 = .291$.

To further analyse the types of responses that participants were making, an error analysis was conducted on participants' responses. Errors were classified as follows: omissions, RWI, phonological errors, IEI, and EEI.

Data was entered into a MANOVA with language group as the fixed factor and day (Day 1, Day 2, Day 7) as the covariate. The results revealed a statistically significant multivariate difference between language groups, $F(2, 129) = 32.41$, $p < .001$, $\eta_p^2 = .342$. As the dependent variables were uncorrelated, separate ANOVAs were conducted on each of the dependent variables (i.e., error types). Results revealed a statistically significant difference between language groups, with bilinguals making more omissions than monolinguals, $F(1, 129) = 21.17$, $p < .001$, $\eta_p^2 = .141$. There was also a statistically significant difference between language groups regarding EEI errors, $F(1, 129) = 41.49$, $p < .001$, $\eta_p^2 = .241$, with the monolinguals making more EEI errors than bilinguals. A statistically significant difference between language groups was also found regarding IEI errors, $F(1, 129) = 24.87$, $p < .001$, $\eta_p^2 = .159$, with monolinguals making significantly more IEI errors than bilinguals. No statistically significant differences were found in terms of errors types between days, $F(1, 132) = 17.15$, $p = .804$, $\eta_p^2 = .018$.

Table 8. 3 Response patterns in recall at Day 1, Day 2, and Day 7 by group

	Monolingual						Bilingual					
	Correct	Errors					Correct	Errors				
		Omissions	RWI	Phonological	EEI	IEI		Omissions	RWI	Phonological	EEI	IEI
Day 1	.260 (.32)	.239	.083	.104	.198	.073	.283 (.21)	.461	.106	.094	.042	.015
Day 2	.240 (.30)	.323	.094	.104	.188	.052	.212 (.17)	.522	.128	.086	.044	.007
Day 7	.323 (.29)	.292	.052	.073	.167	.094	.187 (.16)	.527	.111	.086	.032	.017

Another statistically significant difference was found between language groups' performance on IEI, $F(1, 129) = 25.61$, $p = .001$, with monolingual participants making on average more IEI errors than bilingual counterparts. There was no significant difference between language groups in terms of RWI, $F(1, 129) = 1.87$, $p = .174$, $\eta_p^2 = .042$ and no differences in phonological errors, $F(1, 129) = .082$, $p = .775$, $\eta_p^2 = .212$.

Error pattern analysis suggests that monolinguals were more likely to produce a novel name for a novel object presented even if their answer was not (entirely) correct while bilinguals were more likely to omit the production of the novel word.

8.5.2.2.3.2 Recognition

An independent sample *t*-test showed that there were no statistically significant differences in recognition performance between monolinguals and bilinguals, $t(43) = 1.41$, $p = .242$, $\eta_p^2 = .821$. The overall accuracy was relatively high in both groups with bilinguals scoring at 91% accuracy ($SD = .11$) and monolinguals at 84% accuracy ($SD = .23$) (see Figure 8.2).

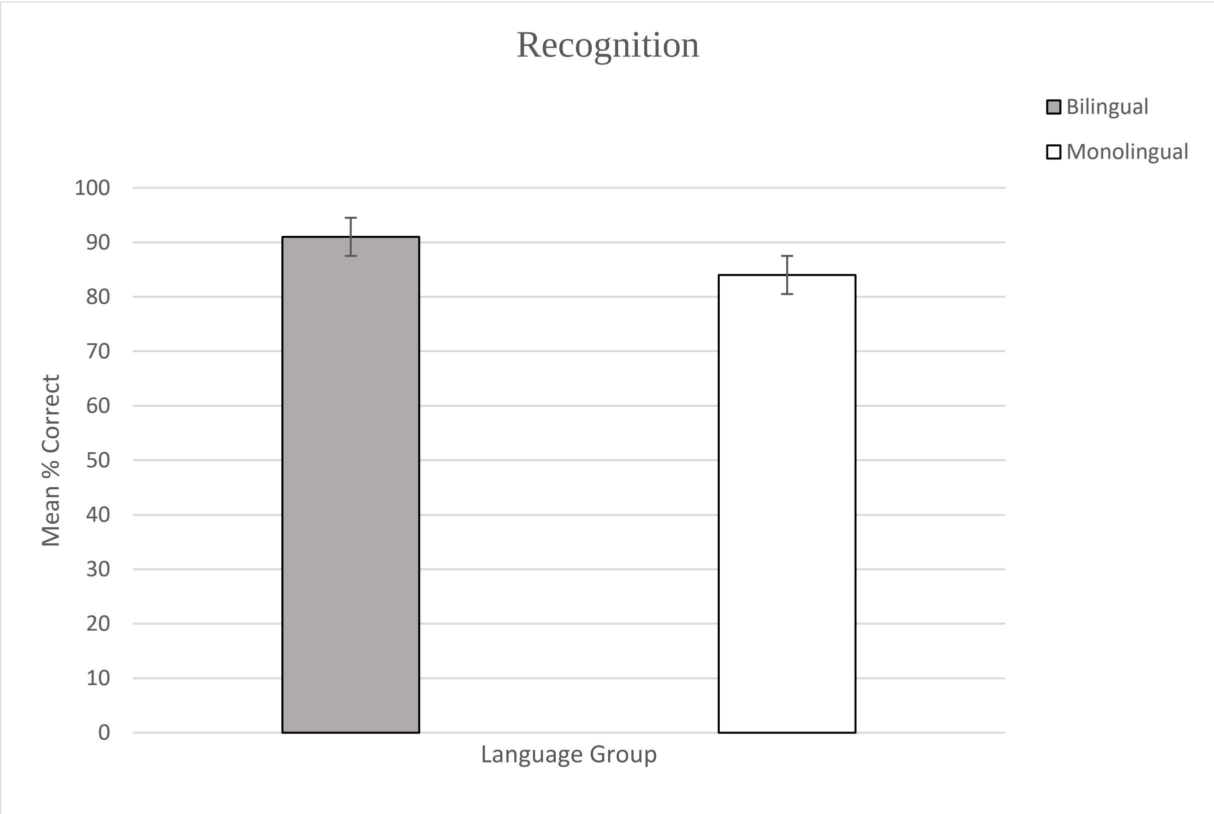


Figure 8. 2 Mean number correct responses in recognition by group

8.5.2.3 What is the Role of Working Memory, Executive Control (i.e., Inhibition/Attentional Control), and Vocabulary in Episodic Memory performance?

For each participant, recall score (i.e., average of accuracy) was calculated for three test days: Day, Day 2, and Day 7. Additionally, an average of accuracy score was computed for recognition. Recall and recognition constitute measures of episodic memory. Working memory measures were divided into two main composite scores: verbal working memory score was computed by adding Digit Recall and Backwards Digit Recall together; similarly, visuospatial working memory score was the sum of Dot Matrix and Mister X. In terms of executive control (i.e., inhibition/attentional control), the Flanker effect and the Simon effect were introduced into correlational analyses. For both groups, vocabulary was measured by the WASI Vocabulary Subtest. Additionally, for bilingual participants, the balance ratio and frequency of English usage (i.e., averaged across English usage at home, at work/school, and with friends) was introduced into correlational analyses.

Correlational analyses were carried out to examine the extent to which episodic memory (as measured by the paired associate learning task), working memory, executive control (i.e., inhibition/attentional control) and vocabulary are related to each other. Due to non-normal distribution of certain variables imputed into correlational analyses, Spearman's rank correlation coefficient was used to investigate relationships between episodic memory (as measured by the paired associate learning task), working memory, executive control, and vocabulary. The analysis included 44 participants (16 monolinguals and 28 bilinguals). For both groups, by participant correlations between all listed measures were computed and are presented in Table 8.4.

For monolinguals, all three recall days correlated strongly and highly significantly with each other (see Table 8.4) and verbal working memory but there was no significant correlation between recall and recognition. In fact, recognition did not correlate significantly with any of the imputed measures. The Flanker effect correlated highly and significantly with vocabulary. No other significant correlations between working memory, executive control, and vocabulary were observed for this language group.

For bilingual group, all paired associate learning measures correlated strongly and highly significantly with each other (three recall days and recognition, see Table 8.4). Additionally, recall Day 2 correlated weakly and significantly with the Flanker effect. There was a moderate and significant correlation between recognition and verbal working, and a strong and highly significant relationship between recall and visuospatial working memory which was not observed in monolinguals.

Bilingual verbal working memory correlated weakly and significantly with the level of bilingualism (i.e., balance ratio) whereas visuospatial memory weakly and significantly with frequency of English usage. The Flanker effect correlated weakly and significantly with the level of bilingualism (i.e., balance ratio) whereas the Simon effect moderately and significantly with frequency of English usage. Vocabulary demonstrated strong and highly significant correlation with frequency of English usage and with the level of bilingualism (i.e., balance ratio).

Together, these findings highlight that episodic memory (as measured by the paired associate learning task) is task-dependent and recall and recognition are significantly related in bilinguals. Moreover, bilingual episodic memory and working memory are also related. However, bilingual executive control is not significantly related to neither working memory nor episodic memory in both language groups studied. Also, performance on executive control measures is constrained by the level of bilingualism (i.e., balance ratio) as well as frequency of English usage.

Table 8. 4 Spearman's correlations between episodic memory, working memory, executive control (i.e., inhibition/attentional control), and vocabulary by group

	1	2	3	4	5	6	7	8	9	10	11
1. Recall Day 1		.651**	.666**	.193	.741**	.237	.151	-.240	-.154		
2. Recall Day 2	.755**		.868**	.317	.581*	.193	.221	-.207	.024		
3. Recall Day 7	.797**	.838**		.180	.527*	.242	.321	-.189	-.248		
4. Recognition	.550**	.593**	.501**		.240	-.020	.326	-.004	-.251		
5. Verbal Working Memory	.273	.221	.228	.426*		.473	.443	-.150	-.018		
6. Visuospatial Working Memory	.170	.076	.178	.621**	.372		.296	.386	-.103		
7. Flanker	-.127	-.390*	-.153	-.161	.015	.162		-.006	-.745**		
8. Simon	.192	.122	.243	-.025	.116	.155	-.149		.033		
9. Vocabulary	.331	.192	.172	.289	.275	.195	-.261	.261			
10. English Usage	.243	.058	.207	.182	.269	.383*	-.093	.475*	.704**		
11. Balance Ratio	.210	.152	.084	.224	.371*	.204	-.387*	.276	.846**	.717**	

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

Note: Correlations for monolingual participants are above the diagonal and bilingual participants below the diagonal.

8.6 Discussion

The aim of this study was to extend our understanding of the impact of bilingualism on long-term memory mechanisms by employing the paired associate learning task to examine episodic memory. In this Experiment, a suite of additional executive control, working memory, and vocabulary measures was also employed to capture the interplay between these indices.

As such, the following research hypotheses were made:

1) There is a bilingual advantage in episodic memory as measured by the paired associate learning task.
2) Bilingual advantage in episodic memory is task-dependent (i.e., recall versus recognition measures).
3) Working memory and executive control (i.e., inhibition/attentional control) play a significant role in bilingual episodic memory performance as measured by the paired associate learning task.
4) Bilinguals' vocabulary knowledge (in English) plays a significant role in bilingual episodic memory performance as measured by the paired associate learning task.

Overall, results revealed that monolinguals performed significantly better in terms of correctly articulating the novel words in the Familiarisation block and they also made significantly fewer real word and phonological errors than bilinguals. This is in line with previous research findings indicating that novel words in a second language that are phonologically similar to participants' first language are indeed learned more easily and efficiently (Storkel & Maekawa, 2005; Majerus, Poncelet, Greffe, & Van der Linden, 2008) than words that differ in this aspect. Although participants were learning novel word-object pairs that they had never encountered before, the novel words were conformed to the phonotactic structure of real English words (i.e. were matched on syllable, phoneme, stress pattern, bigram count and the target sound presence). In other words, they were phonotactically similar to real English words. This enabled the examination of the real process of word learning: bilinguals were arguably learning novel words in a second language and monolinguals in their first language. While novel words in the paired associate learning task were based on modifications to real English words, monolinguals found it easier to repeat and recall more words correctly. Bilingual performance on recall accuracy was worse which might have been due to the fact that Polish, their first language, comprises different sound patterns as the ones present in English.

Interestingly, however, monolingual accuracy performance in the Training block, where they were asked to correctly map novel words onto the novel objects, was significantly worse than bilingual participants' performance. Here, the bilingual advantage in mapping – that is establishing novel associative links between objects and words was found. Zerr et al. (2018) suggested that participants who tend to be more efficient at learning demonstrate faster processing speed and enhanced memory performance among others. One mechanism that may contribute to more efficient learning is inhibition/attentional control. According to Zerr et al., those individuals who learn more effectively are in fact more efficient at allocating attention when they learn new things. Such participants are less affected by interference and demonstrate less forgetting as they are better at allocating their attention to task-relevant information (Unsworth & Spillers, 2010).

In terms of the research hypotheses stated in this Chapter, hypothesis 1 was rejected as no bilingual advantage was found in episodic memory as measured by recall and recognition. Both monolinguals and bilinguals scored similarly on all three test days. This is inconsistent with the results identified in the systematic review which suggested that there is a bilingual advantage in episodic memory as measured by paired associate learning tasks. However, the results of meta-analyses in Chapter Four indicated substantial heterogeneity between studies and thus the results are inconclusive.

Additional recall error analyses revealed that bilinguals made significantly more omissions and significantly fewer extra- and intra-dimensional errors. Indeed, monolinguals seemed to recall novel word forms even if they were not sure that these are correct forms, whereas bilingual participants tended to refrain from novel word recall instead of attempting to produce a novel word form. There might be a few explanations to this phenomenon: first, bilinguals simply lacked confidence to recall novel words and chose to rather omit the trials. Another explanation might be to do with the fact that they also did not have any phonological representations that might have linked the novel words to contrary to monolinguals in whom novel words might have triggered activation of similarly sounding real words. As the novel words were modified existing English words, it might have been arguably easier for monolinguals to recall the targets or produce similarly-sounding but incorrect responses. Future analyses could investigate features of recalled words to investigate this even further. Fine analysis of words produced was not conducted in this study but it would be interesting to see where the error differences between monolinguals and bilinguals stem from. Such analysis might inform the current state of knowledge regarding differences in phonological processing of novel word forms in various language groups and examine whether such differences in performance are language specific.

Research hypothesis 2 regarding a bilingual advantage in episodic memory being task-dependent was true as correlational results revealed that performance on recall and recognition were moderately and highly significantly related in bilinguals. This relationship was not replicated in monolinguals. The results obtained in meta-analyses in Chapter Four further support the notion of a task-dependent bilingual advantage with different effect sizes observed for recall and recognition as measures of episodic memory.

Research hypothesis 3 was partially supported as bilingual episodic memory (i.e., recognition) and working memory were found to be related. This was not observed for recall. Also, bilingual executive control was not significantly related to neither working memory nor episodic memory in both language groups studied. Additionally, bilingual performance on executive control measures was constrained by the level of bilingualism (i.e., balance ratio) as well as frequency of English usage. This is partially inconsistent with meta-analysis results in Chapter Four indicating that both working memory and executive control (i.e., inhibition/attentional control) play a significant role in bilingual episodic memory performance. However, these meta-analyses comprised a very limited number of studies and indicated substantial between-study heterogeneity therefore the results are inconclusive.

Research hypothesis 4 was rejected as the role of vocabulary in episodic memory performance was found nonsignificant in both language groups. This is not in line with the systematic review findings (Chapter Four) that revealed a statistically significant large effect size indicating that vocabulary plays a significant role in bilingual episodic memory performance. However, only five papers were analysed in this meta-analysis (Chapter Four) and the results revealed substantial heterogeneity between the included studies.

Although the study aimed to recruit a homogenous sample of bilingual Polish-English speakers, a larger sample size of well-defined bilinguals should be recruited to investigate the differences found in the study and examine whether these are still to be observed. Although only proficient bilinguals were sought to participate in this experiment, their proficiency levels were based on self-reported language proficiency which, at times, was found to be not representative of an actual proficiency in the second language. As correlational analyses revealed, bilingual verbal working memory correlated weakly and significantly with the level of bilingualism (i.e., balance ratio) and visuospatial memory weakly and significantly with frequency of English usage. This indicated that working memory performance is closely related to participants' English proficiency and the frequency of using English. Moreover, the Flanker effect correlated weakly and significantly with the level of bilingualism (i.e., balance ratio) whereas the Simon effect moderately and significantly with frequency of English usage. This again suggests that proficiency in English impacted participants' performance on executive control (i.e., inhibition/attentional control) measures. Indeed, the low English proficiency level of the bilingual sample may be accountable for the lack of a bilingual advantage in episodic memory. Therefore, future research into whether bilingualism enhances episodic memory and the role that working memory, executive control, and vocabulary play in this process is much needed.

CHAPTER NINE

Experiment 3 – Is there a Bilingual Advantage in Episodic Memory as measured by Verbal Fluency Tasks?

9.1 Introduction

As outlined in Chapter Two, long-term memory mechanisms can be measured by not only paired associate learning but also by verbal fluency. By asking individuals to produce words that meet the task requirements, one is able not only to assess their executive control, but also test the verbal ability comprising lexical knowledge and ability to retrieve lexical representations from episodic memory.

In verbal fluency, participants are given 1-minute to produce as many words as they can that fit into a given fluency, letter and semantic, which are underpinned by different executive control mechanisms. Semantic fluency relies on existing links between concepts (words) that are related in some way whereas in letter fluency for the relevant words to be retrieved, one needs to first suppress other semantically related words that get activated automatically (Fisk & Sharp, 2004; see Chapter Two). Performance on semantic fluency is therefore largely an automatic and overlearned process (Luo et al., 2010) while letter fluency places increased executive control demands on word production. Contrary to semantic fluency, letter fluency relies on processes that are not often employed by people on a daily basis. According to Henry and Crawford (2004), individual performance on verbal fluency reflects working memory, set-shifting (Abwender et al., 2001), inhibition, and response suppression (Hirshorn & Thompson-Schill, 2006).

As verbal fluency incorporates executive control (i.e., inhibition/attentional control) as well as access to long-term memory systems, there are different ways of investigating individual differences in performance on the task. One of them is largely reliant on episodic memory access and it looks at the number of words accessed and retrieved correctly in letter and semantic fluency. Another perspective involves executive control and examines one's ability to switch and cluster during the task. Clustering refers to the ability to produce series of words one by one that are related to each other to some degree (e.g., when producing fruit, one may start with exotic fruit [banana, mango, pineapple], and then switch to another cluster of fruit [e.g., lemon, tangerine, grapefruit, etc.]). This clustering ability relies heavily on one's working memory. Switching, on the other hand, defines the ability to shift from one cluster to another (see Chapter Two for more details).

A more advanced and complex approach to the assessment of verbal fluency performance is a time-course analysis. The time-course analysis provides a more in-depth insight into the mechanisms that underpin word recall. The examination of the timing of the words generated by participants enables an insight into participants' lexical access speed, vocabulary size, and executive control (Luo, Luk, & Bialystok, 2010). In verbal fluency, the number of words produced declines as a function of time and this decline can be visually presented by plotting the number of retrieved words against time. This further allows an investigation of the slope of the function which provides additional information about participants' word recall patterns.

The time-course analysis provides two types of mean retrieval latencies (Rohrer, Wixted, Salmon, & Butters, 1995), in other words timings of words generated by participants, that form the main central tendency measures obtained in the analysis: first-response latency which is the time interval calculated from the beginning of a task until the onset of the first response, and subsequent-response latency which is the mean value of time intervals between the first response and each subsequent word produced. A short mean latency is indicative of a fast-declining rate of recall as the majority of the words are retrieved at the start of the trial and much fewer are produced later in the task. A long mean latency can be characterised by a slow decline in word retrieval or a relatively slow production of subsequent words throughout the task.

To distinguish between various word recall patterns, the three primary variables are thus taken into account: total number of words recalled, mean latencies, and the decline rate of word retrieval. These three considered together can be indicative of inferior or superior performance on recall (i.e., episodic memory), vocabulary knowledge, and executive control. The examples (Luo, Luk, & Bialystok, 2010) are as follows and will be further discussed in this chapter:

a) Participants who demonstrate slower lexical access produce fewer words and longer mean latencies resulting in a flat curve.
b) Participants with small vocabulary size tend to produce fewer words overall but demonstrate shorter mean latencies, and faster declining rates of word recall.
c) Participants who demonstrate superior executive control produce high numbers of total responses and can be defined by a slower declining retrieval rate.

Luo et al. (2010) proposed that executive control contribution to verbal fluency performance indeed increases with the length of the trial as later in the task participants need to bear in mind the words that they have already produced, and thus suppress interference from the earlier responses, as well as employ the ability to switch between word subgroups (i.e., exhausting the list of pets and starting to recall farm animals).

As research suggests, first language skills have been deemed crucial in second language word learning (e.g., Wang, Cheng, & Chen, 2006). Another body of research, however, points to a barrier to second language acquisition caused by cross-linguistic mismatch between both languages of a bilingual. The aim of this study was to investigate a bilingual advantage in episodic memory as measured by verbal fluency. The results obtained aim to lead to a further discussion regarding the bilingual advantage in verbal executive control mechanisms.

Taking this into account, the main research questions that are addressed in this Chapter include:

1) Is there a bilingual advantage in episodic memory as measured by verbal fluency?
2) Is there a relationship between paired associate learning, verbal fluency, working memory, executive control (i.e., inhibition/attentional control), and vocabulary?

