The effect of bilingualism on lexical learning and memory across two language modalities: Some evidence for a domain-specific, but not general, advantage

David Martinez & Jenny L. Singleton
School of Psychology, Georgia Institute of Technology, Atlanta, USA

This is a preprint of an article published by Taylor & Francis in the Journal of Cognitive Psychology available online at https://www.tandfonline.com/doi/abs/10.1080/20445911.2019.1634080

Martinez, D. & Singleton, J.L. (2019) The effect of bilingualism on lexical learning and memory across two language modalities: some evidence for a domain-specific, but not general, advantage, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2019.1634080

Address for correspondence
David Martinez, Applied Research Laboratory for Intelligence and Security (ARLIS), University of Maryland, College Park, 7005 52nd Ave, College Park, MD 20742.
DMartin5@umd.edu

Abstract

The present study was conducted to replicate bilingual advantages in short-term memory for language-like material and word learning in young adults and extend this research to the sign domain, ultimately with the goal of investigating the domain specificity of bilingual advantages in cognition. Data from 112 monolingual hearing non-signers and 78 bilingual hearing non-signers were analysed for this study. Participants completed a battery of tasks assessing sign and word learning, short-term memory, working memory capacity, intelligence, and a language and demographic questionnaire. Overall, the results of this study suggested a bilingual advantage in memory for speech-like material—no other advantage (or disadvantage) was found. Results are discussed within the context of recent large-scale experimental and meta-analytic studies that have failed to find bilingual advantages in domain-general abilities such as attention control and working memory capacity in young adults.

Keywords: bilingual advantage, working memory capacity, short-term memory, sign language, lexical learning

Introduction

Learning and knowing a second language confers many benefits—economic (Agirdag, 2014; Rumbaut, 2014; Saiz & Zoido, 2005; Shastry, 2012), social (Morgan, 1993; Nelson, 1968), and academic (Cooper, 1987; Cooper et al., 2008; Dangiulli, Siegel, & Serra, 2001)—but recently there have been concerns over reputed *cognitive* benefits, particularly with regard to bilingual advantages in attention control (e.g., de Bruin, Treccani, & Della Sala, 2015; Hilchey & Klein, 2011; Paap, Johnson, & Sawi, 2015) and related abilities, such as working memory capacity (e.g., Ratiu & Azuma, 2015) in young adults. Many published investigations into the matter have observed a bilingual advantage (see Adesope, Lavin, Thompson, & Ungerleider, 2010; Grundy & Timmer, 2017), however, recent studies (Duñabeitia et al., 2014; Paap & Greenberg, 2013) and meta-analyses (Lehtonen et al., 2018; von Bastian, De Simoni, Kane, Carruth, & Miyake, 2017), have observed weaker and, in some cases, nonsignificant results (though see the following for arguments that a bilingual advantage does manifest in young adults as cortical reorganization and in older adults behaviourally: Bialystok, 2017; Bialystok & Grundy, 2018).

While the effect of bilingualism on executive functioning in young adults is equivocal, bilingualism does appear to enhance two interrelated constructs: lexical learning (Bartolotti & Marian, 2012; Kaushanskaya & Marian, 2009a; Kaushanskaya & Rechtzigel, 2012; Kaushanskaya, Yoo, Van Hecke, & Oetting, 2013; Papagno & Vallar, 1995; van Hell & Candia Mahn, 1997) and phonological short-term memory (Kaushanskaya, 2012; Lehtonen et al., 2018; Papagno & Vallar, 1995; Yoo & Kaushanskaya, 2012). Lexical learning refers to learning of verbal material and is often indexed in the laboratory using associative learning tasks in which an unfamiliar item is paired with a familiar item or by presenting unfamiliar lexical items in context, such as in a reading passage—after some amount of time for study, participants are tested on the target lexical items. Phonological short-term memory (STM) is short-term memory for verbal material and is often indexed by tasks such as digit and letter span (e.g., Gupta, 2003; Kaushanskaya, Blumenfeld, & Marian, 2011).

The present study was undertaken to replicate the bilingual advantage in word learning and spoken phonological STM and extend this research to *sign* learning and *signed* phonological STM, ultimately, with the goal of furthering our understanding of the effect of bilingualism on cognition. As will be reviewed below, a number of hypotheses have been proposed to explain these bilingual advantages but a coherent theory has yet to emerge. Investigating the effect of bilingualism on lexical learning and phonological STM across language modalities can offer insight into the mechanisms underlying purported bilingual advantages. For example, if the bilingual advantage in word learning and spoken phonological STM generalises to the sign domain, then it suggests that bilingualism enhances domain-general (e.g., attention control) processes; on the other hand, if the bilingual advantage does not transfer, then the effects are domain-specific, i.e., specific to spoken language processes (for similar reasoning, see Emmorey, Luk, Pyers, & Bialystok, 2008). Other patterns can emerge with concomitant implications.