The data included in the time-course analysis comprised only bilingual data. No data was available for monolinguals due to the data collection time constraints. In the context of this Chapter, it was of great importance to understand what mechanisms underpin bilingual verbal fluency and whether bilingualism impacts episodic memory. Based on the reviewed literature, the following hypotheses guided this Chapter:

Research hypothesis 1: There is a bilingual advantage in episodic memory as measured by verbal fluency.

Research hypothesis 2: There is a relationship between paired associate learning, verbal fluency, working memory, executive control (i.e., inhibition/attentional control), and vocabulary.

9.2 Participants

Forty-seven participants were recruited for this study (33 females, 14 males) with *mean* chronological age of 31 years 8 months (*SD* = 9.17). Of this sample 19 were English monolingual adults (11 females, 8 males; with *mean* chronological age of 32 years and 7 months, *SD* = 5.05) and 28 Polish-English bilingual adults (22 females, 6 males; with *mean* chronological age of 31 years and 7 months, *SD* = 6.26) with no known language or hearing disorders. Bilingual participants who participated in this experiment were the participants from Experiment 1 and their language background characteristics are presented in more detail in Chapter Seven. Monolingual participants were also the same participants as in the previous experiments plus additional three monolinguals were recruited.

Participants were matched on age, *t*46) = -.32, p = .724 and their socioeconomic status as measured by their education level, $t(46) = 3.12$, p = .062.

Participants tested in the UK were recruited by means of snowball sampling as well as through recruitment using the University of Sheffield MyAnnounce System where both staff and students were invited to participate via University mailing system. Subjects tested in Poland were recruited largely via snowball sampling. Researcher has established links at the University of Gdansk where they studied TOEFL as well as in a bilingual high school which they attended. This enabled recruitment and testing of participants who reported to be highly proficient in English.

Interested parties were asked to get in touch with the author to arrange a mutually convenient time for testing. Participants were tested by the author of this thesis. Participants completed the tasks individually in a separate room at the University campus.

9.3 Materials and Design

9.3.1 Background Measures

9.3.1.1 Vocabulary and Fluid Intelligence

The same measures to assess vocabulary and fluid intelligence that were used in Experiment 1 (see Chapter Seven) were used in this experiment.

9.3.2 Experimental Measures

9.3.2.1 Executive Control and Working Memory

The same measures that were used to assess executive control and working memory in Experiment 1 (see Chapter Seven) were used in this experiment.

9.3.2.2 Paired Associate Learning Task

The same measures that were employed to investigate paired associate learning in Experiment 2 (see Chapter Eight) were used in this experiment.

9.3.2.3 Verbal Fluency

Participants' verbal fluency was assessed using letter and semantic fluency. In both, participants were given 1 minute and were asked to produce as many words as they could that meet the requirements.

9.3.2.3.1 Letter Fluency

In letter fluency, participants were asked to produce words that started with a particular letter (P, K, W in Polish and F, A, S in English) within a time limit of one minute. Before the real task, participants were provided with an example of letter fluency ("for letter B in English you could say brown, bear, bounce, beautiful, bird and for letter T in Polish you could say twarz, truskawka, tonąć, twardy"). Participants were instructed that they could produce all parts of speech (nouns, verbs, adjectives, etc.) apart from proper names and place names (such as: Anna, Olaf, as well as KFC, Adidas, etc.). They were also instructed to avoid groups of words such as run, running, runner.

9.3.2.3.2 Semantic Fluency

In semantic fluency, participants were asked to generate as many words as possible that fit into a category (animals and things to eat) in both languages separately. Before the real task, participants were again provided with an example of semantic fluency ("for semantic category of clothes in English you could say skirt, jacket, gloves, scarf, trousers"). Participants were allowed to use all parts of speech (as in letter fluency) apart from proper names and place names as well as groups of words.

9.4 Procedure

Monolingual participants were tested in English only whereas bilinguals were assessed in both languages separately. For this task, correct and incorrect responses were recorded verbatim and were also audio recorded. Participants were tested one at a time in a sound-proof media booth on a university campus. Monolingual data was not audio recorded due to the data collection time constraints.

9.4.1 Treatment of Data

The number of correct responses, excluding repetitions and incorrect responses, was recorded for every participant for three letter (A, F, S in English and P, K, W in Polish) and two semantic (animals and things you can eat) fluencies in English and for bilingual participants additionally in Polish. The scores were also averaged across to provide a single overall average for both types of verbal fluency and these are referred to as the *mean number of correct responses*. The *mean number of correct responses* for each condition in each language separately for monolingual (*N* = 19) and bilingual (*N* = 28) participants are presented in Table 9.2.

9.4.1.1. Clustering and Switching

Following Troyer (2000) data were scored as (1) *total number of correct responses*, (2) *mean cluster size*, and (3) *number of switches*. The first score excluded incorrect responses, such as errors, intrusions and repetitions. For the second score, correct responses generated by each participant were divided based on either letter fluency or semantic fluency. Letter fluency was further divided into several categories: (a) words that begin with first two letters, i.e., *brown, bring*; (b) words that rhyme, i.e., *sand, stand*; (c) words that share first and last sounds, i.e., *pot, plant*; and (d) homonyms. Semantic fluency was distinguished between animal task and things to eat task. The former included categories such as: (a) African animals, (b) Australian animals, (c) Arctic animals, (d) Farm animals, (e) Water animals, amongst others, whereas in the latter the following were created: (a) Fruit, (b) Vegetables, (c) Dairy, (d) Meats, (e) Beverages, etc.

Words that shared the same characteristics constituted a cluster which size was based on the number of words calculated from the second response (n-1). For instance, a cluster comprising words *cow, horse, pig, hen, chicken* had a size of 4 (5-1). Mean cluster size score was generated by adding all the cluster sizes together and dividing the score by the number of clusters (Figure 9.1).

The number of switches was calculated as the number of transitions made by participants between adjacent clusters and single words generated for letter and semantic fluency tasks for both languages.

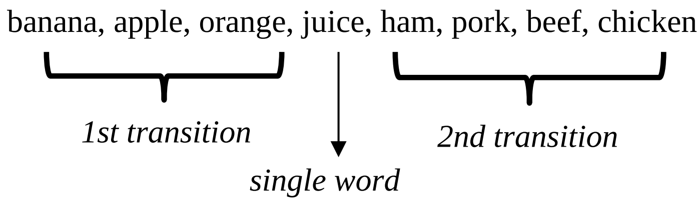


Figure 9. 1 An example of a cluster and a switch in verbal fluency

For the above example of things to eat semantic task, the number of clusters is two (fruits and meats) with a single word *juice* in-between. Thus, the size of the cluster can be calculated by starting from the second item: 2 (*apple,* orange) for fruit and 3 (*pork, beef, chicken*) for meat. Mean cluster size can be then determined by a simple calculation: (2 + 3)/2 which results in 2.5.

The mean cluster size and the number of switches were collapsed across letter and semantic fluency separately and calculated for each participant for each language.

9.4.2 Reliability

Reliability was computed using Cronbach's α. For English, the reliability of letter fluency was good, .85, and acceptable for semantic fluency, .76. For Polish, the reliability of letter fluency was also good, .82, and excellent for semantic fluency, .91.

9.5 Results

9.5.1 Performance on Background Measures

An independent sample *t*-test revealed that participants were matched on fluid intelligence as measured by WASI Matrix Reasoning, $t(46) = -1.17$, $p = .247$.

Table 9. 1 Mean (and standard deviation) performance on background measures by group

	Monolinguals		Bilinguals	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*
WASI Matrix (Raw)	22.15 (3.07)	14-26	22.97 (2.13)	19-28
WASI Vocab (Raw)	40.13 (6.38)	30-50	34.72 (6.99)	19-45

However, an independent sample t-test revealed that English monolinguals' vocabulary scores were higher than Polish-English bilinguals, $t(46) = 2.45$, $p = .018$ (see Table 9.1).

9.5.2 Performance on Experimental Measures

9.5.2.1 Executive Control and Working Memory

All the same participants (with the addition of two monolingual participants) from Experiment 1 did the tasks in Experiment 3. Even with two participants more, the pattern of results was the same as reported in Chapter Seven. Therefore, these will not be reported in detail here.

9.5.2.2 Paired Associate Learning Task

Similarly, all the same participants (with the addition of three monolingual participants) who took part in Experiment 2 did the tasks in Experiment 3. Due to the increased number of participants, the analyses were repeated and the pattern of results was the same as reported in Chapter Eight. Therefore, these will not be reported in detail here.

9.5.2.3 Is there a Bilingual Advantage in English Verbal Fluency?

9.5.2.3.1 Accuracy

Verbal fluency data was checked for normality and was approximately normally distributed for both groups for letter and semantic in English. Table 9.2 summarises the performance on verbal fluency for monolinguals and bilinguals.

Table 9. 2 Mean (and standard deviation) percentage of correct responses for verbal fluency by group

	Bilinguals	Monolinguals
	Mean (SD)	*Mean (SD)*
English Letter	12.50 (3.77)	15.33 (4.38)
English Semantic	19.37 (5.12)	25.42 (4.92)
Polish Letter	17.13 (3.92)	
Polish Semantic	27.24 (5.58)	

Note: Number of words produced averaged across subcategories: Letter (A+F+S)/3; Semantic (Animals+Food)/2.

A 2 (language group: monolingual and bilingual) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of verbal fluency, $F(1, 49) = 219.99$, $p < .001$, $\eta_p^2 = .815$. There was also a statistically significant main effect of language group, $F(1, 49) = 14.12$, $p < .001$, $\eta_p^2 = .220$, and a statistically significant interaction between language group and verbal fluency, $F(1, 49) = 7.89$, $p = .007$, $\eta_p^2 = .136$.

Post-hoc analysis with an independent samples *t*-test revealed that monolinguals produced significantly more correct responses in both verbal fluency in English compared to bilingual counterparts: letter, $t(46) = 2.45$, $p = .018$, and semantic, $t(46) = 4.15$, $p < .001$) (see Table 9.2).

9.5.2.3.2 Mean Cluster Size

A 2 (language group: monolingual and bilingual) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of the language group, $F(1, 49) = 107.48$, $p < .001$, $\eta_p^2 = .687$. The main effect of verbal fluency was significant, $F(1, 49) = 32.53$, $p < .001$, $\eta_p^2 = .399$, and an interaction of verbal fluency and language group was also significant, $F(1, 49) = 13.75$, $p = .001$, $\eta_p^2 = .219$.

Post-hoc analysis employing an independent samples *t*-test revealed the mean cluster size was significantly bigger for monolinguals than bilinguals for both fluency categories: letter, $t(49) = 11.66$, $p < .001$, $\eta_p^2 = .512$, and semantic, $t(49) = 5.06$, $p < .001$, $\eta_p^2 = .542$ (see Table 9.3 and 9.4).

9.5.2.3.3 Number of Clusters

A 2 (language group: monolingual and bilingual) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a significant main effect of verbal fluency, $F(1, 49) = 92.81$, $p < .001$, $\eta_p^2 = .654$, and a nonsignificant main effect of language group, $F(1, 49) = .95$, $p = .335$, $\eta_p^2 = .019$. An interaction of verbal fluency and language group was statistically significant, $F(1, 49) = 32.46$, $p < .001$, $\eta_p^2 = .398$.

Post-hoc analyses using an independent samples *t*-test revealed that the number of clusters was significantly bigger for bilinguals than monolinguals in letter fluency, $t(49) = -4.39$, $p < .001$, $\eta_p^2 = .763$. However, it was significantly bigger for monolinguals than bilinguals in semantic fluency (see Table 9.3 and 9.4).

9.5.2.3.4 Number of Switches

A 2 (language group: monolingual and bilingual) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of fluency, $F(1, 49) = 22.78$, $p < .001$, $\eta_p^2 = .317$, with more switches being generated in letter fluency than in semantic fluency (see Table 9.3). The main effect of the language group was also statistically significant, $F(1, 49) = 12.14$, $p < .001$, $\eta_p^2 = .199$, with monolingual participants generating on average significantly more switches compared to bilinguals (see Table 9.3 and 9.4). An interaction of fluency category and language group was not statistically significant, $F(1, 49) = 3.22$, $p = .079$, $\eta_p^2 = .062$.

Table 9. 3 Mean (and standard deviations) for English verbal fluency for monolinguals

	English	
	Letter *Mean* (*SD*)	Semantic *Mean* (*SD*)
Mean correct responses	15.33 (4.38)	25.42 (4.92)
Mean cluster size	3.21 (.88)	3.46 (.94)
Number of clusters	4.21 (1.75)	12.00 (3.56)
Number of switches	30.63 (9.01)	26.95 (5.58)

9.5.2.4 Verbal Fluency in First Language

9.5.2.4.1 Accuracy

Verbal fluency data was checked for normality and was approximately normally distributed for letter and semantic in English and in Polish.

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA with language group as a between-factor was utilised to compare verbal fluency performance in participants' first languages: English for English monolingual speakers and Polish for Polish-English bilinguals. Results revealed a non-significant main effect of verbal fluency, $F(1, 28) = 3.71$, $p = .065$, $\eta_p^2 = .117$. There was a statistically significant main effect of language, $F(1, 28) = 90.32$, $p < .001$, $\eta_p^2 = .763$, and a non-significant interaction between language and verbal fluency, $F(1, 28) = 1.64$, $p = .212$, $\eta_p^2 = .550$.

9.5.2.4.2 Mean Cluster Size

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA with language group as a between-factor revealed a statistically significant main effect of language, $F(1, 27) = 221.85$, $p < .001$, $\eta_p^2 = .891$, and a statistically significant main effect of verbal fluency, $F(1, 27) = 15.26$, $p < .001$, $\eta_p^2 = .421$. An interaction of verbal fluency and language was non-significant, $F(1, 27) = 2.47$, $p = .127$, $\eta_p^2 = .084$.

9.5.2.4.3 Number of Clusters

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA with language group as a between-factor revealed a statistically significant main effect of language, F(1, 27) = 29.78, p < .001, $\eta_p^2 = .524$, and a statistically significant main effect of verbal fluency, $F(1, 27) = 65.75$, $p < .001$, $\eta_p^2 = .709$. An interaction of verbal fluency and language was non-significant, $F(1, 27) = 3.15$, $p = .087$, $\eta_p^2 = .104$.

9.5.2.4.4 Number of Switches

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA with language group as a between-factor revealed a statistically significant main effect of language, F(1, 27) = 18.53, p < .001, $\eta_p^2 = .329$, and a statistically significant main effect of verbal fluency, $F(1, 27) = 53.75$, $p < .001$, $\eta_p^2 = .732$. An interaction of verbal fluency and language was non-significant, $F(1, 27) = 4.12$, $p = .092$, $\eta_p^2 = .107$.

9.5.2.4 Lexical Access in Bilinguals – Cross-Linguistic comparisons

9.5.2.4.1 Accuracy

Verbal fluency data was checked for normality and was approximately normally distributed for letter and semantic in English and in Polish.

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of language, $F(1, 27) = 84.55$, $p < .001$, $\eta_p^2 = .751$. Additionally there was a statistically significant main effect of verbal fluency, $F(1, 27) = 204.81$, $p < .001$, $\eta_p^2 = .880$. The interaction between language and verbal fluency was statistically significant, $F(1, 27) = 16.72$, $p < .001$, $\eta_p^2 = .374$.

Table 9. 4 Mean (and standard deviations) for Polish and English verbal fluency for bilinguals

	Polish		English	
	Letter *Mean* (*SD*)	Semantic *Mean* (*SD*)	Letter *Mean* (*SD*)	Semantic *Mean* (*SD*)
Mean correct responses	17.13 (3.92)	27.24 (5.58)	12.50 (3.77)	19.37 (5.12)
Mean cluster size	1.57 (.43)	2.89 (.68)	1.26 (.26)	2.40 (.55)
Number of clusters	10.56 (4.02)	12.31 (2.10)	7.81 (3.29)	9.81 (3.09)
Number of switches	33.96 (7.07)	22.40 (6.18)	26.71 (8.87)	18.59 (5.52)

Post-hoc analysis using a paired sample *t*-test revealed significantly more correct responses being recorded for both verbal fluency in participants' first language, Polish, compared to their performance in English: letter fluency, $t(27) = 6.79$, $p < .001$, $\eta_p^2 = .854$, and semantic fluency, $t(27) = 9.77$, $p < .001$, $\eta_p^2 = .732$ (see Table 9.4). Also, bilingual participants generated significantly more words for the semantic fluency than for the letter fluency in both languages (Polish, $t(27) = -14.63$, $p < .001$, $\eta_p^2 = .897$, and English, $t(27) = -10.51$, $p < .001$, $\eta_p^2 = .788$) (Table 9.4).

9.5.2.4.2 Mean Cluster Size

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of language, $F(1, 27) = 15.52$, $p < .001$, $\eta_p^2 = .334$, and a statistically significant main effect of verbal fluency, $F(1, 27) = 240.99$, $p < .001$, $\eta_p^2 = .886$. The cluster size was significantly bigger for Polish than for English and also significantly bigger for semantic than for letter fluency (see Table 9.4). The language and verbal fluency interaction was not statistically significant; $F(1, 27) = 1.23$, $p = .276$, $\eta_p^2 = .038$.

9.5.2.4.3 Number of Clusters

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of language, $F(1, 27) = 27.84$, $p < .001$, $\eta_p^2 = .473$. and a statistically significant main effect of verbal fluency, $F(1, 27) = 17.22$, $p <$

.001, η_p^2 = .357. Number of clusters was significantly bigger for Polish than for English and also significantly bigger for semantic than for letter fluency (see Table 9.4). The language and verbal fluency interaction was not statistically significant; $F(1, 27) = .08$, $p = .784$, $\eta_p^2 = .002$.

9.5.2.4.4 Number of Switches

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effect of language, $F(1, 27) = 31.40$, $p < .001$, $\eta_p^2 = .503$, and a statistically significant main effect of verbal fluency, $F(1, 27) = 75.94$, $p < .001$, $\eta_p^2 = .710$. Number of switches was significantly bigger for Polish than for English and also significantly bigger for letter than for semantic fluency (see Table 9.4). The language and verbal fluency interaction was not statistically significant, $F(1, 27) = 94.53$, $p = .070$, $\eta_p^2 = .102$.

9.5.2.5 The Relationship between Episodic Memory (as measured by Paired Associate Learning task and Verbal Fluency), Working Memory, Executive Control, and Vocabulary

For both groups of participants, correlational analyses were conducted to investigate the relationship between episodic memory (as measured by the paired associate learning task and verbal fluency), working memory, executive control, and vocabulary by means of Spearman's rank correlation coefficient due to non-normal distributions of certain variables.

Verbal fluency for English correlates comprised mean correct responses in letter and semantic fluency, mean cluster size, number of clusters, as well as the number switches (for bilinguals these all were additionally entered in Polish). For each participant, recall score (i.e., average of accuracy) was calculated for three test days: Day 1, Day 2, and Day 7. Additionally, an average of accuracy score was computed for recognition. Verbal fluency, recall, and recognition constituted measures of episodic memory. For each participant, working memory was divided into two main composite scores that were calculated separately. Verbal working memory score was computed by adding Digit Recall and Backwards Digit Recall together; similarly, visuospatial working memory score was the sum of Dot Matrix and Mister X. In terms of executive control (i.e., inhibition/attentional control), the Flanker effect as well as the Simon effect were included in correlational analyses and additionally, for bilingual participants, the level of bilingualism (i.e., balance ratio) and the frequency of English usage.

Correlational analyses were carried out to examine the extent to which episodic memory, working memory, executive control (i.e., inhibition/attentional control), and vocabulary are related to each other. The analysis included 45 participants (17 monolinguals and 28 bilinguals). For both groups, by participant correlations between all listed measures were computed and are presented in Table 9.5.

9.5.2.5.1 Verbal Fluency, Working Memory, Executive Control (i.e., Inhibition/Attentional Control), and Vocabulary

Table 9.5 provides a summary of correlations between verbal fluency, working memory, executive control (i.e., inhibition/attentional control), and vocabulary. For both groups, English letter and semantic fluency correlated with each other significantly. Additionally, a series of significant correlations between verbal fluency indices were observed (see Table 9.5).

For bilinguals, letter fluency correlated very strongly and highly significantly with the number of clusters and switching for letter. Both letter and semantic fluency in English correlated highly significantly with vocabulary and the level of bilingualism (i.e., balance ratio). None of the verbal fluency measures did not correlate with executive control (i.e., the Flanker effect or the Simon effect). Number of clusters (as well as switching) in semantic verbal fluency correlated significantly with visuospatial memory.

For monolinguals, letter and semantic verbal fluency correlated highly and significantly, both were also significant correlates with number of clusters and switching for letter fluency. No other significant relationships between included variables were observed in monolinguals. In terms of other significant correlations between working memory, executive control, and vocabulary, these yielded the same results as presented in Chapter Eight.

9.5.2.5.2 Paired Associate Learning, Verbal Fluency, and Vocabulary

Table 9.5 provides an overview of correlations between the paired associate learning task, verbal fluency, and vocabulary for monolingual and bilingual participants.

In monolinguals, all recall days were correlated highly significantly. Additionally, there was a significant correlation between recall Day 1 and semantic switching (Table 9.6). Also, letter and semantic verbal fluency were found to be significant correlates. Letter and semantic fluency correlated strongly and highly significantly with letter switching.

In bilinguals, all recall days were correlated strongly and highly significantly. Recall Day 2 and 7 correlated with semantic mean cluster size significantly and highly significantly, respectively. There was a highly significant correlation between letter and semantic fluency in English, and significant correlations with letter and semantic verbal fluency in Polish. English letter and semantic were found to be highly significant correlates with switching in English, and significant correlates with Polish semantic switching. Additionally, letter and semantic fluency in English correlated highly significantly with vocabulary, level of bilingualism (i.e., balance ratio), and frequency of English usage. There was also a highly significant correlation between switching in English letter and semantic fluency and level of bilingualism (i.e., balance ratio), as well as significant between English switching and vocabulary.

Together, these findings highlight that verbal fluency is significantly related to English proficiency in bilinguals. Bilingual verbal fluency and working memory are partially correlated (i.e., switching and visuospatial memory). However, bilingual executive control is not significantly related to neither working memory nor verbal fluency in both language groups studied.

Table 9. 5 Spearman’s correlations between verbal fluency in English, working memory, executive control (i.e., inhibition/attentional control), and vocabulary by group

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Letter Fluency		.543*	.438	-.001	.797**	.068	.797**	.068	.346	.087	.286	-.190	.239	
2. Semantic Fluency	.643**		.015	.027	.650**	.363	.650**	.363	.408	-.243	.181	-.072	.102	
3. Mean Cluster Size (Letter)	-.157	-.098		.280	.179	.082	.179	.082	-.107	.050	-.077	-.339	.122	
4. Mean Cluster Size (Semantic)	.186	.044	-.202		-.193	.203	-.193	.203	-.127	.164	.125	.350	.107	
5. Number of Cluster (Letter)	.947**	.647**	-.241	.211		.207	1.00**	.207	.432	-.147	.284	-.361	.184	
6. Number of Cluster (Semantic)	.463*	.608**	.077	-.326	.477*		.207	1.00**	.349	.100	.231	.088	-.022	
7. Switches (Letter)	.947**	.647**	-.241	.211	1.00**	.477*		.207	.432	-.147	.284	-.361	.184	
8. Switches (Semantic)	.463*	.608**	.077	-.326	.477*	1.00**	.477*		.349	.100	.231	.088	-.022	
9. Verbal Working Memory	-.035	.357	-.227	.242	-.015	.219	-.015	.219		.473	.443	-.150	-.018	
10. Visuospatial Working Memory	.084	.186	.096	-.315	-.003	.423*	-.003	.423*	.372		.296	.386	-.103	
11. Flanker	-.136	.030	.186	.039	-.154	-.097	-.154	-.097	.015	.162		-.006	-.745**	
12. Simon	.222	.091	.196	.343	.164	-.088	.164	-.088	.116	.155	-.149		.033	
13. Vocabulary	.695**	.689**	-.031	.061	.644**	.463*	.644**	.463*	.275	.195	-.261	.261		
14. Balance Ratio	.592**	.668**	-.220	-.042	.580**	.484**	.580**	.484**	.371*	.204	-.387*	.276	.846**	
15. English Usage	.419*	.576**	-.029	.025	.343	.412*	.343	.412*	.269	.383*	-.093	.475*	.704**	.717**

Note: Correlations for monolingual participants are above the diagonal and bilingual participants below the diagonal

Table 9. 6 Spearman's correlations between episodic memory (the paired associate learning task and verbal fluency), and vocabulary in monolinguals

	1	2	3	4	5	6	7	8	9	10	11
1. Recall Day 1											
2. Recall Day 2	.651**										
3. Recall Day 7	.666**	.868**									
4. Letter Fluency	-.065	.240	.003								
5. Semantic Fluency	.174	.191	.117	.543*							
6. Mean Cluster Size (Letter)	-.066	-.151	-.390	.438	.015						
7. Mean Cluster Size (Semantic)	-.121	.135	-.098	-.001	.027	.280					
8. Number of Cluster (Letter)	-.047	.074	-.143	.797**	.650**	.179	-.193				
9. Number of Cluster (Semantic)	.581*	.162	.054	.068	.363	.082	.203	.207			
10. Switches (Letter)	-.047	.074	-.143	.797**	.650**	.179	-.193	1.000**	.207		
11. Switches (Semantic)	.581*	.162	.054	.068	.363	.082	.203	.207	1.000**	.207	

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

Table 9. 7 Spearman's correlations between episodic memory (the paired associate learning task and verbal fluency) and vocabulary in bilinguals

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Recall Day 1																			
2. Recall Day 2	.755**																		
3. Recall Day 7	.797**	.838**																	
Verbal fluency in English																			
4. Letter fl.	.016	.012	-.096																
5. Semantic fl.	.179	.001	-.030	.643**															
6. MCS (Letter)	-.007	-.099	.034	-.157	-.098														
7. MCS (Semantic)	.110	.069	.075	.186	.044	-.202													
8. NC (Letter)	.094	.053	-.076	.947**	.647**	-.241	.211												
9. NC (Semantic)	.045	.108	-.014	.463*	.608**	.077	-.326	.477*											
10. S (Letter)	.094	.053	-.076	.947**	.647**	-.241	.211	1.00**	.477*										
11. S (Semantic)	.045	.108	-.014	.463*	.608**	.077	-.326	.477*	1.00**	.477*									
Verbal fluency in Polish																			
12. Letter fl.	-.198	-.216	-.353	.382*	.367*	-.132	.079	.432*	.230	.432*	.230								
13. Semantic fl.	.041	-.223	-.222	.434*	.666**	.093	-.160	.434*	.478*	.434*	.478*	.702**							
14. MCS (Letter)	.006	-.009	-.128	.307	.290	.050	-.088	.340	.357	.340	.357	.460*	.307						
15. MCS (Semantic)	-.188	-.445*	-.505**	.153	.222	.176	-.092	.200	.134	.200	.134	.549**	.580**	.328					
16. NC (Letter)	-.091	-.069	-.215	.224	.123	-.170	.203	.266	.131	.266	.131	.507**	.330	-.247	.115				
17. NC (Semantic)	.340	.237	.324	.379*	.447*	.032	-.006	.314	.226	.314	.226	.278	.478*	.343	-.164	.087			
18. S (Letter)	-.091	-.069	-.215	.224	.123	-.170	.203	.266	.131	.266	.131	.507**	.330	-.247	.115	1.00**	.087		
19. S (Semantic)	.340	.237	.324	.379*	.447*	.032	-.006	.314	.226	.314	.226	.278	.478*	.343	-.164	.087	1.00**	.087	
20. Balance Ratio	.210	.152	.084	.592**	.668**	-.220	-.042	.580**	.484**	.580**	.484**	-.043	.289	.230	-.066	-.088	.362	.088	.362

* p < .05 ** p < .01

Note: MCS = Mean Cluster Size, NC = Number of Cluster, S = Switches

9.5.2.6 Time-Course Analysis - The Timings of the Words Produced

9.5.2.6.1 Data Preparation

A time-course analysis and additional analyses were conducted for the bilingual group only due to the lack of recorded verbal fluency data for the monolingual group.

Verbal fluency data were recorded digitally in a media booth and processed using Audacity software. The researcher played the recordings and marked timestamp of every response produced with the accuracy of 1 millisecond. Based on the recorded timestamps, the responses were divided into twelve 5-sec bins for every participant and their 1-minute verbal fluency performance. The responses were then assigned a *serial number* stating the position of the word produced in the trial (e.g. if the first word produced was *dog*, it was assigned number 1, etc.); a *latency* which comprised a time interval from the start of the task until the onset of the verbal response (i.e., word *cat* produced 23.34 seconds from the start, word *cow* produced 45.34 seconds from the onset) and *bin number* indicative of the 5-sec bin to which the response was assigned (see Figure 9.2).

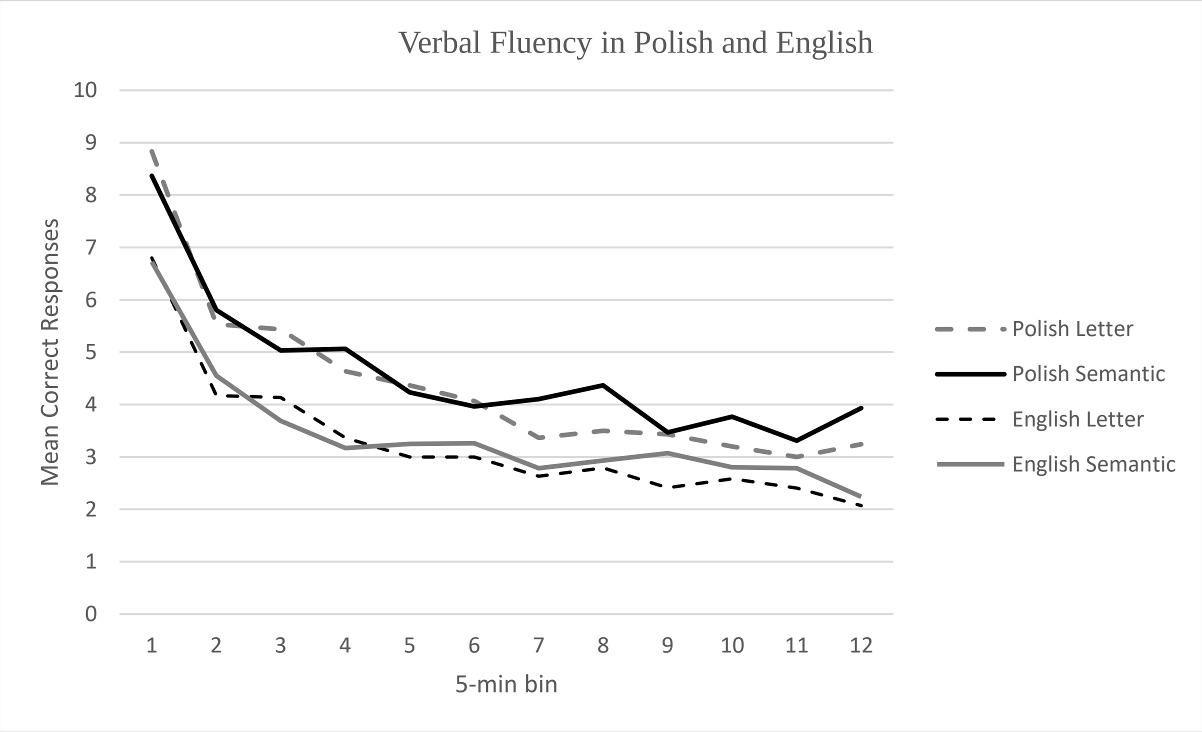


Figure 9. 2 Letter and semantic verbal fluency performance by language in bilinguals

For each task, two scores were calculated for each participant for the latency: (1) *first-response latency* – the time interval from the onset of a trial until the onset of the first response, and (2) *subsequent-response latency* which is the mean value of time intervals between the first response and each subsequent response. For each participant, the mean first-response latency was averaged across all letter trials (P, K, W for Polish and A, F, S for English) and across semantic trials (animals and things to eat) in order to provide an overall mean score for each task in each language. Additionally, the mean scores of the total number of correct responses for both letter and semantic fluency were computed separately, and these three means comprised the basis of the further time-course analysis.

The summary of mean latencies and standard deviations is presented in Table 9.8.

Table 9. 8 Mean (and standard deviation) first- and subsequent-response latencies for bilinguals

Task	First-response latency (sec) *Mean* (*SD*)	Subsequent-response latency (sec) *Mean* (*SD*)
Polish		
Letter	.65 (.34)	25.45 (2.41)
Semantic	.42 (.18)	26.26 (2.02)
English		
Letter	1.13 (1.01)	25.15 (2.57)
Semantic	.52 (.22)	25.06 (2.82)

9.5.2.6.2 Results

9.5.2.6.2.1 Exponential and Linear Functions

The mean number of correct responses across participants was then plotted as a function of twelve 5-second bins for each language and verbal fluency: Polish letter, Polish semantic, English letter, and English semantic, and presented in Figures 9.3. 9.4, 9.5, and 9.6.

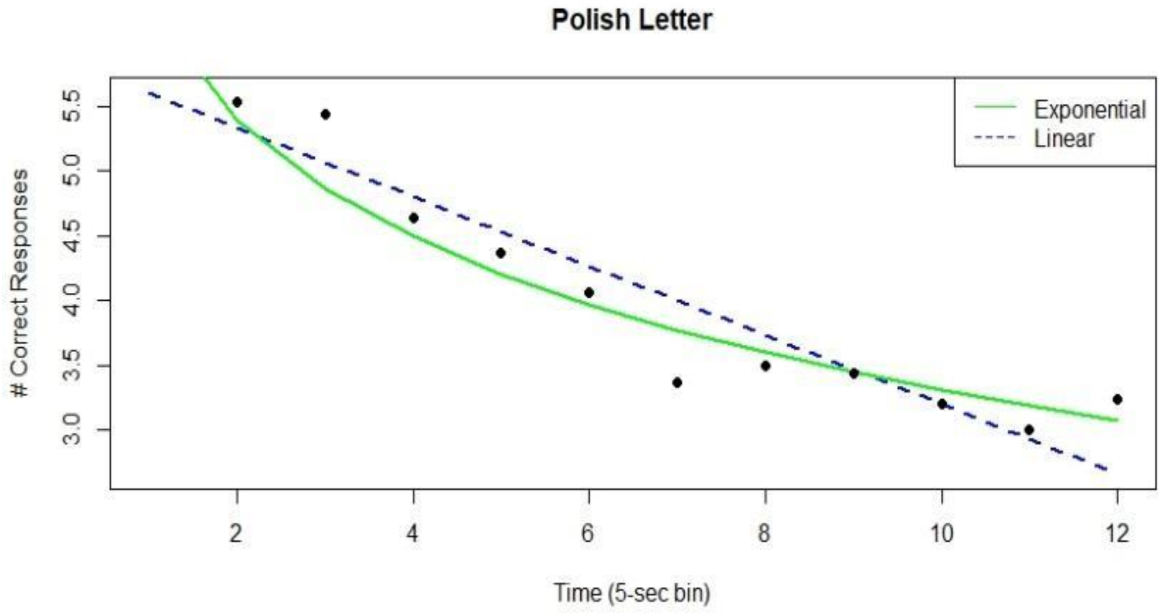


Figure 9. 3 Exponential and linear functions plotted for Polish letter fluency for bilinguals

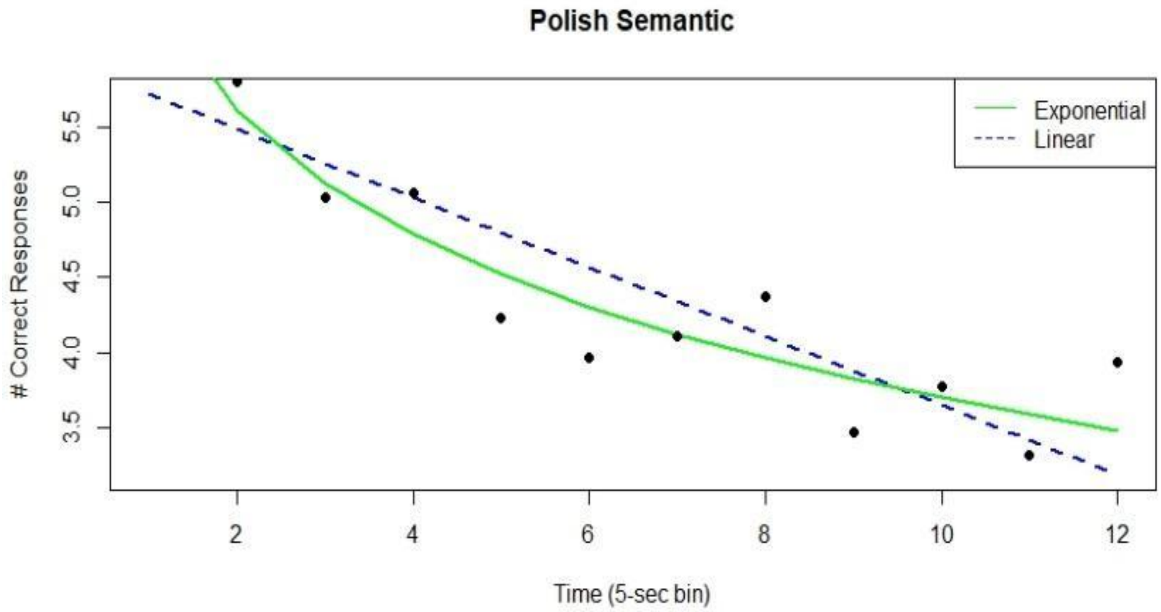


Figure 9. 4 Exponential and linear functions plotted for Polish semantic fluency for bilinguals

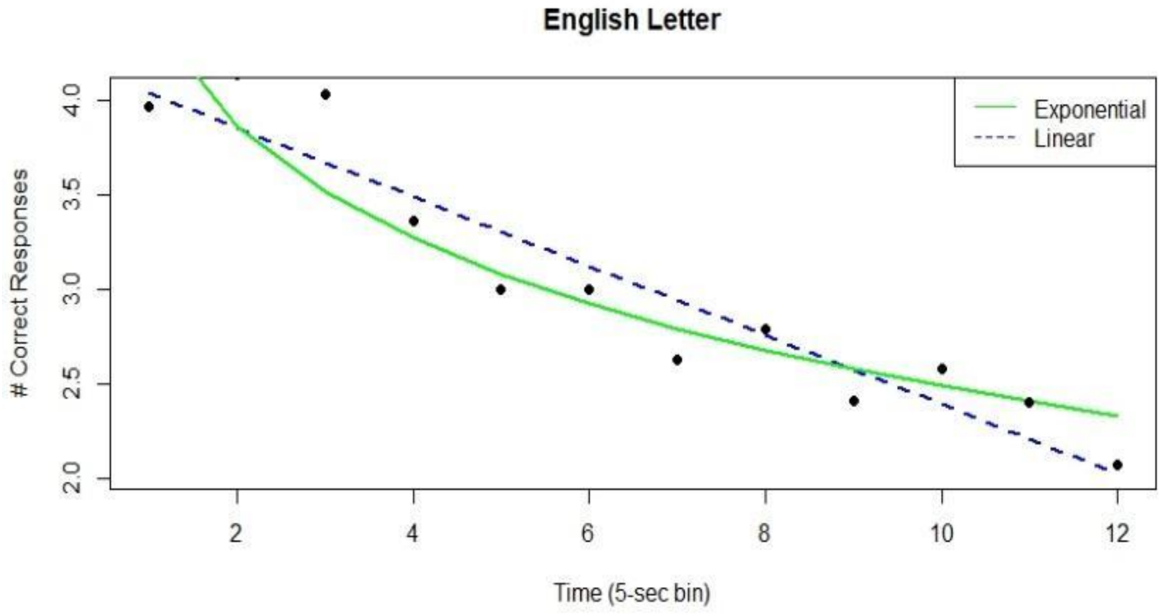


Figure 9. 5 Exponential and linear functions plotted for English letter fluency for bilinguals

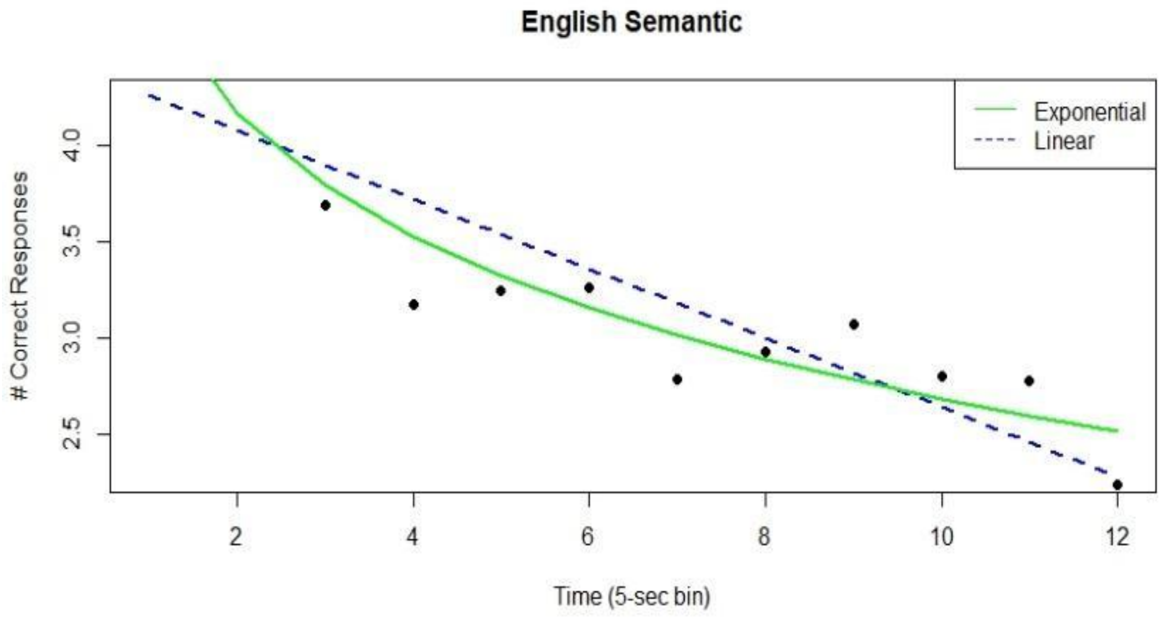


Figure 9. 6 Exponential and linear functions plotted for English semantic fluency for bilinguals

Research has shown that the number of words being produced by an individual in both verbal fluency declines as a function of time (Luo, Luk, & Bialystok, 2010). Verbal fluency usually begins with the best performance (i.e., the highest number of items produced), then the word retrieval declines unless an asymptote is reached assuming unlimited time. Wixted and Rohrer

(1993) proposed that the declining curve is exponential. Figures 9.3-9.6 present scatterplots with logarithmic and exponential functions for bilinguals for verbal fluency in both languages, Polish and English. As indicated in these figures, bilinguals retrieved more words at the beginning of both verbal fluency in Polish compared to English. No substantial differences were noted during a visual graph inspection. Table 9.9 provides the fitted functions for the mean scores for bilinguals with *y* representing the estimated value from the function at various time levels (*t*).

To identify the best fitting model that explains the rate of decline in lexical retrieval, exponential, and logarithmic models were plotted. Both functions have been used in verbal fluency performance in the past (Rohrer et al., 1995) and best describe the asymptotic. Best fitting functions for the time-course of verbal fluency output are presented in Table 9.9.

Table 9. 9 Exponential and linear functions for Polish and English verbal fluency in bilinguals

Task	Exponential		Logarithmic	
	Estimated function	R^2	*Estimated function*	R^2
Polish Letter	$y = (6.05\ )1.80e^{-.063t}$	.92	y = 6.28 - 1.29 ln(t)	.92
Polish Semantic	$y = (5.99)\ 1.79e^{-.050t}$	.79	y = 6.43 - 1.19 ln(t)	.91
English Letter	$y = (4.39)\ 1.48e^{-.060t}$	.92	y = 4.46 - .85 ln(t)	.86
English Semantic	$y = (4.53)\ 1.51e^{-.053t}$	.82	y = 4.80 - .92 ln(t)	.90

Note: Exponential and logarithmic functions are fitted with 12 observation means. Exponential [lm(log(y)~x)] and Logarithmic [lm(y~log(x))]

In this study, the fast-declining rate of recall can be best described by a logarithmic function as it accounts for a larger proportion of variance (Table 9.9). The function is presented above by plots with most items produced at the beginning leaving few items by the end of the time (Rohrer et al., 1995; Luo, Luk, & Bialystok, 2010). As presented in Table 9.9, logarithmic functions seem to account for a larger proportion of variance in the data (with an exception of English Letter, where exponential function accounts for 92% of variance, $R^2 = .92$).

As the functions presented in Table 9.9 are computed on the basis of 12 means, further steps were taken to gain more in-depth insight into within-subject variance and thus linear mixed effect modelling was employed. The entire model comprised 12 means (1 per each 5-second bin) multiplied by the number of participants ($N = 28$) which resulted in 336 observations. Here, the differences within each participant's performance over the course of the 1-minute task were of interest and the total number of words recalled in each bin was calculated.

Additional analysis of first and subsequent response latencies were conducted for bilingual participants. The time-course data to conduct these analyses were not available for monolingual participants.

9.5.2.6.2.2 First Response Latencies

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed a statistically significant main effects of language, $F(1, 27) = 6.69$, $p = .015$, $\eta_p^2 = .190$ and verbal fluency, $F(1, 27) = 10.93$, $p = .003$, $\eta_p^2 = .280$. For letter fluency, faster responses were recorded for Polish letter than English letter. For semantic fluency, participants demonstrated faster first-response latency in Polish than in English. No statistically significant interaction of language and verbal fluency was found, $F(1, 27) = 2.96$, $p = .096$, $\eta_p^2 = .100$ (Table 9.8).

9.5.2.6.2.3 Subsequent Response Latencies

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) repeated measures ANOVA revealed no statistically significant main effect of either language, $F(1, 27) = 3.98$, $p = .056$, $\eta_p^2 = .125$, or verbal fluency, $F(1, 27) = .58$, $p = .451$, $\eta_p^2 = .020$. No statistically significant interaction of language and verbal fluency was found, $F(1, 27) = 1.05$, $p = .315$, $\eta_p^2 = .036$ (see Table 9.8).

9.5.2.7 Slope Comparison Analysis

Further analyses were conducted to investigate the differences between slopes for verbal fluency in both languages of bilingual participants. These were run using an Analysis of Covariance (ANCOVA) with a dependent variable of the mean correct responses produced in each 5-second bin and a bin number introduced as a covariate. The interaction results between the fixed factor and the bin number comprised comparisons of slopes.

9.5.2.7.1 Within-Language Comparisons

9.5.2.7.1.1 Polish Letter and Polish Semantic

An ANCOVA with verbal fluency as the fixed factor and bin number as the covariate revealed a statistically significant main effect of bin number, $F(1, 20) = 41.21$, $p = .001$, $\eta_p^2 = .250$ and a nonsignificant main effect of verbal fluency, $F(1, 20) = .11$, $p = .745$, $\eta_p^2 = .102$. Verbal fluency and bin number interaction was not statistically significant suggesting that the slopes are homogenous, $F(1, 20) = .50$, $p = .487$, $\eta_p^2 = .202$.

9.5.2.7.1.2 English Letter and English Semantic

An ANCOVA with verbal fluency as the fixed factor and bin number as the covariate revealed a nonsignificant main effect of verbal fluency, $F(1, 20) = .02$, $p = .886$, $\eta_p^2 = .013$, and a statistically significant main effect of bin; $F(1, 20) = 37.55$, $p = .001$, $\eta_p^2 = .011$. Verbal fluency and bin number interaction was not statistically significant suggesting that the slopes are homogenous, $F(1, 20) = .73$, p = .402, $\eta_p^2 = .072$.

9.5.2.7.2 Between-Language Comparisons

9.5.2.7.2.1 Polish Letter vs English Letter

An ANCOVA with language as the fixed factor and bin number as the covariate revealed a statistically significant main effect of the bin number, $F(1, 20) = 45.06$, $p = .001$, $\eta_p^2 = .135$ and a nonsignificant main effect of language, $F(1, 20) = 5.02$, $p = .037$, $\eta_p^2 = .072$. Language and bin number interaction was not statistically significant suggesting that the slopes are homogenous, $F(1, 20) = .76$, $p = .473$, $\eta_p^2 = .201$.

9.5.2.7.2.2 Polish Semantic and English Semantic

An ANCOVA with language as the fixed factor and bin number as the covariate revealed a statistically significant main effect of language, $F(1, 20) = 4.41$, $p = .049$, $\eta_p^2 = .103$, and a significant main effect of bin number; $F(1, 20) = 33.61$, $p = .001$, $\eta_p^2 = .024$. Language and bin number interaction was not statistically significant suggesting that the slopes are homogenous, $F(1, 20) = .26$, $p = .617$, $\eta_p^2 = 0.17$.

9.5.2.8. Quartile Distribution Analysis

9.5.2.8.1 Data Preparation

Further analysis was conducted to investigate the differences between the number of words produced in the quartiles. The number of words in each quartile was averaged across all three letter conditions and two semantic conditions for each bilingual participant separately for Polish and English. This was then averaged across participants (see Figure 9.7 for Polish and 9.8 for English).

Table 9. 10 Mean correct responses by quartile for Polish and English verbal fluency in bilinguals

Quartile criteria	Polish		English	
	Letter	Semantic	Letter	Semantic
1st n(%)	7.1	10.4	5.33	7.95
2nd n(%)	4.18	6.48	3.12	4.63
3rd n(%)	3.51	5.73	2.67	4.4
4th n(%)	2.91	4.78	2.21	3.06

Note: Average across subcategories

The means were further divided into four quartiles with words produced in the first quartile (from 0 to 15 seconds), the second quartile (from 16 secs to 30), the third (30 to 45 secs) and the fourth quartile (45 to 60). The results are presented in Table 9.10.

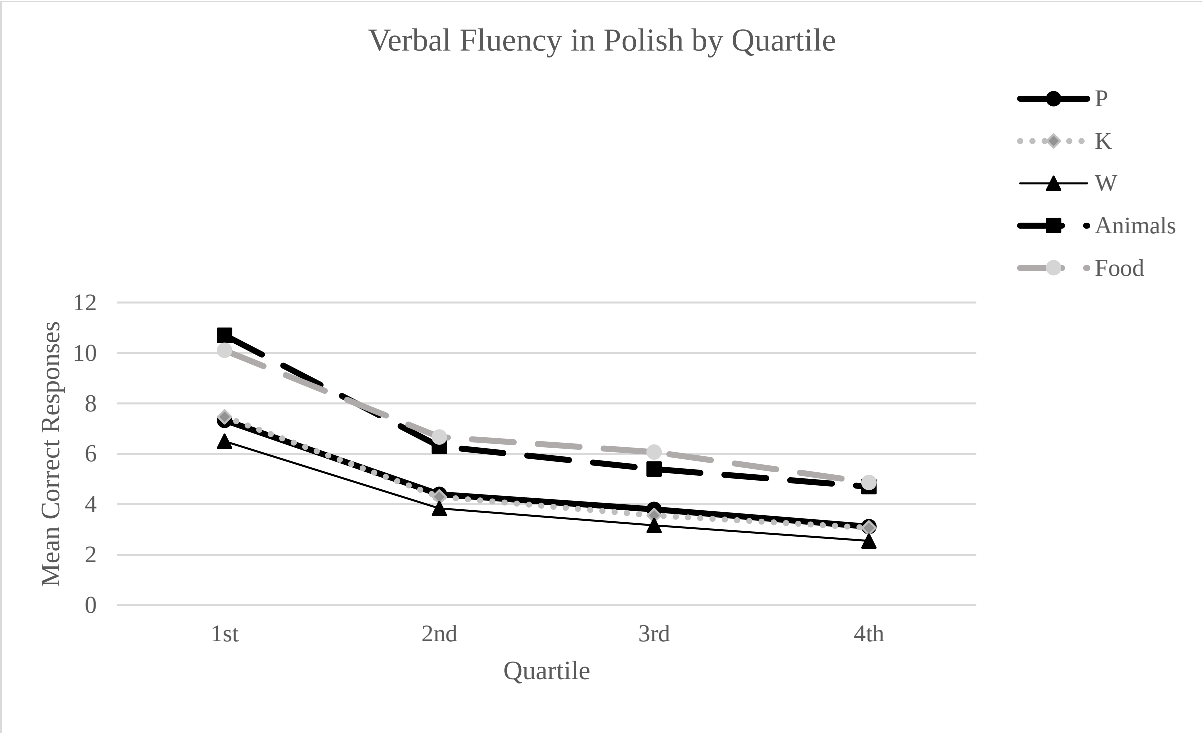


Figure 9. 7 Mean correct responses by quartile for letter and semantic verbal fluency in Polish

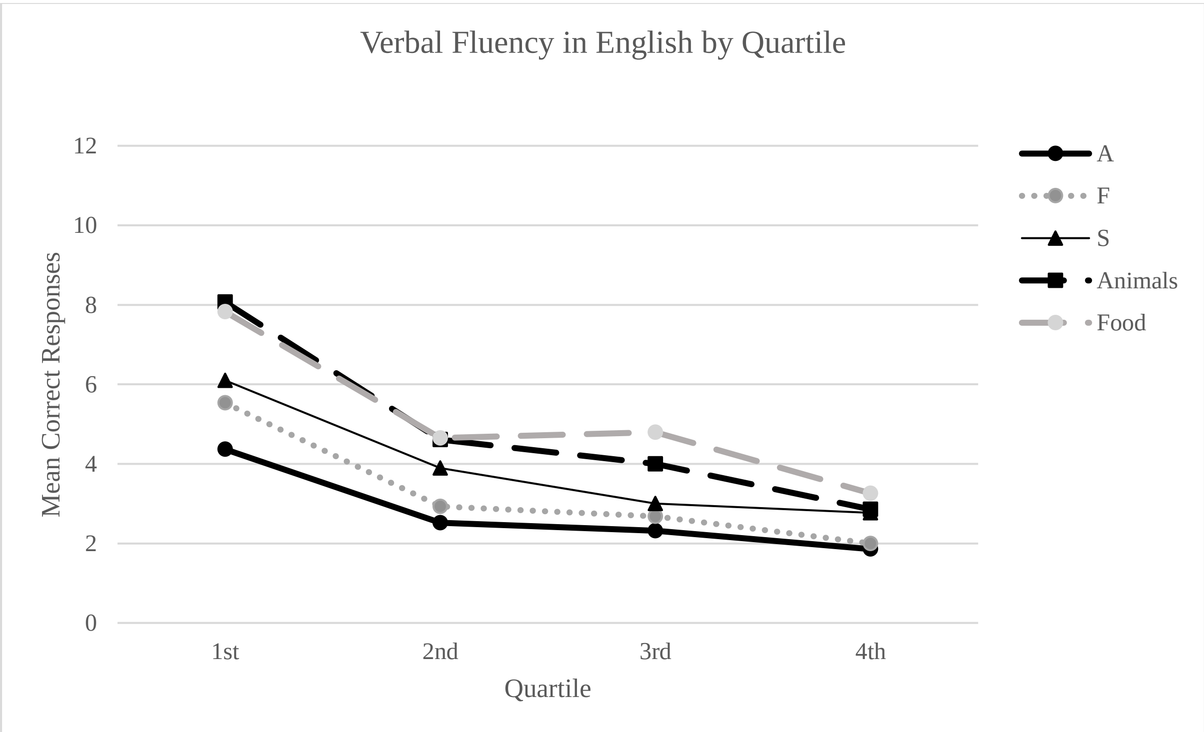


Figure 9. 8 Mean correct responses by quartile for letter and semantic verbal fluency in English

9.5.2.8.2 Results

A 2 (language: Polish and English) x 2 (verbal fluency: letter and semantic) x 4 (quartile: first, second, third, and fourth) repeated measures ANOVA revealed a statistically significant main effect of language, $F(1, 26) = 61.41$, $p = .001$, $\eta_p^2 = .601$, a statistically significant main effect of verbal fluency, $F(1, 26) = 293.93$, $p = .001$, $\eta_p^2 = .483$, and a statistically significant main effect of quartile, $F(3, 87) = 376.99$, $p = .001$, $\eta_p^2 = .309$. Additionally, the following interactions were found to be statistically significant: language and verbal fluency, $F(1, 26) = 5.28$, $p < .001$, $\eta_p^2 = .280$; language and quartile, $F(3, 87) = 3.82$, $p = .013$, $\eta_p^2 = .490$; and verbal fluency and quartile, $F(3, 87) = 9.41$, $p < .001$, $\eta_p^2 = .011$.

Further post-hoc analyses, after Bonferroni correction ($p = .050/8$, adjusted *p*-value level set to .006), revealed statistically significant differences between corresponding quartiles in letter and semantic verbal fluency between Polish and English in bilinguals with more words being generated in each quartile in participants' first language, Polish (see Table 9.10).

Similar results of significance, with Bonferroni correction of adjusted p-value at .006, were obtained for within-language comparisons for both Polish and English with significantly more words being generated in corresponding quartiles in semantic fluency as opposed to letter fluency (Table 9.11 and 9.12).

Table 9. 11 Mean (and standard deviation) with *t*- and *p*-values for cross-linguistic comparisons by quartile in bilinguals

Letter				
Quartile	Polish *Mean* (*SD*)	English *Mean* (*SD*)	*t*-value	*p*-value
1st	7.10 (1.46)	5.33 (1.51)	5.604	<.001
2nd	4.18 (1.13)	2.99 (1.05)	5.291	
3rd	3.51 (.99)	2.52 (1.14)	3.805	
4th	2.71 (1.03)	1.81 (.81)	4.658	

Semantic				
Quartile	Polish *Mean* (*SD*)	English *Mean* (*SD*)	*t*-value	*p*-value
1st	10.40 (2.11)	8.02 (2.07)	5.484	<.001
2nd	6.48 (1.68)	4.58 (1.50)	4.543	
3rd	5.73 (1.62)	4.37 (1.31)	4.373	
4th	4.78 (1.86)	2.90 (1.49)	4.664	

Table 9. 12 Mean (and Standard Deviation) with *t*- and *p*-values for within-language comparisons by quartile in bilinguals

Letter				
Quartile	Polish *Mean* (*SD*)	Polish *Mean* (*SD*)	*t*-value	*p*-value
1st	7.10 (1.46)	10.40 2.11)	-10.927	<.001
2nd	4.18 (1.13)	6.48 (1.68)	-9.060	
3rd	3.51 (.99)	5.73 (1.62)	-8.457	
4th	2.71 (1.03)	4.78 (1.86)	-6.140	

Semantic				
Quartile	English *Mean* (*SD*)	English *Mean* (*SD*)	*t*-value	*p*-value
1st	5.33 (1.51)	8.02 (2.07)	-6.95	<.001
2nd	2.99 (1.05)	4.58 (1.50)	-7.64	
3rd	2.52 (1.14)	4.37 (1.31)	-6.85	
4th	1.81 (.81)	2.90 (1.49)	-3.65	

9.6 Discussion

The aim of this study was to extend our understanding of whether there is a bilingual advantage in episodic memory as measured by verbal fluency. In this Experiment, letter and semantic verbal fluency were employed to investigate participants' word retrieval speed, their vocabulary knowledge, and executive control. Additionally, a suite of other measures (i.e., working memory and executive control) were examined as correlates of long-term memory (as measured by the paired associate learning task and verbal fluency).

As such, the following research hypotheses were made:

1) There is a bilingual advantage in episodic memory as measured by verbal fluency.
2) There is a relationship between the paired associate learning task, verbal fluency, working memory, executive control (i.e., inhibition/attentional control), and vocabulary.

Research hypothesis 1 was rejected as results revealed that on average monolingual participants produced significantly more words in both letter and semantic category in English than bilinguals. Bilinguals also produced significantly more words in both categories in their first language than in English.

In English verbal fluency analysis, monolinguals were found to have significantly bigger cluster size in both categories, which suggests that they produced more words in each cluster than bilinguals. Also, the number of clusters was significantly bigger for bilinguals indicating that bilinguals switched more often and created more but smaller clusters than monolinguals. This may be explained by the fact that bilinguals tend to have smaller vocabulary sizes in their second languages (Bialystok & Feng, 2008). As a result, they produce fewer lexical items as word retrieval is more effortful than in monolingual counterparts. Thus, the bigger number of clusters in bilinguals can be attributed to creating more smaller clusters with fewer words in each cluster when compared to monolinguals. More switches were found to be made in letter than semantic category in both language groups which is in line previous research findings suggesting that words generated in the former category are not semantically related and thus are more difficult to retrieve in a sequence (Luo et al., 2010, Katzev et al., 2013). This indeed leads to a greater number of switches made by participants. The fact that monolinguals demonstrated on average more switches than bilinguals may also stem from the better-established word and more accessible representations in their mental lexicon in English which is in fact their first language. As the word retrieval in the first language is less effortful than in the second language, monolinguals can be proposed to exhibit an advantage in word retrieval in phonemic fluency in their first language.

Bilinguals produced more words in their first language, Polish, than monolinguals in their first language, English. In first language verbal fluency, where monolinguals were tested in English and bilingual participants in Polish, the latter gave significantly more correct responses than the former. Additionally, their mean cluster size and the number of clusters were also significantly bigger in both categories which means that they were able to generate more words and clusters than monolinguals. What is more, the number of switches was also significantly larger for bilinguals than for monolinguals. These results suggest that bilinguals outperformed monolinguals when their first language performance was concerned as they produced more words and were more efficient at switching between clusters. As already discussed, this was not replicated in English.

These findings further support the idea that bilinguals tested in this Experiment were not balanced bilinguals, in other words, they were more proficient in their first (dominant) language which was Polish than in their second language – English. This is partially in line with previous verbal fluency research indicating that bilingual participants have smaller vocabulary sizes in both their languages (Bialystok & Luk, 2012) – this is true for English but not for Polish language. Such an unequal distribution may be accounted for by the relatively low English proficiency level of the bilingual sample as indicated by the level of bilingualism (i.e., balance ratio). Also, as their main language environment and main language of use was Polish, it can be assumed that their Polish vocabulary size was indeed larger than English. This might explain the superior accuracy performance on both verbal fluency in Polish and inferior performance in terms of vocabulary retrieval in English. Additionally, as in previous experiments in Chapter Seven and Eight, the low English proficiency of bilinguals included in this study may be accountable for the lack of presence of the bilingual advantage in episodic memory (as measured by verbal fluency). First language verbal fluency revealed that there is a disproportion between Polish and English proficiency in the bilingual sample and thus it would be interesting to recruit more heterogeneous bilingual samples in future research.

In terms of within-language differences in bilinguals, results revealed that they generated significantly more correct words in Polish (compared to English) and semantic (compared to letter) fluency, as well as their mean cluster size and the number of clusters was bigger for these two aspects. Switching, on the other hand, was again significantly bigger for Polish but for letter rather than semantic fluency. This means that bilinguals produced more clusters per se and more words within clusters in both semantic fluency as well as their first language. More switches were, however, made in letter fluency which is in line with previous research suggesting increased executive control demands on word production when retrieving words that are not semantically related (Luo et al., 2010, Katzev et al., 2013).

Additional cross-linguistic differences in verbal fluency for Polish and English in bilinguals were analysed by fitting letter and semantic fluency responses as a function of time. The time-course analysis with estimated exponential and logarithmic functions was presented in Table 9.9 with logarithmic functions found to account for the larger proportion of variance in bilingual data. Faster first response latencies were found for both letter and semantic verbal fluency in Polish compared to English which indicates faster and less effortful onset of word retrieval in participants' first language. Indeed, generating second language words took participants significantly longer. The difference was not replicated for subsequent mean latencies suggesting that there were no differences between letter and semantic fluency performance in Polish and English in this aspect.

This analysis also demonstrated that the number of words produced by participants declines exponentially as a function of time (Wixted & Rohrer, 1993). In other words, participants tend to generate the highest number of words at the beginning of the trial (within the first quartile) and this number declines as the time passes. This is explained by the fact that the more words participants retrieve, the more engaged executive control processes become. As proposed by Luo et al. (2010), executive control contribution to verbal fluency performance increases with the length of the trial as later in the task participants need to remember the words that they have already produced, and thus suppress interference from the earlier responses, as well as employ the ability to switch between word clusters. The slope analysis was conducted for bilinguals to examine the decay that occurs over time (i.e, and thus implicates increased use of controlled processes towards the end of the task as opposed to automatic processes at the beginning that include an initial burst of breadth of words available to an individual). Slope comparisons did not find any significant interactions between the number of the bin and verbal fluency within and between language comparisons indicating that the slopes were indeed homogeneous. Although bilinguals generated significantly more words at the onset of the trials in Polish, no difference between their further performance on verbal fluency in Polish and English was noted. Taken together, this indicates better performance at the start of the trial with similar slopes across verbal fluency in both languages, which indicates similar ability in this bilingual sample to sustain a constant word generation during the task.

Also, the quartile distribution analysis investigating the number of words generated in each one of four quartiles revealed that bilinguals produced significantly more words in every quartile in both categories in Polish when compared to English. This, once again, indicates better and more accurate performance in participants' first language suggesting an easier and less effortful access to Polish words within long-term memory mechanisms.

Research hypothesis 2 was partially rejected as no significant relationship was found between verbal fluency and paired associate learning, working memory, or executive control (i.e., inhibition/attentional control). However, there was a significant relationship between letter and semantic verbal fluency in English with vocabulary, level of bilingualism (i.e., balance ratio), and frequency of English usage.

The results obtained in this experiment may be due to a relatively small sample size. Moreover, the recruited bilingual sample can be characterised as of relatively low English proficiency as indicated by the level of bilingualism (i.e., balance ratio). Future studies should aim to recruit a well-defined group of more proficient bilinguals and ensure that the level of bilingualism (i.e., balance ratio) is not significantly different from a value that might be taken to indicate perfect balance ratio (i.e., a ratio of 1.0). Also, it would be of great interest to compare both language groups on the time-course analysis. This was not possible in this study due to the lack of recordings of verbal fluency performance for monolinguals.

CHAPTER TEN

General Discussion

10.1 Introduction

This thesis examined whether there is a bilingual advantage in working memory, executive control (i.e., inhibition/attentional control), and episodic memory (as measured by the paired associate learning task and verbal fluency tasks). Its main purpose was to extend the current understanding of the bilingual advantage by investigating not only executive control and working memory, but also long-term memory. Additionally, the thesis examined the role of working memory, executive control (i.e., inhibition/attentional control), and English vocabulary knowledge in bilingual episodic memory performance. The summary of main experimental findings will be presented and strengths and limitations of the thesis discussed in this Chapter. This Chapter will also contextualise the key findings within the field of bilingual advantage research.

10.2 Summary of Experiments

10.2.1 Is there a Bilingual Advantage in Working Memory and Executive Control (i.e., Inhibition/Attentional Control)?

Bilingual advantage in working memory and executive control (i.e., inhibition/attentional control), and the relationship between them was investigated in Experiment 1 in Chapter Seven. Seventeen monolinguals and 28 bilingual adults were recruited and tested on a suite of measures tapping executive control (i.e., the Flanker task and the Simon task) and working memory (i.e., Digit Recall, Backward Digit Recall, Dot Matrix, and Mister X).

Results revealed no bilingual advantage in working memory which is in line with studies indicating similar performance between monolinguals and bilinguals (Martin-Rhee & Bialystok, 2008; Luo, Luk, & Bialystok, 2010; Engel de Abreu, 2011) and bilingual disadvantage (Ratiu & Azuma, 2015). Interestingly, results obtained in the systematic meta-analytic review in Chapter Four indicate that there is a bilingual advantage in working memory, however, this is based on a very limited number of studies included in the analyses. In the context of bilingual advantage in working memory, it is worth mentioning that verbal short-term and verbal working memory assessments in Experiment 1 were conducted in English which might explain better monolingual performance. Additionally, as a significant relationship between bilingual performance on verbal working memory measure and their level of bilingualism (i.e., balance ratio) was found, it may be suggested that bilinguals with higher English proficiency performed better at Digit Recall and Backward Digit Recall. This provides additional evidence towards the claim that bilingual advantage in working memory was not evident in this experiment due to low English proficiency levels of the bilingual sample.

In terms of executive control, bilinguals were significantly faster than monolinguals to resolve conflicts in the Simon task which is in line with previous research findings (Bialystok et al., 2004; Bialystok, Craik, & Ryan, 2006; Bialystok, 2017). This suggests that bilingual participants demonstrated more efficient ways of conflict resolution than monolinguals. In the Flanker task, however, no bilingual advantage was found. Although there was a smaller Flanker effect in bilinguals, the language-group difference was not statistically significant. Similarly to results obtained on working memory measures, the lack of bilingual advantage in executive control may come as a result of low English proficiency levels within the bilingual sample. Indeed, correlational analyses revealed a highly significant and a strong relationship between the Flanker effect and level of bilingualism (i.e., balance ratio) as well as frequency of English usage (*rho* = .704 and *rho* = .846, respectively). This may account for no bilingual advantage in executive control being detected in Experiment 1. This is further supported by the fact that the score obtained for the balance ratio of the bilingual sample differed significantly from a value that might be taken to indicate perfect bilingual balance. As proposed by previous research findings, the Flanker effect may fail to manifest itself in less proficient bilinguals whose executive control mechanisms engaged in language control (i.e., efficient managing of two languages in terms of activation of target language and inhibiting of non-target language, see Hilchey & Klein, 2011) might be not enhanced.

10.2.2 Is there a Bilingual Advantage in Episodic Memory as measured by the Paired Associate Learning Task? Is the Bilingual Advantage in Episodic Memory Task-Dependent (i.e., how the Paired Associate Learning is assessed: Recall versus Recognition Measures)?

The impact of bilingualism on long-term memory mechanisms (i.e., episodic memory) was assessed by employing the paired associate learning task in Experiment 2 in Chapter Eight. For the purpose of this Experiment, a bespoke paired associate learning task was developed and piloted to ensure that it is a reliable tool to investigate episodic memory performance (see Chapter Six). Sixteen monolingual and 28 bilingual adults were tested on the paired associate learning task. Additional executive control, working memory, and vocabulary measures were included in correlational analyses to capture the interplay between the indices.

Results revealed no bilingual advantage in episodic memory as measured by the paired associate learning task (i.e., recall and recognition) as both language groups scored similarly on all three test days. This is inconsistent with the results obtained in the systematic review which suggest that there is a bilingual advantage in episodic memory as measured by paired associate learning tasks. In fact, in Experiment 2, bilingual performance on recall accuracy was worse than monolingual which might be linked to the fact that Polish, their first language, comprises different sound patterns as the ones present in English. Moreover, as indicated by previous research (Storkel & Maekawa, 2005; Majerus, Poncelet, Greffe, & Van der Linden, 2008), second language novel words which are phonologically similar to participants' first language are learned more easily and efficiently than words that differ in this aspect. Taking into consideration the fact that novel words in the paired associate learning task were based on modifications to real English words, monolingual participants might have found it easier to repeat and recall more words correctly.

In terms of bilingual advantage in episodic memory being task-dependent, Experiment 2 revealed that this might be true as correlational analyses showed that performance on recall and recognition were moderately and highly significantly related in bilinguals. This is further corroborated by systematic review results in Chapter Four that indicate a task-dependent bilingual advantage with different effect sizes observed for recall and recognition as measures of episodic memory. This is also consistent with previous findings in terms of the discrepancies between recall and recognition as they rely on different mechanisms and there are substantial differences between them in terms of specific processes they employ to complete the task (Reynolds & Romano, 2016). Recall is more effortful and requires a higher depth of processing that relies on semantic and phonological information access, retrieval, and production, whereas recognition relies mainly on information identification often supported by an external cue.

10.2.3 What is the Role of Working Memory and Executive Control (i.e., Inhibition/Attentional Control) in Bilingual Episodic Memory Performance, as measured by the Paired Associate Learning Task?

The research hypothesis of the role of working memory in bilingual episodic memory performance was partially supported in Experiment 2. Correlational analyses revealed that although bilingual episodic memory (i.e., recognition) and working memory were found to be related, this result was not replicated for recall. Moreover, bilingual executive control was not significantly related to neither working memory nor episodic memory in both language groups studied. Based on the correlational results, it was found that bilingual executive control performance was constrained by their level of bilingualism (i.e., as measured by balance ratio) and frequency of English use. This is to some extent inconsistent with meta-analytic results in Chapter Four obtained for working memory and executive control (i.e., inhibition/attentional control) which suggest their significant role in episodic memory performance in bilinguals.

10.2.4 What is the Role of Bilinguals' Vocabulary Knowledge (in English) in Bilingual Episodic Memory Performance?

The role of bilinguals' vocabulary knowledge (in English) in bilingual episodic memory was measured by verbal fluency in Experiment 3 in Chapter Nine. In this Experiment, letter and semantic verbal fluency were employed to investigate participants' word retrieval speed, their vocabulary knowledge, and executive control. Additionally, a suite of other measures (i.e., working memory and executive control) were examined as correlates of long-term memory (as measured by the paired associate learning task and verbal fluency). Nineteen monolinguals and 28 bilingual adults were tested on letter and semantic verbal fluency.

Results revealed no significant role of English vocabulary in episodic memory performance in both language groups. This result is not consistent with the systematic review findings (Chapter Four) where a statistically significant large effect size suggesting that vocabulary plays a significant role in bilingual episodic memory performance was found. Although, it needs to be noted that only five papers were included in this meta-analysis (Chapter Four) and results revealed substantial heterogeneity between the included studies.

10.3 Contextualisation of the Key Findings

This study sought to investigate the role of bilingualism in long-term memory mechanisms, specifically episodic memory as measured by paired associate learning and verbal fluency. The aim of the following section is to contextualise the above presented answers to key research questions within results obtained by previous bilingual studies that are relevant to the project.

This study did not replicate bilingual advantage in working memory which was indicated by some previous research (Ratiu & Azuma, 2015; Anton, Carreiras, & Dunabeitia, 2019).

However, a significant relationship was found between bilingual working memory and episodic memory which indicates that there is a shared feature that is common across these two types of memory. This is in line with Miyake and Friedman's (2012) unity and diversity framework explaining commonalities and disparities between memory components.

Interestingly, monolingual advantage was found for verbal short-term memory performance as measured by Digit Recall (i.e., recalling a sequence of numbers) in Experiment 1 which is consistent with bilingual research that points to similar between language-group performance (Martin-Rhee & Bialystok, 2008; Luo, Luk, & Bialystok, 2010; Engel de Abreu, 2011) or monolingual advantage (Ratiu & Azuma, 2015) (see Chapter Three). Interestingly, the proposal that language may impact episodic memory performance (Loftus & Palmer, 1974) was replicated in Experiment 2. In line with previous findings (Fernandes, Craik, Bialystok, & Kreuger, 2007), bilinguals tend to perform worse than monolinguals on tasks that require explicit recall of words. This is due to smaller vocabularies in any one of their languages (or both) when compared to monolinguals which has been claimed to result in poorer lexical access and worse memory retrieval (Fernandes et al.). For this reason, bilinguals have been found to be at a disadvantage when decoding and encoding purely linguistic information. Indeed, in Experiment 2, bilinguals demonstrated worse performance on recall across Day 2 and Day 7 when compared to monolinguals.

It needs to be noted that verbal short-term and verbal working memory assessments used in this experiment employed stimuli comprising English numbers (i.e., monolingual participants' first language) and thus might explain English monolinguals' better performance on these tasks. At the same time, it may also account for the lack of bilingual advantage found for the bilingual sample. Moreover, as suggested by research (Storkel & Maekawa, 2005; Majerus, Poncelet, Greffe, & Van der Linden, 2008), the process of learning or recalling second language novel words that are phonologically similar to individual's first language is less effortful and more efficient than that of words that differ in this aspect. In the context of the paired associate learning task employed in Experiment 2, this argument may arguably account for the differences between language groups with bilinguals performing worse on recall. Novel words employed in the paired associate learning task were based on modified existing English words (see Chapter Six), and therefore they might have been easier and less effortful for English monolinguals to retrieve and/or recall correctly.

On the other hand, bilinguals have been observed to demonstrate superior non-linguistic recall (Bialystok, Craik, Klein, & Viswanathan, 2004). In Experiment 2, bilinguals were more accurate at recognition than monolinguals. Moreover, there was a strong and highly significant relationship between visuospatial working memory and recognition in bilinguals. This corroborates findings by Schroeder and Marian (2012) who employed a linguistically-reduced version of picture recall task. Results of their study revealed that bilinguals performed better at picture recall than their monolingual counterparts with factors such as early acquisition and more extensive bilingual experience associated with better episodic memory performance. In Experiment 3, a weak but significant relationship was revealed between visuospatial working memory and the frequency of English usage (although not English vocabulary as measured by WASI or level of bilingualism as defined by the balance ratio) which indicates that more proficient bilinguals demonstrate superior performance on recognition tasks where there is lower linguistic load than monolinguals. This is in line with results obtained by Ratiu and Azuma (2015) who found that working memory is not related to bilingual status (i.e., balance ratio). It can be proposed that the more often bilinguals use their second language, the better their performance on visuospatial memory. This suggests that daily language switching can contribute to an enhanced recognition ability. There was no relationship between working memory and executive control as indicated by some previous research (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010) which found a very strong correlation between working memory and executive control (r = .97) in a large sample of 200 adults.

This study did not find a bilingual advantage in executive control. Bilingual performance on the Simon task aligns with the Bilingual Inhibitory Control Advantage proposed by Hilchey and Klein (2011) who state that bilinguals demonstrate more efficient inhibitory control when compared to monolinguals. This is arguably due to the fact that they have to manage language interference from a non-target language on a daily basis which results in an enhanced ability to resolve conflict. Additionally, this framework also proposes that bilinguals have faster reaction times in incongruent trials which demonstrates their better ability to suppress interference and thus indicates enhanced inhibitory control. In the context of Experiment 1, this was replicated in the Simon task where bilingual participants were faster than monolinguals in incongruent trials, however, this difference was not statistically significant.

Moreover, evidence that would further support bilingual advantage in inhibitory control would mean that both target measures tapping inhibition/attentional control, i.e., the Simon task and the Flanker task in this study, are significant correlates if they indeed share common features and both tap an underlying skill associated with inhibitory control. Experiment 1 did not reveal a significant relationship between the Simon task and the Flanker task. This finding is however consistent with results obtained by Stins et al. (2005) who observed small and nonsignificant correlations between Simon, Flanker, and Stroop when testing 12-year-olds. Also, Kousaie and Phillips (2012) and Keye, Wilhelm, Oberauer, and van Ravenzwaaij (2009) found no associations between the Simon task and the Flanker task which implies that both tasks may indeed tap different executive control aspects. This finding is in line with the unity and diversity framework distinguishing commonalities and disparities within executive control (Miyake & Friedman, 2012).

On the other hand, Bialystok et al. (2014) explains that bilingual advantage may be associated with enhanced ability to monitor attention as an answer to fast changing task demands rather than inhibitory control itself. Along these lines, Costa et al. (2009) also suggests that bilinguals may be more efficient at monitoring attention and thus better at dealing with mixed congruent/incongruent trials. This means that they demonstrate advantage in moving between conflicting and conflict-free trials. As bilinguals need constant monitoring of their languages, this may result in their enhanced conflict monitoring skills.

As indicated by correlational analyses, there were significant relationships for both the Simon task and the Flanker task and variables associated with the level of bilingualism and the frequency of second language usage (i.e., the balance ratio and English usage, respectively). Varying second language proficiency levels within the bilingual sample may led to difficulties in capturing bilingual advantage in executive control. According to Martin-Rhee and Bialystok (2008), the enhanced ability to selectively attend to target cues in situations involving conflict resolution is a characteristic specific to highly proficient bilinguals. In fact, such enhancement may not be evident in lower proficiency bilingual populations. This is consistent with results obtained in Experiment 1. Although the bilingual sample was not highly proficient, a bilingual advantage was found in the Simon effect. Therefore, an alternative explanation is that both measures, the Simon task and the Flanker task, tap different underlying executive control components. As proposed by Paap and Greenberg (2013), numerous tasks tapping the same executive control skill should be utilised in studies examining bilingual advantage to ensure that it is captured more comprehensively.

No bilingual advantage in episodic memory as measured by paired associate learning was found in Experiment 2. As discussed above from the perspective of memory, bilingual participants demonstrated worse recall and better recognition ability which is consistent with previous research findings (Fernandes, Craik, Bialystok, & Kreuger, 2007). There was a moderate and highly significant relationship between recall and recognition performance in bilinguals, thus indicating some shared underlying ability between these two memory variables.

Remarkably, monolinguals performed significantly worse in the Training block in which their task was to learn novel word - novel object mappings. Bilingual demonstrated an advantage in correct mappings, i.e., establishing novel associations between novel words and objects. According to Zerr et al. (2018), participants more efficient at learning are also faster in terms of their processing speed and demonstrate enhanced memory performance among others. A crucial mechanism that contributes to one' ability to learn efficiently is inhibition/attentional control. This means that effective learners are more efficient at allocating attention during learning. Therefore, such individuals are less affected by interference as they have an enhanced ability to allocate attention to task-relevant information (Unsworth & Spillers, 2010).

Analysis of errors made by participants during recall in Experiment 2 also yielded interesting results. Results showed that overall bilinguals made significantly more omissions and significantly fewer extra- and intra-dimensional errors when recalling novel words. On the other hand, monolingual participants recalled novel word forms more confidently even if they were not sure whether the words produced are correct. Bilinguals seemed to refrain from recalling novel words instead of trying to produce them. One possible explanation to this may be that bilinguals simply lacked confidence to recall novel words and preferred to omit the trials. Another possibility is that they did not have any established phonological representations linked to the novel words as opposed to monolinguals in whom novel words might have triggered activation of similarly sounding real words. As previously mentioned, the novel words were modifications of existing English words and therefore it might have been arguably easier for monolinguals to recall the targets or produce similarly-sounding even though at times incorrect responses.

This study did not reveal the bilingual advantage in episodic memory as measured by verbal fluency. This is contrary to previous findings by Luo et al. (2010) and Friesen et al. (2015) who found the bilingual advantage in letter verbal fluency in bilinguals. It needs to be noted however that their bilingual sample was highly proficient in English. This experiment partially replicates Sandoval et al.' (201) study with bilinguals generating fewer words in letter fluency than monolinguals. They also report longer subsequent response times in bilinguals, but this comparison could not be made in Experiment 3 due to the lack of audio-recorded data for the monolingual sample. Also, it needs to be reminded that target samples differed in terms of their performance on the WASI Vocabulary Subtest with monolinguals scoring significantly better.

Bilinguals demonstrated better performance in Polish in which they produced more words in letter and semantic verbal fluency. Interestingly, they produced more words in their first language than monolinguals in theirs. Bilingual performance in English verbal fluency was however worse than monolingual word production. This finding is partially consistent with previous findings indicating that bilinguals tend to have smaller vocabulary sizes in both their languages (Bialystok & Luk, 2012). However, in Experiment 3, they generated more words in Polish than monolinguals in English which is not in line with the vocabulary size proposal. Such an unequal between language performance in verbal fluency may again be linked to second language proficiency of the bilingual sample. Indeed, there was a strong and significant relationship for both letter and semantic verbal fluency in English and vocabulary (as measured by WASI Vocabulary Subtest), level of bilingualism (i.e., balance ratio), and frequency of English use. This suggests that bilinguals exhibited stronger ability in Polish and weaker ability in English which is further corroborated by results obtained from the time-course analysis including response latencies, slope, and quartile distribution comparisons. Additionally, graphical representations of word retrieval for Polish and English clearly show discrepancies in bilingual verbal fluency performance. As reported in language background questionnaires, Polish was indicated to be their main language of use and main language within participants' daily environment. Thus, it can be assumed that their Polish vocabulary size was larger than English. Lower English proficiency levels (i.e., unequal balance ratio) may account for the superior accuracy performance on verbal fluency in Polish and inferior performance in terms of vocabulary retrieval in English.

To further investigate between-language group differences in English, clustering and switching ability was also examined. Monolinguals demonstrated significantly bigger cluster sizes in both letter and semantic verbal fluency which means that they generated more words in each cluster than bilinguals. The number of clusters was also significantly bigger for bilinguals which indicates that they made more switches and thus produced more but smaller clusters than monolinguals. As discussed earlier, due to lower English proficiency of the sample, it can be assumed that they had a smaller vocabulary size for their second language (Bialystok & Feng, 2008). As a result, they were able to generate fewer words as word retrieval was more effortful than for monolinguals and the cross-linguistic interference was more difficult to suppress than it usually is for highly proficient bilinguals. This implies more smaller clusters with fewer words in each cluster when compared to monolinguals. Additionally, more switches were observed in letter than semantic verbal fluency in both language groups which is consistent with research findings. Generating words in letter verbal fluency are not semantically related and therefore it is more difficult to retrieve them in a sequence (Luo et al., 2010, Katzev et al., 2013) which results in a greater number of switches made by participants. In Experiment 3, monolinguals made on average more switches than bilinguals which may be due to well-established and more accessible word representations in their mental lexicon in English which is in fact their first language. As a matter of fact, word retrieval in the first language is claimed to be less effortful than in the second language (Majerus, Poncelet, Greffe, & Van der Linden, 2008),.

In terms of within-language differences in bilinguals, results revealed that they generated significantly more correct words in Polish (compared to English) and semantic (compared to letter) fluency which is consistent with previous studies (Fisk & Sharp, 2004). As semantic retrieval is largely an automatic and an overlearned process as opposed to more effortful letter verbal fluency (Luo et al., 2010), one can expect to see more words being generated in the former. Similarly to monolinguals, more switching was observed in letter fluency which is in line with previous research suggesting increased executive control demands on word production when retrieving words that are not semantically related (Luo et al., 2010, Katzev et al., 2013).

As mentioned earlier in the context of language disproportions exhibited by the bilingual sample, additional between language differences in verbal fluency were examined by fitting letter and semantic fluency responses as a function of time. The time-course analysis began with calculating estimated exponential and logarithmic functions with the latter found to account for the larger proportion of variance in bilingual data (i.e., 92%, $R^2 = .92$). This type of analysis allows to present how the number of words retrieved by participants decline exponentially as a function of time (Wixted & Rohrer, 1993). This was replicated for the bilingual sample in Experiment 3. As already discussed, previous research findings (Perret, 1974; Luo et al., 2010) propose that participants tend to produce the highest number of words at the beginning of the trial (within its first quartile – i.e., from the onset to the time point of 15 seconds) and this number declines as the trial progresses. Luo et al. (2010) explains that executive control contribution to verbal fluency performance increases with the length of the trial as later in the task participants need to remember the words that they have already produced, and thus suppress interference from the earlier responses, as well as employ the ability to switch between word clusters.

Faster first response latencies were found for Polish letter and semantic verbal fluency compared to English which suggests faster and easier onset of word retrieval in participants' first language. First response timing refers to the time interval between the beginning of the task and the onset of the first response and therefore indicates the amount of time needed for an individual to retrieve and produce the first word. For English, producing the first response took bilinguals significantly longer. No significant difference was observed for subsequent mean latencies therefore indicating that there are no differences between letter and semantic word retrieval of subsequent responses in Polish and English.

Additional quartile distribution analysis revealed that bilinguals produced significantly more words in every quartile (i.e., 15-sec bin) in Polish letter and semantic verbal fluency than in English. This provides further evidence towards superior and more accurate performance in participants' first language than in their second language. This suggests that bilingual access to Polish words within long-term memory mechanisms was less effortful.

Although only highly proficient Polish-English bilinguals were sought to participate in this experiment, their proficiency levels were based on self-reported language proficiency which, at times, were later found to be not representative of an actual proficiency in the second language. As revealed by correlational analyses, bilingual verbal working memory was weakly and significantly related to the level of bilingualism (i.e., balance ratio) and visuospatial memory weakly and significantly related to the frequency of English usage. This indicated that working memory performance is closely related to participants' English proficiency and frequency of using this language. Moreover, there was a weak and significant relationship between the Flanker effect and the level of bilingualism (i.e., balance ratio) whereas a moderate and significant relationship between the Simon effect and the frequency of using English. This again suggests that proficiency in English impacted participants' performance on executive control (i.e., inhibition/attentional control) measures. Correlational analyses indicated that the lack of bilingual advantage on certain tasks as well as worse bilingual performance is linked to low English proficiency (measured by WASI Vocabulary Subtest), level of bilingualism (i.e., balance ratio), and frequency of English use. Indeed, the low English proficiency level of the bilingual sample may be accountable for the lack of bilingual advantage in episodic memory. Therefore, future research into the way bilingualism enhances episodic memory and the role that working memory, executive control, and vocabulary play in this process is much needed.

Overall, one of the key findings revealed by this thesis is selecting and securing an appropriate bilingual sample. By appropriate, the author means a sample of bilingual participants that demonstrate high proficiency in the second language (as measured by standardised assessments such as WASI Vocabulary Subtest), and a balanced level of bilingualism (as indicated by the balance ratio) that does not differ significantly from the value that might be taken to indicate perfect bilingual balance (i.e., a ratio of 1.0). Additionally, it needs to be ensured that bilinguals demonstrate a balanced frequency of using both languages to ensure they have to manage two linguistic systems equally often. Therefore, it is crucial to consider different ways of conceptualising bilingualism and thus its definitions that are often used in research (see Chapter Three). Some other important factors to consider that are often mentioned in the literature are: age of acquisition, patterns of languages used at home, as well as language environment among others.

Taken together, future studies should aim to recruit a well-defined group of highly proficient bilinguals and ensure that their level of bilingualism and frequency of language use do not differ significantly between the languages. This might increase the possibility of observing a bilingual advantage across memory and executive control as suggested by previous research (Hilchey & Klein, 2011). Also, for the paired associate learning task, future analyses into phonological features of words recalled could be conducted to investigate the differences in language processing between monolinguals and bilinguals. Fine analysis of words produced was not conducted in this study but it would be interesting to see where error differences between language groups stem from. Such analysis might inform the current state of knowledge regarding differences in phonological processing of novel word forms in various language groups and examine whether such differences in performance are language specific.

Additionally, to extend findings of this thesis, it would be of great interest to compare both language groups recruited in these experiments on the time-course analysis. This was not possible in this study due to the lack of recordings of verbal fluency performance for monolinguals.

10.4 Strengths and Limitations

The following section will summarise the strengths and limitations of this research project from the perspective of what factors should be taken into consideration when planning similar studies in the future.

The main strength of the current study is that it investigates a phenomenon that is clearly under-researched as indicated by the systematic review of studies within the field of bilingual research. There is a very limited number of studies that examine the way bilingualism impacts long-term memory mechanisms by means of paired associate learning tasks. This study aimed to extend the understanding of whether there is a bilingual advantage in episodic memory.

Another strength of this study is the fact that it employed a systematic review and a meta-analytic approach to investigate the current state of research with regards to the bilingual advantage in working memory, executive control, and episodic memory. As such, it examined 22 studies to address the main research questions by providing a synthesis of the findings.

The study is innovative in a sense that it presents the development and piloting of a bespoke paired associate learning task that was employed to investigate episodic memory performance. Phase 4 and Phase 5 in Chapter Six present results from the pilot studies that were conducted to ensure that the task is a reliable measure of participants' episodic memory. A detailed description of the stimuli used and modifications made to the initial task design are presented in detail to allow replication.

Additionally, a single population of Polish-English bilingual adults was recruited to participate in this project. The selection of this particular language group is supported by the fact that Polish constitutes the main language other than English that is widely spoken in England. Moreover, it is an under-represented language in bilingualism and cognition literature and thus research with Polish-English bilinguals is of great importance. This procedure also aimed to minimise variability associated with cultural diversity.

A further strength of this thesis lies in the fact that the same set of bilingual participants was assessed in all three experiments (i.e., 28 Polish-English adults) which allowed investigation of various processes across experiments within the same language group. Based on these, the study is considered to make a relevant contribution in offering a novel insight into the relationships between executive control, working memory, episodic memory, and vocabulary.

It is also argued that current findings are largely generalisable in terms of executive control, working memory, and episodic memory performance in bilingual samples that share language proficiency characteristics of the included bilingual sample. For such language groups, no bilingual advantage in executive control, working memory, and episodic memory may be found due to the lack of enhanced executive control and superior memory which is often observed for highly proficient bilinguals who are more efficient at interference suppression and conflict resolution.

One of the main limitations of this thesis is the relatively small sample size of both monolingual and bilingual participants. It is thus possible that the results obtained in experimental chapters (i.e, the lack of bilingual advantage) are affected by this. Moreover, the monolingual sample size varies across experiments. Therefore, it seems plausible to aim to replicate the study with a larger and well-matched sample of monolingual and bilingual participants.

A further crucial limitation is English proficiency of the bilingual sample which might compromise the results in experimental chapters. Bilingual participants were asked to complete a language background questionnaire where they were asked to self-assess their proficiency in both languages and frequency of daily usage of both languages in various contexts (see Appendix G). Recruitment of participants took place based on their self-reported English proficiency and no additional language screening was conducted at recruiting. Experiments aimed to recruit highly proficient Polish-English bilinguals, however, correlational analyses revealed that the lack of bilingual advantage on certain tasks as well as worse bilingual performance may be linked to low English proficiency (measured by WASI Vocabulary Subtest), their level of bilingualism (i.e., balance ratio), and the frequency of English use. The balance ratio for this sample was .78 (SD = .15) which differs significantly from a value that is often taken to indicate perfect bilingual balance (i.e., a ratio of 1.0) (Warmington, Kandru-Pothineni, & Hitch, 2019).

This was further corroborated by the time-course verbal fluency analysis for bilingual sample which revealed that their performance was significantly better in Polish than English. Although the project aimed to recruit a monolingual and bilingual sample matched on vocabulary and fluid intelligence, the former assumption was not met. There were statistically significant differences between language groups in all three experiments with English monolinguals' vocabulary scores being higher than Polish-English bilinguals (i.e., Experiment 1: p = .039; Experiment 2: p = .022; and Experiment 3: p = .018. In the future, it would be beneficial to utilise additional measures to assess language background characteristics at the point of recruitment to ensure an equal distribution of language proficiency and balanced frequency of language usage in bilinguals.

Finally, as noted in Experiment 3 in Chapter Nine, monolingual data for verbal fluency was not audio recorded due to the time constraints. It would be interesting to investigate language-group differences in terms of time-course analysis performance as well as quartile distribution analysis to observe whether there are differences in monolingual and bilingual episodic memory mechanisms.

Although there are certain limitations to this study, it can be viewed as an insightful and critical step towards an understanding of a bilingual advantage in long-term memory mechanisms.

10.5 Conclusions

The aim of this section is to collate evidence from the experimental chapters and draw the general conclusion. As presented in Chapter Three, bilingual advantage in executive control and working memory has been of interest for a number of years with research yielding inconsistent results (for a review of bilingual advantage see Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012: however, see De Bruin, Treccani & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson & Sawi, 2015). Taken together, there is a lack of consensus regarding bilingual advantage within these two domains. As executive control is also implicated in long-term memory mechanisms (i.e, episodic memory processes), it is of great importance to examine how bilingualism impacts long-term memory mechanisms. Currently, there is a very limited understanding of whether bilingualism impacts episodic memory as evidenced by the number of studies included in the systematic review presented in Chapter Four. Overall, twenty-two eligible studies were included in the meta-analytic considerations with ten studies examining recall and eleven studies investigating recognition as measures of episodic memory performance. Thus, the aim of this thesis was to extend the current state of knowledge regarding bilingual advantage by investigating not only executive control and working memory, but also long-term memory (i.e., episodic memory) as measured by the paired associate learning task and verbal fluency.

The experiments conducted as a part of this thesis yielded interesting outcomes. Experiment One yielded results partially in line with the results obtained in the meta-analysis where working memory but not executive control bilingual advantage was found. In this experiment, no bilingual advantage was found for either working memory or executive control (i.e., apart from significant between-language group differences for the Simon effect). Results obtained in Experiment Two are not consistent with results from the meta-analysis that revealed a bilingual advantage in episodic memory in terms of recall and recognition. Bilingual advantage in paired associate learning was not replicated in the second experiment. Finally, Experiment 3 did not replicate bilingual advantage in episodic memory as measured by verbal fluency which is consistent with research findings concerning lower proficiency second language speakers.

Indeed, experimental results obtained in this thesis are largely inconsistent with the results obtained from the systematic review in Chapter Four which sheds different light on our understanding of the bilingual advantage in episodic memory, working memory and executive control. It might be due to the above discussed factors such as lower second language proficiency levels of the bilingual sample than initially anticipated or relatively small sample size of both language groups. However, there was also high between-study heterogeneity in meta-analyses included in Chapter Four which might have yielded not reliable results.

Nevertheless, for future studies it is of great importance to select and secure a bilingual sample that demonstrates high proficiency in the second language and a balanced level of bilingualism. Additionally, it needs to be ensured that bilinguals demonstrate a balanced frequency of using both languages to ensure they have to manage two linguistic systems equally often. Also, different ways of conceptualising bilingualism need to be considered and as well as other factors such as age of acquisition, patterns of languages used at home, and language environment among others. As the level of bilingualism seems to mediate working memory, executive control, and episodic memory, it is crucial to employ bias free measures to assess these components. In other words, participants who demonstrate imperfect balance ratio (i.e., there is a significant difference between their language proficiency levels), should be assessed in both of their languages as they are likely to underperform in their weaker language. Testing in the second language exclusively may not uncover their actual potential.

Taken together, this thesis contributed to the existing state of knowledge regarding the impact of bilingualism on episodic memory, working memory, and executive control. It also emphasises a great importance for future studies to ensure that bilingual sample is clearly defined and of the second language proficiency level that enables capturing a potential bilingual advantage in episodic memory, working memory, and executive control.

APPENDICES

Appendix A. Chapter Four: Search terms for the Systematic Review for Web of Science<1864 to 11 March 2021.

1 bilingual$ or bilingualism. ti, ts. (38076)

2 multilingual$ or multilingualism. ti, ts. (22830)

3 English as an Additional Language or EAL. ti, ts. (1984)

4 English as a Second Language or ESL. ti, ts. (10361)

5 Second Language Learn*. ti, ts. (3705)

6 or/1-5 (49213)

7 (word or lexical) and learning. ti, ts. (64230)

8 (novel or new) and word learning. ti, ts. (13775)

9 (new or novel) and word and (acquisit$ or acquir$). ti, ts. (889)

10 non$word learning. ti, ts. (17)

11 statistical learning. ti, ts. (5907)

12 paired-associate learning. ti, ts. (5009)

13 didactic learning. ti, ts. (172)

14 associative learning. ti, ts. (9368)

15 lexical consolidation. ti, ts. (20)

16 fictitious and (word or lexical) and learning. ti, ts. (0)

17 (implicit or explicit) and word learning. ti, ts. (2396)

18 (implicit or explicit) and lexical learning. ti, ts. (521)

19 or/7-18 (74447)

20 (executive or attentional or inhibitory) and control. ti, ts. (206270)

21 cognition. ti, ts. (348642)

22 cognitive control. ti, ts. (15750)

23 inhibition or response inhibition. ti, ts. (1942550)

24 attention or selective attention. ti, ts. (194421)

25 interference suppression or interference. ti, ts. (1085519)

26 working memory or verbal short-term memory. ti, ts. (92292)

27 executive function$. ti, ts. (42740)

28 or/20-27 (3866883)

29 language development. ti, ts. (27034)

30 oral and (language or proficiency or skill$ or abilit$). ti, ts. (61664)

31 (expressive or receptive) and (vocabulary or word$). ti, ts. (6624)

32 word knowledge. ti, ts. (671)

33 linguistic and (competence or ability$). ti, ts. (10941)

34 or/29-33 (96417)

35 age*of*acquisition or acquisition age or AOA. ti, ts. (55916)

36 (speech or language or communication) and (delay or disorder or impairment or difficult$). ti, ts. (492850)

37 developmental disabilit$. ti, ts. (4096)

38 attention deficit or ADHD. ti, ts. (100442)

39 dyslexia. ti, ts. (24446)

40 autism or ASD. ti, ts. (130579)

41 cochlear implant$. ti, ts. (30272)

42 deaf or hearing impair$. ti, ts. (74450)

43 brain and (damage or injury). ti, ts. (422082)

44 traumatic brain and (damage or injury). ti, ts. (107329)

45 syndrome or disability or disorder or impair$. ti, ts. (8459268)

46 low birth weight. ti, ts. (76030)

47 prematur$. ti, ts. (327672)

48 case and (report or stud$ or series). ti, ts. (4316712)

49 or/36-48 (10651402)

50 6 and 19 and 28 (556)

51 6 and 34 and 28 (665)

52 50 not 49 (390)

53 51 not 49 (407)

****

Appendix B. Departmental ethics approval for the experiments

The University Of Sheffield.

Downloaded: 08/02/2019
Approved: 30/08/2018

Marta Ciesielska
Registration number: 170219080
Human Communication Sciences
Programme: Postgraduate research programme

Dear Marta

PROJECT TITLE: Does L1 vocabulary knowledge influence L2 lexical learning? - existing lexical knowledge in L1 shapes the ability to learn new words in L2.
APPLICATION: Reference Number 022681

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 30/08/2018 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 022681 (dated 22/08/2018).
- Participant information sheet 1050510 version 2 (22/08/2018).
- Participant information sheet 1050509 version 2 (22/08/2018).
- Participant information sheet 1050508 version 2 (22/08/2018).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Jonathan Stray
Ethics Administrator
Human Communication Sciences

Appendix C. Additional sixteen novel objects developed for the purpose of the study

Appendix D. Forty novel object pairings created in Phase Two

Pair 1. enlədiːv

Pair 2. brɪpə

Pair 3. rɜːl

Pair 4. keəhɒs


Pair 5. wɪlmɪkt


Pair 6. fluːəl


Pair 7. stɒd


Pair 8. kaɪnæfəl


Pair 9. kɒtəmægən

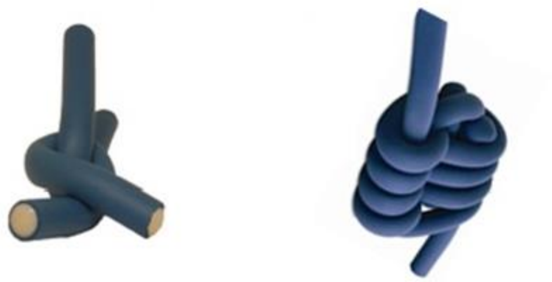

Pair 10. tʃrɪbdraɪvə


Pair 11. mɒnveɪ


Pair 12. feəp

Pair 13. wændaʊ

Pair 14. ɒtrɪpt



Pair 15. ʌsfrɒlə

Pair 16. pɪkeɪtə


Pair 17. kætəbɜːgə


Pair 18. trɔː


Pair 19. klaɪəs


Pair 20. drəʊks


Pair 21. hemıteltə


Pair 22. blempıt


Pair 23. kengsuːn

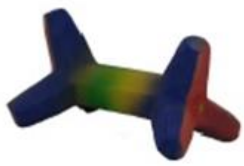

Pair 24. leməfeın


Pair 25. preɪn


Pair 26. keɪntgrʌʃ


Pair 27. aɪrəʊtreɪt


Pair 28. feləsuːsən

Pair 29. fænvə


Pair 30. bæskəl


Pair 31. treɪdʒ


Pair 32. kæŋbreʃ

Pair 33. eməfens

Pair 34. kraɪpɒkəbɑː

Pair 35. bræb

Pair 36. streɪbetʃi

Pair 37. əfɑːdiən

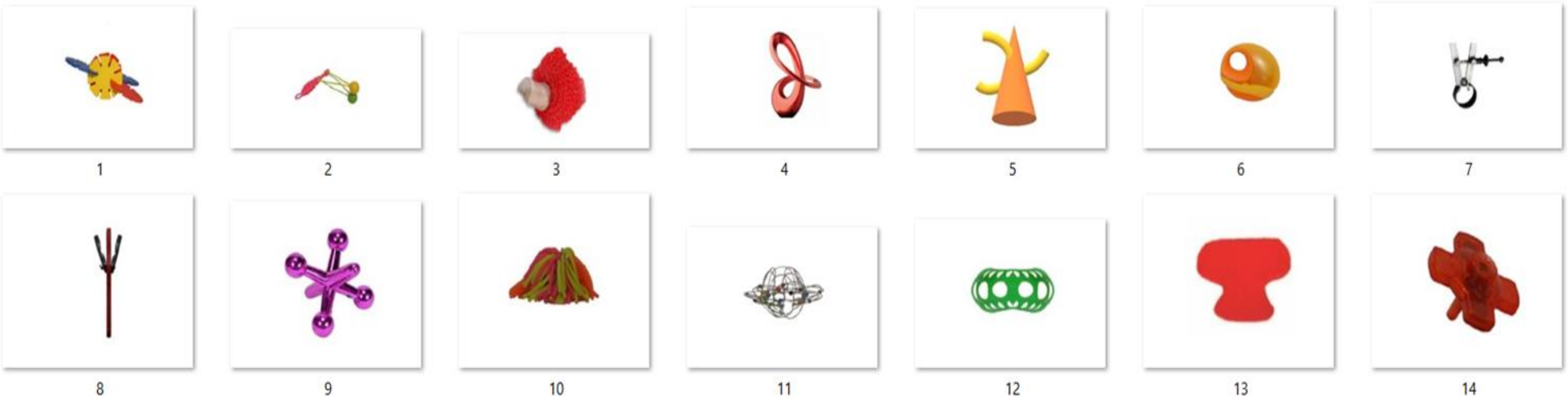
Pair 38. klaʊp

Pair 39. skwɪrɪt

Pair 40. krəʊt

Appendix E. Final set of items

Set A Version 1



Set A Version 2

Set B Version 1

Set B Version 2

Appendix F. Novel words in Set A and Set B (including two lists with reverse mappings)

Set A (20 items)

List A

1. enlədiːv
2. kɒtəmægən
3. kaɪnæfəl
4. stɒd
5. fluːəl
6. wɪlmɪkt
7. keəhɒs
8. rɜːl
9. brɪpə
10. tʃrɪbdraɪvə
11. ɒtrɪpt
12. wændaʊ
13. feəp
14. mɒnveɪ
15. krəʊt
16. skwɪrɪt
17. klaʊp
18. əfɑːdiən
19. streɪbetʃi
20. bræb

Items 1–5: Positive feedback

Items 6–10: Negative feedback

Items 11–15: Positive feedback

Items 16–20: Negative feedback

List B

1. enlədiːv
2. kɒtəmægən
3. kaɪnæfəl
4. stɒd
5. fluːəl

Negative feedback

6. wɪlmɪkt
7. keəhɒs
8. rɜːl
9. brɪpə
10. tʃrɪbdraɪvə

Positive feedback

11. ɒtrɪpt
12. wændaʊ
13. feəp
14. mɒnveɪ
15. krəʊt

Negative feedback

16. skwɪrɪt
17. klaʊp
18. əfɑːdiən
19. streɪbetʃi
20. bræb

Positive feedback

Set B (20 items)

List A

1. kraɪpɒkəbɑː
2. eməfens
3. kæŋbreʃ
4. treɪdʒ
5. bæskəl

Positive feedback

6. fænvə
7. feləsuːsən
8. aɪrəʊtreɪt
9. keɪntgrʌʃ
10. preɪn

Negative feedback

11. leməfeɪn
12. kengsuːn
13. blempɪt
14. hemɪteltə
15. drəʊks

Positive feedback

16. klaɪəs
17. trɔː
18. kætəbɜːgə
19. pɪkeɪtə
20. ʌsfrɒlə

Negative feedback

List B

1. kraɪpɒkəbɑː
2. eməfens
3. kæŋbreʃ
4. treɪdʒ
5. bæskəl
6. fænvə
7. feləsuːsən
8. aɪrəʊtreɪt
9. keɪntgrʌʃ
10. preɪn
11. leməfeɪn
12. kengsuːn
13. blempɪt
14. hemɪteltə
15. drəʊks
16. klaɪəs
17. trɔː
18. kætəbɜːgə
19. pɪkeɪtə
20. ʌsfrɒlə

Items 1–5: Negative feedback

Items 6–10: Positive feedback

Items 11–15: Negative feedback

Items 16–20: Positive feedback

Appendix G. Language Background Questionnaire

Language Background Questionnaire

Please answer the following questions to the best of your knowledge.

1. Please specify the age (a rough estimate) at which you started to learn each language in the following situations.

English	**Polish**
At home ____________	At home ____________
At school ____________	At school ____________

2. How did you learn English up to this point? (check all that apply)

Mainly through formal education ________
Mainly through interacting with people _________
Other (please specify) ___________________

3. Rate your proficiency in each language (circle the appropriate choice).

English

a. Very Poor
b. Poor
c. Fair
d. Functional
e. Good
f. Very good
g. Native-like

Polish

a. Very Poor
b. Poor
c. Fair
d. Functional
e. Good
f. Very good
g. Native-like

4. Rate how often you use each language per day at home (Circle the appropriate choice).

English

Never All the time

1 2 3 4 5

Polish
Never All the time
1 2 3 4 5

5. Rate how often you use each language per day with friends.

English
Never All the time
1 2 3 4 5

Polish
Never All the time
1 2 3 4 5

6. Rate how often you use each language per day for work or study related activities (e.g. going to classes, interacting with colleagues or classmates, writing assignments/papers etc.)

English
Never All the time
1 2 3 4 5

Polish
Never All the time
1 2 3 4 5

7. Do you have a working knowledge of any other language? YES NO

If yes, list languages and number of years of experience with each one.

Language name	Years of experience
_________________________	_______________
_________________________	_______________
_________________________	_______________

8. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

____

REFERENCES

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23* (4), 557–582. https://doi.org/10.1080/01690960801920602

Abwender, D. A., Swan, J., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative Analysis of Verbal Fluency Output: Review and Comparison of Several Scoring Methods. *Assessment*, *8*, 323–38. https://doi.org/10.1177/107319110100800308

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37* (1), 44–59. https://doi.org/10.1037/a0021205

Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36* (1), 17–36. https://doi.org/10.1037/a0017679

Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia*, *50* (13), 3004–3014. https://doi.org/10.1016/j.neuropsychologia.2012.07.038

Alloway, T. P. (2007). *Automated working memory assessment*. London: Psychological Corporation.

Alloway, T. P., Gathercole, S. E., Adams, A.-M., & Willis, C. (2005). Working memory abilities in children with special educational needs. Educational and Child Psychology, 22(4), 56–67.

Andreou, G., & Trott, K. (2013). Verbal fluency in adults diagnosed with attention-deficit hyperactivity disorder (ADHD) in childhood. *Attention deficit and hyperactivity disorders*, *5*. https://doi.org/10.1007/s12402-013-0112-z

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). https://doi.org/10.1371/journal.pone.0206770

Antón, E., Yuriem, G., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly? *Journal of Memory and* Language, 90, 147–160. https://doi.org/10.1016/j.jml.2016.04.007

Aron, A. R. (2008). Progress in Executive-Function Research: From Tasks to Functions to

Regions to Networks. *Current Directions in Psychological Science*, *17* (2), 124–129. https://doi.org/10.1111/j.1467-8721.2008.00561.x

Arora, S., Lawrence, M. A., & Klein, R. M. (2020). The attention network test database: ADHD and cross-cultural applications, *Frontiers in Psychology, 11*.

Artuso, C., & Palladino, P. (2019). Long-term memory effects on working memory updating development. *PLoS ONE*, *14* (5). https://doi.org/10.1371/journal.pone.0217697

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence, *The psychology of learning and motivation* (pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

August, D. (2006). Demographic Overview. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 43–49). Lawrence Erlbaum Associates Publishers.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. (1995). The psychology of memory. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 3–25). John Wiley & Sons.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *49A* (1), 5–28. https://doi.org/10.1080/027249896392784

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105* (1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47–90). Academic Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *36A* (2), 233–252. https://doi.org/10.1080/14640748408402157

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Baddeley, A., Vallar, G., & Wilson, B. (1987). Sentence comprehension and phonological memory: Some neuropsychological evidence. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 509–529). Lawrence Erlbaum Associates.

Baddeley, A. D., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40* (10), 1737–1743. https://doi.org/10.1016/S0028-3932(01)00146-4

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General*, *133* (1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33* (3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Bechara, A., & Martin, E. M. (2004). Impaired Decision Making Related to Working Memory Deficits in Individuals with Substance Addictions. *Neuropsychology*, *18* (1), 152–162. https://doi.org/10.1037/0894-4105.18.1.152

Bergmann, C., Paulus, M., & Fikkert, P. (2012). Preschoolers' comprehension of pronouns and reflexives: the impact of the task. *Journal of Child Psychology*, *39* (4), 777–803. https://doi.org/10.1017/S0305000911000298.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (2), 317–333. https://doi.org/10.1037/a0014873

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60* (1), 68–79. https://doi.org/10.1037/cjep2006008

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

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45* (2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10* (3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence from the Simon Task. *Psychology and Aging*, *19* (2), 290-303. https://doi.org/10.1037/0882-7974.19.2.290

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

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

Bialystok, E., Martin, M., & Viswanathan, M. (2005). Bilingualism across the Lifespan: The Rise and Fall of Inhibitory Control. *International Journal of Bilingualism*, *9*, 103–119. https://doi.org/10.1177/13670069050090010701.

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

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15* (11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology, 128,* 105-119. doi.org/10.1016/j.jecp.2014.06.007

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *An Introduction to Meta-Analysis. Introduction to Meta-Analysis*. https://doi.org/10.1002/9780470743386.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108* (3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Bretherton-Furness, J., Ward, D., & Saddy, D. (2016). Creating a non-word list to match 226 of the Snodgrass standardised picture set. *Journal of Phonetics and Audiology, 2* (1), 1-4.

Buac, M., Gross, M., & Kaushanskaya, M. (2016). Predictors of processing-based task performance in bilingual and monolingual children. *Journal of Communication Disorders, 62*, 12–29. https://doi.org/10.1016/j.jcomdis.2016.04.001

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33* (3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130* (3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61* (3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

Campoy, G. (2012). Evidence for decay in verbal short-term memory: A commentary on Berman, Jonides, and Lewis (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38* (4), 1129–1136. https://doi.org/10.1037/a0026934

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *Neuroimage, 54*, 550–559.

Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence. *Aging, Neuropsychology, and Cognition, 14* (6), 631–646. https://doi.org/10.1080/13825580600987660

Cheung, H. (1996). Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology, 32* (5). 867–873. https://doi.org/10.1037/0012-1649.32.5.867.

Chlapana, E., & Tafa, E. (2014). Effective practices to enhance immigrant kindergarteners' second language vocabulary learning through storybook reading. *Reading and Writing: An Interdisciplinary Journal, 27* (9), 1619–1640. https://doi.org/10.1007/s11145-014-9510-7

Chomsky, N. (1965). *Aspects of the theory of syntax.* M.I.T. Press.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence, 40* (3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America, 106*(5), 2805–2811. https://doi.org/10.1121/1.428106

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

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

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97* (3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Cohen, M. J., Morgan, A. M., Vaughn, M., Riccio, C. A., Hall, J. (1999). Verbal fluency in children: developmental issues and differential validity in distinguishing children with attention-deficit hyperactivity disorder and two subtypes of dyslexia. *Archives of Clinical Neuropsychology, 14,* 433–443.

Collette, F., Van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the Unity and Diversity of the Neural Substrates of Executive Functioning. *Human Brain Mapping, 25* (4), 409–423. doi.org/10.1002/hbm.20118

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34* (2), 302–312. https://doi.org/10.1037/0278-7393.34.2.302

Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41* (3), 614–635. https://doi.org/10.1037/xlm0000065

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology*, *55* (4), 429–432. http://dx.doi.org/10.1111/j.2044-8295.1964.tb00928.x

Corsi, P. M. (1973). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34* (2), 891.

Costa, A. & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, *2*, 231–244

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), 135–49. https://doi.org/10.1016/j.cognition.2009.08.001.

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

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41* (3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Cowan, N. (2005). *Working memory capacity.* Psychology Press. https://doi.org/10.4324/9780203342398

Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. S. C. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86–107). Psychology Press.

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51* (1), 42–100. http://doi: 10.1016/j.cogpsych.2004.12.001.

Craik, F.I.M., & Byrd, M. (1982) Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and Cognitive Processes* (pp. 191–211). Plenum Press, New York.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125* (2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11* (6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Crone, E. A., Bunge, S. A., van der Molen, M. W., & Ridderinkhof, K. R. (2006). Switching between tasks and responses: A developmental study. *Developmental Science, 9* (3), 278–287. https://doi.org/10.1111/j.1467-7687.2006.00490.x

Cummins, J. (1976). The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses. *Working Papers on Bilingualism*, *9*, 1–43.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19* (4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9* (4), 561–584. https://doi.org/10.1037/0278-7393.9.4.561

Davidson, D. J., & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*(38).

de Bruin, A., Treccani, B., & Della Sala, S. (2015). The connection is in the data: We should consider them all. *Psychological Science, 26* (6), 947–949. doi.org/10.1177/0956797615583443

Deeks, J.J., Higgins, J.P.T., & Altman, D. G. (2019). *Analysing data and undertaking meta-analyses*. https://doi.org/10.1002/9781119536604.ch10.

DeFries, J. C., Plomin, R., Vandenberg, S. G., & Kuse, A. R. (1981). Parent–offspring resemblance for cognitive abilities in the Colorado Adoption Project: Biological, adoptive, and control parents and one-year-old children. *Intelligence, 5* (3), 245–277. https://doi.org/10.1016/S0160-2896(81)80012-8

Dennis, J. P., & Vander Wal, J. S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34* (3), 241–253. https://doi.org/10.1007/s10608-009-9276-4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi:10.1146/annurev-psych-113011-143750.

Diana, R. A., Van den Boom, W., Yonelinas, A. P., & Ranganath, C. (2011). ERP correlates of source memory: Unitized source information increases familiarity-based retrieval. *Brain Research, 1367,* 278–286. doi.org/10.1016/j.brainres.2010.10.030

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11* (9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Lawrence Erlbaum Associates Publishers.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5* (3), 175–197. https://doi.org/10.1017/S1366728902003012

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41* (5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

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. https://doi: 10.3758/s13423-019-01567-z.

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. Psychological Science, 18(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duncan, J., & Miller, E. K. (2002). Cognitive focus through adaptive neural coding in the primate prefrontal cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 278–291). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195134971.003.0018

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23* (10), 475–483. https://doi.org/10.1016/S0166-2236(00)01633-7

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test - revised*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test Third Edition*. Circle Pines, MN: American Guidance.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale*. Windsor, England: NFER-Nelson.

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody: TVIP: Adaptacion Hispanoamericana (Peabody Picture Vocabulary Test: PPVT: Hispanic-American Adaptation)*. Circle Pines, MN: American Guidance Service (AGS), Inc.

Dunn, L. M., Thériault-Whalen, C.M., & Dunn, L.M. (1993). *Echelle de Vocabulaire en Images Peabody*. Toronto: Psycan.

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56* (2), 455–63.

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, *74,* 1–15. https://doi.org/10.1016/j.jml.2014.03.006

Egger, M., Davey Smith, G., & Altman, D. G. (2001). *Systematic reviews in health care: meta analysis in context*. London: BMJ Books.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in Meta-Analysis Detected by a Simple, Graphical Test. *British Medical Journal*, *315* (7109): 629–34.

Engel de Abreu, P. M. (2011). Working memory in multilingual children: is there a bilingual effect? *Memory*, *19* (5), 529–37. https://doi: 10.1080/09658211.2011.590504.

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory (pp*. 145–199). Elsevier Science.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.007

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128* (3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

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

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology, 20* (5), 607–613. https://doi.org/10.1037/0894-4105.20.5.607

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage*, *18* (1), 42–57.

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. https://doi.org/10.1162/089892902317361886

Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. (2007) G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, 39, 175-191. http://dx.doi.org/10.3758/BF03193146

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego, CA: Singular Publishing Group.

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

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psycholog*y, *63* (3), 665–694. doi.org/10.1348/000711010X502733

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, foreign-language vocabulary learning. *Journal of Memory and Language*, *30* (3), 331–347. https://doi.org/10.1016/0749-596X(91)90040-Q

Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General, 133* (1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*, 172–179.

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology, 47* (5), 1410–1430. https://doi.org/10.1037/a0023750

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137* (2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

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. doi.org/10.1080/23273798.2014.918630

Fuster, J. M. (2003). *Cortex and mind: Unifying cognition.* Oxford University Press.

García-Pentón, L., García, Y. F., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). "The neuroanatomy of bilingualism: how to turn a hazy view into the full picture": Corrigendum. *Language, Cognition and Neuroscience, 31* (3). https://doi.org/10.1080/23273798.2016.1144398

Garlitch, S. M., & Wahlheim, C. N. (2020). The role of reminding in retroactive effects of memory for older and younger adults. *Psychology and Aging, 35* (5), 697–709. https://doi.org/10.1037/pag0000427

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological Short-Term Memory and New Word Learning in Children. *Developmental Psychology*, *33*, 966–979.

Gathercole, S.E., & Baddeley, A.D. (1993). *Working Memory and Language*. Hove: Erlbaum.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during early school years: A longitudinal study. *Developmental Psychology, 28,* 887–898.

Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and* Psychiatry, *45* (4), 836–54.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology, 28* (1), 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33* (7), 1220–1234. https://doi.org/10.3758/BF03193224

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

Gough. D. (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22* (2), 213-228, https://doi.org/10.1080/02671520701296189

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 207–220). North-Holland. https://doi.org/10.1016/S0166-4115(08)61496-X

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1* (2), 67–81.

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. https://doi.org/10.1016/j.cognition.2017.02.010

Grundy, J. G., & Timmer, K. (2016). Cognitive mechanisms underlying performance differences between monolinguals and bilinguals. In J. W. Schwieter (Ed.), *Cognitive control and consequences of multilingualism* (pp. 375–396). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.2.16gru

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13* (2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia*, *14* (1), 29–37.

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, *58* (1), 1–31. https://doi.org/10.1111/j.1467-9922.2007.00433.x

Hamada, M., & Koda, K. (2011). Similarity and difference in learning L2 word-form. *System*, *39*, 500–509.

Hansen, L. B., Morales, J., Macizo, P., Duñabeitia, J. A., Saldaña, D., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2017). Reading comprehension and immersion schooling: Evidence from component skills. *Developmental Science*, *20* (1), 1–16. https://doi.org/10.1111/desc.12454

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press. ISBN 978-0-367-61007-4.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (*pp. 193–225). Academic Press. https://doi.org/10.1016/S0079-7421(08)60041-9

Hazeltine, E., & Wifall, T. (2011). Searching working memory for the source of dual-task costs. *Psychological Research, 75* (6), 466–475. https://doi.org/10.1007/s00426-011-0343-6

Hedden, T., & Gabrieli, J. D. (2010). Shared and selective neural correlates of inhibition, facilitation, and shifting processes during executive control. *Neuroimage, 51*, 421–431.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behavior Research Methods, 50(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin*, *96* (3), 573–580. https://doi.org/10.1037/0033-2909.96.3.573

Hedges, L. V., & Vevea, J. L. (1996). Estimating Effect Size Under Publication Bias: Small Sample Properties and Robustness of a Random Effects Selection Model. *Journal of Educational and Behavioral Statistics*, *21* (4), 299–332.

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.

Henry, J. D., & Crawford, J. R. (2004). A Meta-Analytic Review of Verbal Fluency Performance in Patients with Traumatic Brain Injury. *Neuropsychology*, *18* (4), 621–628. https://doi.org/10.1037/0894-4105.18.4.621

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. https://doi.org/10.1017/S1366728909990010

Hernández, M., Costa, A., & Humphreys, G. W. (2012). Escaping capture: Bilingualism modulates distraction from working memory. *Cognition*, *122* (1), 37–50. https://doi.org/10.1016/j.cognition.2011.08.002

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133* (4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem Solving. *Science*, *140*, 1190–92.

Hester, R., & Garavan, H. (2005). Working memory and executive function: the influence of content and load on the control of attention. *Memory and Cognition*, *33* (2), 221–33. http://doi: 10.3758/bf03195311. PMID: 16028577.

Higgins, J. P. T., & Thompson. S. G. (2002). Quantifying Heterogeneity in a Meta-Analysis. *Statistics in Medicine*, *21* (11). 1539–58.

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. https://doi.org/10.3758/s13423-011-0116-7

Hilchey, M. D., Saint-Aubin, J., & Klein, R. (2015). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In J. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing*. Cambridge: Cambridge University Press. doi.org/10.1017/CBO9781107447257.026

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44* (12), 2547-57. 10.1016/j.neuropsychologia.2006.03.035

Hoeven, N., & De Bot, K. (2012). Relearning in the Elderly: Age-Related Effects on the Size of Savings. *Language Learning*, *62*. doi.org/10.1111/j.1467-9922.2011.00689.x.

Horst, J. S., & Hout, M. C. (2015). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*. https://doi.org/10.3758/s13428-015-0647-3

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13* (2), 128–157. https://doi.org/10.1080/15250000701795598

Hübner, R., & Töbel, L. (2019). Conflict resolution in the Eriksen flanker task: Similarities and differences to the Simon task. PLoS ONE, 14(3), Article e0214203. https://doi.org/10.1371/journal.pone.0214203

Hudson, S. R., & Baddeley, A. D. (1996). Working memory in chess. *Memory & Cognition*, *24* (1), 83–93. https://doi.org/10.3758/BF03197274

Huizinga, M., Dolan, C.V., & van der Molen, M. W. (2006). Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia*, *44* (11), 2017–2036. https://doi:10.1016/j.neuropsychologia.2006.01.010.

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology, 30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Jacques, S., & Zelazo, P. D. (2005). Language and the Development of Cognitive Flexibility: Implications for Theory of Mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 144–162). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0008

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39* (7), 1198–1210. https://doi.org/10.3758/s13421-011-0094-z

James, W. (1890). *The principles of psychology, Volume 1.* Henry Holt and Co. https://doi.org/10.1037/10538-000

Jonides, J. (2003). Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia*, *41* (3), 357–370. https://doi.org/10.1016/S0028-3932(02)00167-7

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224. doi: 10.1146/annurev.psych.59.103006.093615.

Jubenville, K., Sénéchal, M., & Malette, M. (2014). The moderating effect of orthographic consistency on oral vocabulary learning in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *126*, 245–263. doi.org/10.1016/j.jecp.2014.05.002

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kan, P. F., & Kohnert, K. (2012). A growth curve analysis of novel word learning by sequential bilingual preschool children. *Bilingualism: Language and Cognition*, *15* (3), 452–469.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130* (2), 169–183. https://doi.org/10.1037/0096-3445.130.2.169

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132* (1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144* (11), 1147–1185. doi: 10.1037/bul0000160.

Katzev, M., Tüscher, O., Hennig, J., Weiller, C., & Kaller, C. P. (2013). Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: The crucial role of task demands and individual ability. *The Journal of Neuroscience*, *33* (18), 7837–7845. doi.org/10.1523/JNEUROSCI.3147-12.2013

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, *15* (3), 470–489. https://doi.org/10.1017/S1366728911000472

Kaushanskaya, M., & Marian, V. (2009). Bilingualism reduces native-language interference during novel-word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (3), 829–835. https://doi.org/10.1037/a0015275

Kaushanskaya, M., Yoo, J., & Van Hecke, S. (2013). World learning in adults with second-language experience: Effects of phonological and referent familiarity. *Journal of Speech, Language, and Hearing Research*, *56* (2), 667–678. https://doi.org/10.1044/1092-4388.

Kelley, T. A., & Lavie, N. (2010). Working memory load modulates distractor competition in primary visual cortex. *Cerebral Cortex*, *21* (3), 659–65. doi: 10.1093/cercor/bhq139.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55* (4), 352–358. https://doi.org/10.1037/h0043688.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112* (2), 240–264. https://doi.org/10.1037/0096-3445.112.2.240

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. American Psychological Association. doi.org/10.1037/10693-000

Kohnert, K. (2008). *Language disorders in bilingual children and adults*. San Diego, CA: Plural Publishing.

Konstantinou, N., Beal, E., King, J.-R., & Lavie, N. (2014). Working memory load and distraction: Dissociable effects of visual maintenance and cognitive control. *Attention, Perception, & Psychophysics, 76* (7), 1985–1997. doi.org/10.3758/s13414-014-0742-z

Konstantinou, N., & Lavie, N. (2013). Dissociable roles of different types of working memory load in visual detection. *Journal of Experimental Psychology: Human Perception and Performance*, *39* (4), 919–924. http://dx.doi.org/10.1037/a0033037

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology, 65* (2), 356–369. doi.org/10.1080/17470218.2011.604788

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. https://doi.org/10.3389/fnbeh.2014.00250.

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *106* (16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, *11* (2), 203–223. https://doi.org/10.1017/S1366728908003386

Kroll, J. F., & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. *Journal of cognitive psychology*, *25* (5). https://doi.org/10.1080/20445911.2013.799170.

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence*, *12* (4), 393–421. doi.org/10.1016/0160-2896(88)90004-9.

Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General*, *120* (1), 57–79. https://doi.org/10.1037/0096-3445.120.1.57

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9* (2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, *133* (3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lechevalier, B. & Habas, Christophe. (2020). Mémoire procédurale et mémoire déclarative. *Bulletin de l'Académie Nationale de Médecine*, *205*. https://doi.org/10.1016/j.banm.2020.12.011.

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), 394–425. https://doi.org/10.1037/bul0000142

Lenhard, W. & Lenhard, A. (2015). *Calculation of Effect Sizes*. Psyhometrica, Bibergau (Germany). http://www.psychometrica.de/effect_size.htlm

Lenneberg, E.H. (1967). *Biological foundations of language.* Wiley.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior*, *13* (5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). Academic Press.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91* (3), 295–327. https://doi.org/10.1037/0033-295X.91.3.295

Logie, R. H. (1995). *Visuo-spatial working memory.* Lawrence Erlbaum Associates, Inc.

Logie, R. H. (1999). Working memory. *The Psychologist*, *12* (4), 174–178.

López, B. G., Luque, A., & Piña-Watson, B. (2021). Context, intersectionality, and resilience: Moving toward a more holistic study of bilingualism in cognitive science. *Cultural Diversity and Ethnic Minority Psychology*. Advance online publication. https://doi.org/10.1037/cdp0000472

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*, 174–207. https://doi:10.3758/BF03210959

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

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

Mahone, E. & Koth, C., Cutting, L., Singer, H. & Denckla, M. (2001). Executive function in fluency and recall measures among children with Tourette syndrome or ADHD. *Journal of the International Neuropsychological Society*, *7*, 102–11. https://doi.org/10.1017/S1355617701711101.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107* (2), 395–419. https://doi.org/10.1016/j.cognition.2007.10.003

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87* (3), 252–271. https://doi.org/10.1037/0033-295X.87.3.252

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50* (4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martin, R. C., Hamilton, A. C., Lipszyc, M., & Potts, G. F. (2014). *Manipulation of Inhibition Demands in a Working Memory Task: Evidence from Patient and ERP Data*. Poster presented at the 11th annual meeting of the Cognitive Neuroscience Society. San Francisco, CA, USA.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11* (1), 81–93. https://doi.org/10.1017/S1366728907003227

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122* (3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24* (2), 222–243. https://doi.org/10.1037/a0017619.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 155–200). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(06)46005-9

Meijer, J. H., Schmitz, N., Nieman, D. H., Becker, H. E., van Amelsvoort, T. A. M. J., Dingemans, P. M., Linszen, D. H., & de Haan, L. (2011). Semantic fluency deficits and reduced grey matter before transition to psychosis: A voxelwise correlational analysis. *Psychiatry Research: Neuroimaging*, *194*, 1–6.

Menjivar, J., & Akhtar, N. (2016). Language experience and preschoolers' foreign word learning. *Bilingualism: Language and Cognition*, *20*, 1–7. doi.org/10.1017/S1366728916001103.

Merriman, W. E. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp. 331–358). Lawrence Erlbaum Associates Publishers.

Michael, E. B., & Gollan, T. H. (2005). Being and Becoming Bilingual: Individual Differences and Consequences for Language Production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 389–407). Oxford University Press.

Miller, A. L., & Unsworth, N. (2018). Individual differences in working memory capacity and search efficiency. *Memory & Cognition*, *46*, 1149–1163.

Miller, A. L., & Unsworth, N. (2020). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46* (12), 2277–2294. https://doi.org/10.1037/xlm0000797

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63* (2), 81–97. https://doi.org/10.1037/h0043158

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., Kross, E., Teslovich, T., Wilson, N. L., Zayas, V., & Shoda, Y. (2011). 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6* (2), 252–256. https://doi.org/10.1093/scan/nsq081

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. https://doi.org/10.1006/cogp.1999.0734.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7* (3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Morales, J., Calvo, A., & Bialystok, E. (2012). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114* (2):187–202. https://doi.org/10.1016/j.jecp.2012.09.002.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2016). Multi-component perspective of cognitive control in bilingualism. In J. W. Schwieter (Ed.*), Cognitive control and consequences of multilingualism* (pp. 271–296). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.2.12mor

Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review*, *20* (1), 163–170. https://doi.org/10.3758/s13423-012-0313-z

Morini, G. (2014). *Interactions between language experience and cognitive abilities in word learning and word recognition* (Publication No. AAI3644000) [Doctoral Dissertation, University of Maryland]. University Libraries database. https://doi.org/10.13016/M2M881

Morris, R. G., Downes, J. J., Sahakian, B. J., Evenden, J. L., Heald, A., & Robbins, T. W. (1988). Planning and spatial working memory in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *51* (6), 757–766. https://doi.org/10.1136/jnnp.51.6.757

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005a). *Batería III Woodcock-Muñoz*. Itasca, IL: Riverside Publishing.

Murphy, V., & Unthiah, A. (2015). *A systematic review of intervention research examining English language and literacy development in children with English as an Additional Language (EAL).* Retrieved from www.naldic.org.uk/Resources/NALDIC/Research and Information/Documents/eal-systematic-review-prof-v-murphy.pdf

Nair, Vishnu & Biedermann, Britta & Nickels, Lyndsey. (2015). Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism*, *20*. https://doi.org/10.1177/1367006914567005.

Nakai, S., Lindsay, S., Ota, M. (2015) A prerequisite to L1 homophone effects in L2 spoken-word recognition. *Second Language Research*, *31*, 29–52.

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, *13* (5), 597–616. https://doi.org/10.1080/13670050.2010.488288

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, *6* (2), 136–142. https://doi.org/10.1111/1467-7687.00263

Nee, D. E., & Jonides, J. (2013). Trisecting representational states in short-term memory. *Frontiers in Human Neuroscience, 7* (796).

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51* (2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73,* 15–30. doi.org/10.1016/j.jml.2014.02.003.

Paap, K. R., Anders-Jefferson, R., Mason, L., Alvarado, K., & Zimiga, B. (2018). Bilingual Advantages in Inhibition or Selective Attention: More Challenges. *Frontiers in Psychology*, *15* (9), 1409. https://doi: 10.3389/fpsyg.2018.01409.

Paap, K. R., Anders-Jefferson, R., Zimiga, B., Mason, L., & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, Simon, and spatial Stroop tasks? *Cognitive Research: Principles and Implications*, *5* (7). https://doi.org/10.1186/s41235-020-0207-y

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. https://doi.org/10.1016/j.cogpsych.2012.12.002

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

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *74,* 305–314. https://doi.org/10.1016/j.cortex.2015.09.010

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

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, (962). doi.org/10.3389/fpsyg.2014.00962

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology*, *44*, 47–67.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *48A* (1), 98–107. https://doi.org/10.1080/14640749508401378

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30* (3), 331–347. https://doi.org/10.1016/0749-596X(91)90040-Q

Patra, A., Bose, A., & Marinis, T. (2020). *Verbal fluency in Bengali-English bilinguals with aphasia and bilingual healthy controls*. University of Reading Dataset. https://doi.org/10.17864/1947.261

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* Oxford Univ. Press.

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, *3* (6), 346–353. https://doi.org/10.1111/j.1467-9280.1992.tb00045.x

Pelham, S. D., & Abrams, L. (2013). Cognitive advantages and disadvantages in early and late bilinguals. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40* (2), 313–25. https://doi.org/10.1037/a0035224.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia*, *12* (3), 323–330. https://doi.org/10.1016/0028-3932(74)90047-5

Persuh, M., Genzer, B., & Melara, R. D. (2012). Iconic memory requires attention. *Frontiers in Human Neuroscience, 6* (126). https://doi.org/10.3389/fnhum.2012.00126

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, *20* (3), 249–262. https://doi.org/10.1016/0028-3932(82)90100-2

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29* (2), 187–204. https://doi.org/10.1037/a0036085

Pettit, L., McCarthy, M., Davenport, R., & Abrahams, S. (2013). Heterogeneity of letter fluency impairment and executive dysfunction in Parkinson's disease. *Journal of the International Neuropsychological Society*, *19* (9), 986–994. https://doi.org/10.1017/S1355617713000829

Perani, D., & Abutalebi, J. (2015). Bilingualism, dementia, cognitive and neural reserve. *Current Opinion in Neurology*, *28* (6), 618–25. doi.org/10.1097/WCO.0000000000000267

Pino Escobar, G., Kalashnikova, M., & Escudero, P. (2018). Vocabulary matters! The relationship between verbal fluency and measures of inhibitory control in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *170*, 177–189. https://doi.org/10.1016/j.jecp.2018.01.012

Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science, 4* (3–4), 293–312. doi.org/10.1080/09540099208946620

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113* (4), 535–551. https://doi.org/10.1016/j.jecp.2012.06.013

Pomper, R., & Saffran, J. (2018). Familiar Object Salience Affects Novel Word Learning. *Child Development*, *90*. https://doi.org/10.1111/cdev.13053.

Posner, M. I., & Snyder, C. R. R. (*1975*). Attention and Cognitive Control. In R. L. Solso (Ed.), *Information Processing and Cognition* (pp. 55–85). The Loyola Symposium: Lawrence Erlbaum.

Postle, B. (2006). Working memory as an emergent property of the mind. *Neuroscience, 139*, 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005.

Prior, A., & Gollan, T. (2011). Good Language-Switchers are Good Task-Switchers: Evidence from Spanish-English and Mandarin-English Bilinguals. *Journal of the International Neuropsychological Society*, 17, 682–91. doi.org/10.1017/S1355617711000580.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13* (2), 253–262. doi.org/10.1017/S1366728909990526

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory & Cognition*, *29* (8), 1073–1080. doi.org/10.3758/BF03206375

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, *14*, 1027–1042. https://doi.org/10.3758/BF03193088

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior Temporal, Prefrontal, and Hippocampal Contributions to Visual Working Memory Maintenance and Associative Memory Retrieval. *The Journal of Neuroscience*, *24* (16), 3917–3925. https://doi.org/10.1523/JNEUROSCI.5053-03.2004

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

Raven, J. C., Court, J. H., & Raven J. (1983). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 4 Advanced Progressive Matrices).

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44* (4), 501–526. https://doi.org/10.1037/xlm0000450

Reynolds, G. D., & Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Frontiers in Systems Neuroscience*, *10* (15).

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnyhough, C., Henson, R., & Hudson, S. R. (1996). Working memory in chess. *Memory and Cognition*, *24* (1), 83–93. https://doi.org/10.3758/bf03197274.

Roca, M., Manes, F., Chade, A., Gleichgerrcht, E., Gershanik, O., Arévalo, G. G., et al. (2012). The relationship between executive functions and fluid intelligence in Parkinson's disease. *Psychological Medicine*, *42*, 2445–2452. doi.org/10.1017/S0033291712000451

Rohrer, D., Wixted, J. T., Salmon, D. P., & Butters, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21* (5), 1127–1139. https://doi.org/10.1037/0278-7393.21.5.1127

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science*, *23* (11), 1345–55. https://doi.org/10.1177/0956797612444902.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126* (3), 211–227. https://doi.org/10.1037/0096-3445.126.3.211

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118* (2), 183–192. https://doi.org/10.1037/0033-2909.118.2.183

Rudner, M., & Rönnberg, J. (2008). The role of the episodic buffer in working memory for language processing. *Cognitive Processing*, *9* (1), 19–28. doi.org/10.1007/s10339-007-0183-x

Salvatierra, J., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*, 15, 26–37. doi.org/10.1177/1367006910371021.

Sanchez-Azanza, V. A., López-Penadés, R., Buil-Legaz, L., Aguilar-Mediavilla, E., Adrover-Roig, D. (2017). Is bilingualism losing its advantage? A bibliometric approach. *PLoS ONE*, *12*. https://doi.org/e0176151.

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: Language and Cognition*, *13* (2), 231–252. doi.org/10.1017/S1366728909990514

Sasisekaran J, & Weisberg S. (2013). The effects of cognitive: linguistic variables and language experience on behavioural and kinematic performances in nonword learning. *Journal of Psycholinguistic Research, 42* (2),175–90. https://doi.org/10.1007/s10936-012-9218-2.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1089–1096. https://doi.org/10.1037/a0015730

Schrauf, R. W., & Rubin, D. C. (2003). On the bilingual's two sets of memories. In R. Fivush & C. A. Haden (Eds.), *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives* (pp. 121–145). Lawrence Erlbaum Associates Publishers.

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

Schwartz, S., & Baldo, J. (2001). Distinct patterns of word retrieval in right and left frontal lobe patients: A multidimensional perspective. *Neuropsychologia*, *39* (11), 1209–1217. https://doi.org/10.1016/S0028-3932(01)00053-7

Schwarzer, G. (2007). *Meta: An R package for meta-analysis*. R News, 7, 40–5.

Seibert, P. S., & Ellis, H. C. (1991). Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory & Cognition*, *19* (5), 507–513. doi.org/10.3758/BF03199574

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition (CELF-4)*. San Antonio, TX: Psychological Corporation.

Shallice, T., & Vallar, G. (1990). The impairment of auditory-verbal short-term storage. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 11–53). Cambridge University Press.

Shallice, T., & Warrington, E. K. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. *Brain and Language*, *4* (4), 479–491. doi.org/10.1016/0093-934X(77)90040-2

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5* (772). https://doi.org/10.3389/fpsyg.2014.00772

Simos, P. G., Sideridis, G. D., Protopapas, A., & Mouzaki, A. (2011). Psychometric evaluation of a receptive vocabulary test for Greek elementary students. *Assessment for Effective Intervention*, *37* (1), 34–49. https://doi.org/10.1177/1534508411413254

Skinner, B. F. (1957). *Verbal behavior.* Appleton-Century-Crofts. doi.org/10.1037/11256-000

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33* (1), 5–42. https://doi.org/10.1006/cogp.1997.0658

Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition, 38* (5), 651–658. doi.org/10.3758/MC.38.5.651

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74* (11), 1–29. https://doi.org/10.1037/h0093759

Sperling, G., & Speelman, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 149–202). New York: Academic Press.

Spiers, H. J., Maguire, E. A., & Burgess, N. (2001). Hippocampal amnesia. *Neurocase*, *7* (5), 357–382. https://doi.org/10.1076/neur.7.5.357.16245

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5* (1), 60–78. https://doi.org/10.1002/jrsm.1095

Staresina, B. P., & Davachi, L. (2006). Differential Encoding Mechanisms for Subsequent Associative Recognition and Free Recall. *The Journal of Neuroscience*, *26* (36), 9162–9172. https://doi.org/10.1523/JNEUROSCI.2877-06.2006

Sterne, J. A. C., Egger, M., & Moher, D. (2008). Addressing reporting biases. In: J. P. T. Higgins & S. Green (Eds.) *Cochrane handbook for systematic reviews of interventions*. Wiley.

Stins, J. F., Polderman, J. C. (T.), Boomsma, D. I., & de Geus, E. J. C. (2005). Response interference and working memory in 12-year-old children. *Child Neuropsychology*, *11* (2), 191–201. https://doi.org/10.1080/092970490911351

Stocco, A., & Prat, C. S. (2014). Bilingualism trains specific brain circuits involved in flexible rule selection and application. *Brain and Language, 137*, 50–61. https://doi.org/10.1016/j.bandl.2014.07.005

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18* (6), 643–662. https://doi.org/10.1037/h0054651

Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society*, *17* (5), 759–765. https://doi.org/10.1017/S1355617711000695

Sylvester, C.-Y., Wagner, T. D., Lacey, S. C., Hernandez, L., Nichols, T. E., Smith, E. E., & Jonides, J. (2003). Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia, 41* (3), 357–370. doi.org/10.1016/S0028-3932(02)00167-7

Takács, Á., Kóbor, A., Tárnok, Z., & Csépe, V. (2013). Verbal fluency in children with ADHD: strategy using and temporal properties. *Child Neuropsychology*, *20* (4), 415–29. https://doi.org/10.1080/09297049.2013.799645.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45* (1), 37–55. https://doi.org/10.1037/xlm0000578

Tirre, W. C. (1991). Knowledge and ability factors underlying simple learning by accretion. *Genetic, Social, and General Psychology Monographs*, *117* (1), 91–126.

Treccani, B., & Mulatti, C. (2015). No matter who, no matter how… and no matter whether the white matter matters. Why theories of bilingual advantage in executive functioning are so difficult to falsify. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *73*, 349–351. https://doi.org/10.1016/j.cortex.2015.07.015

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22* (3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53* (1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Unsworth, N. (2010). Interference control, working memory capacity, and cognitive abilities: A latent variable analysis. *Intelligence*, *38* (2), 255–267. doi.org/10.1016/j.intell.2009.12.003

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

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38* (6), 1765–1772. doi.org/10.1037/a0028075

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

Vandek, M., Gabrić, P., Kužina, I., Erdeljac, V., & Sović, M. (2018). *Verbal fluency and working memory interaction*. [poster presentation] doi.org/10.5281/zenodo.4681276.

van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S. H., & Lim, S. (2019). Does the Bilingual Advantage in Cognitive Control Exist and If So, What Are Its Modulating Factors? A Systematic Review. *Behavioural Science*, *9*, 27.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136* (4), 601–626. https://doi.org/10.1037/a0019791 van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39* (3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Vaughan L, Giovanello K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology of Aging*, *25* (2), 343–55. https://doi.org/10.1037/a0017729.

Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (4), 1012–1028. https://doi.org/10.1037/a0015859

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40* (4), 1072–1095. doi.org/10.1037/a0035779

Vīnerte, S., & Sabourin, L. (2019). Reviewing the bilingual cognitive control literature: Can a brain-based approach resolve the debate? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *73* (2), 118–134. https://doi.org/10.1037/cep0000174

Vogel, E. K., McCullough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438* (7067), 500–503. https://doi.org/10.1038/nature04171

von Bastian, C. C., De Simoni, C., Kane, M., Carruth, N., & Miyake, A. (2017). *Does being bilingual entail advantages in working memory? A meta-analysis*. Presented at the 58th Annual Meeting of Psychonomic Society, Vancouver.

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, *145* (2), 246–258. https://doi.org/10.1037/xge0000120

Vygotsky, L. S. (*1986*). *Thought and language*. Cambridge, MA MIT Press.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: PRO-ED.

Ware, A. T., Kirkovski, M., & Lum, J. A. G. (2020). Meta-Analysis Reveals a Bilingual Advantage That Is Dependent on Task and Age. *Frontiers in Psychology*, *24*, 1411–1458. https://doi.org/10.3389/fpsyg.2020.01458.

Warmington, M., Kandru-Pothineni, S., & Hitch, G. (2019). Novel word learning, executive control and working memory: A bilingual advantage. *Bilingualism: Language and Cognition*, *22*. https://doi.org/10.1017/S136672891800041X.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13* (6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Weaver, B., Bédard, M., & Mcauliffe, J. (2013). Evaluation of a 10-minute Version of the Attention Network Test. *The Clinical neuropsychologist*, *27*. 10.1080/13854046.2013.851741.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t15170-000

Wickelgren, W. A. (1965). Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *4* (1), 53–61. https://doi.org/10.1016/S0022-5371(65)80067-6

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44* (2), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54* (8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Willcutt, E. G., Pennington, B. F., Boada, R., Ogline, J. S., Tunick, R. A., Chhabildas, N. A., & Olson, R. K. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *110* (1), 157–72. https://doi.org/10.1037//0021-843x.110.1.157.

Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Wiseheart, M., Viswanathan, M., & Bialystok, E. (2016). Flexibility in task switching by monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *19* (1), 141–146. https://doi.org/10.1017/S1366728914000273

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, *19* (5), 1024–1039. https://doi.org/10.1037/0278-7393.19.5.1024

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson® III Test*. Riverside Publishing. Company. Itasca, IL.

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In W. Damon (Ed.), *Handbook of child psychology*: Vol. 2. Cognition, perception, and language (pp. 371–420). John Wiley & Sons Inc.

Wright, R. W., Brand, R. A., Dunn, W., & Spindler, K. (2007). How to write a systematic review. *Clinical Orthopaedics and Related Research*, *455*. 23-29. https://doi: 10.1097/BLO.0b013e31802c9098.

Xu, J., & Metcalfe, J. (2016). Studying in the region of proximal learning reduces mind wandering. *Memory & Cognition*, *44* (5), 681–695. https://doi.org/10.3758/s13421-016-0589-8

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, *2* (210). https://doi.org/10.3389/fpsyg.2011.00210

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 293–357). Lawrence Erlbaum Associates Publishers.

Zerr, C. L., Berg, J. J., Nelson, S. M., Fishell, A. K., Savalia, N. K., & McDermott, K. B. (2018). Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. *Psychological Science*, *29*, 1436–1450.

Zimmer, H. D. (2008). Visual and spatial working memory: From boxes to networks. *Neuroscience and Biobehavioral Reviews*, *32* (8), 1373–1395. https://doi.org/10.1016/j.neubiorev.2008.05.016