Background

An abundance of research has indicated that phonological STM and lexical learning are related, with the dominant view being that lexical learning is determined, in part, by phonological STM (Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006; Gupta & Tisdale, 2009; Majerus, Poncelet, Elsen, & van der Linden, 2006; Page & Norris, 2009). Though theories vary in their details, the general explanation for this relationship is that phonological

STM allows individuals to briefly maintain accurate phonological representations in mind while more durable representations are consolidated in long-term memory (Baddeley et al., 1998; Gupta, 2003; Norris, Page, & Hall, 2018)—thus, greater phonological STM capacity results in greater lexical learning ability (Hummel & French, 2016; Linck et al., 2013; Majerus, Poncelet, Van der Linden, & Weekes, 2008; K. I. Martin & Ellis, 2012; O'Brien, Segalowitz, Collentine, & Freed, 2006).

Importantly, phonological STM *performance* is affected by language experience (Baddeley, 1964; Bartolotti & Marian, 2017; Botvinick, 2005; Gathercole, 1995; Gathercole, Pickering, Hall, & Peaker, 2001; Hulme, Maughan, & Brown, 1991; Majerus, Martinez Perez, & Oberauer, 2012; Thorn & Frankish, 2005; Thorn & Gathercole, 1999; Woodward, Macken, & Jones, 2008), and, as a consequence, so is lexical learning (Bartolotti & Marian, 2017; Gupta & Tisdale, 2009; Majerus, Poncelet, et al., 2008; Nair, Biedermann, & Nickels, 2017; Storkel, Armbrüster, & Hogan, 2006). Language experience endows individuals with phonological, phonotactic, and semantic knowledge which supports phonological STM—possibly through slower trace decay rates resulting from more integrated and resonant STM and language processing networks (Kaushanskaya et al., 2013; Klein, Watkins, Zatorre, & Milner, 2006; Majerus, 2008, 2013; N. Martin & Saffran, 1992) and/or redintegration—in which partial representations and prior knowledge (e.g., expectancies) are used to reconstruct whole items (Bower & Glass, 1976; Schweickert, 1993). For example, it is much easier for monolingual English speakers to remember a list of names such as John, Mary, George, Chris, Leah than a list composed of Nahuatl (Aztec language) names such as Ahuatzi, Xochitl, Cuautemoc, Tenoch, and Yaretzi. Not only is a native English speaker likely to know or know of individuals with names such as John and Mary (providing semantic information), these names also follow familiar (phonotactic) patterns—the Nahuatl names on the other hand, do not. Accordingly, performance is greater when phonological STM tasks utilize known words rather than pseudowords (Hulme et al., 1991) and when pseudowords match the phonotactic constraints of a familiar language rather than an unfamiliar language (Botvinick, 2005; Gathercole, 1995). Similar effects have been found using lexical learning paradigms (Majerus, Poncelet, et al., 2008; Ordonez Magro, Attout, Majerus, & Szmalec, 2018; Storkel et al., 2006).

The fact that language experience is related to phonological STM and lexical learning appears to present a paradox for the bilingual advantage—after all, bilingual individuals tend to have less experience in their L2 than monolingual individuals and there is evidence that this leads to a bilingual *disadvantage* in measures of lexical access and retrieval (Bialystok, 2009; Lehtonen et al., 2018). However, it is important to note that lexical access and retrieval tasks are speeded tasks, while lexical learning and phonological STM, though they often have time limits, are not. Thus the effect of experience with a given language may be only a minor factor when bilinguals are given sufficient time. Second, it is important to distinguish the effect of bilingualism from other sources of group differences, such as socioeconomic status and age of acquisition. When studies are conducted that control for confounds, as stated above, bilinguals *do* tend to outperform monolinguals (Buac, Gross, & Kaushanskaya, 2016; Delcenserie & Genesee, 2017; Hirosh & Degani, 2018; Pierce, Genesee, Delcenserie, & Morgan, 2017).

Explanations for the bilingual advantage in phonological STM and lexical learning are varied and, in the case of phonological STM, fairly scarce. One explanation for the bilingual advantage in phonological STM performance is that learning multiple languages is an enriching experience that leads to cognitive and neurological adaptations resulting in a more efficient phonological STM system (Pierce et al., 2017). Alternatively, the bilingual advantage may result indirectly from the reported bilingual advantage in the domain-general processes of *attention control*. Of the two theories, the latter is the most developed. Briefly, the purported bilingual advantage in attention control abilities is said to arise from bilingual individuals' need to manage two (or more languages); at any one time, a bilingual individual must control attention to select and process linguistic representations from one language while inhibiting representations from another (Green, 1998; Kroll, Van Hell, Tokowicz, & Green, 2010). The claim is that bilingualism *trains* one's ability to control attention (Bialystok, 2017). Phonological STM tasks (and STM tasks generally) place demands on the ability to control attention (Colom, Rebollo, Abad, & Shih, 2006; Cowan, 2008; Kane et al., 2004) and so bilingualism is believed to indirectly affect phonological STM through its effect on attention control.

With regard to the bilingual advantage in lexical learning, theories are much more plentiful and varied. Proposed explanations include superior attention control (Bartolotti & Marian, 2012; Kaushanskaya, 2012; Yoo & Kaushanskaya, 2012); greater phonological STM capacity (Papagno & Vallar, 1995; van Hell & Candia Mahn, 1997); a more flexible and tolerant phonological system (Kaushanskaya & Marian, 2009a); activation of a richer semantic network (Kaushanskaya & Rechtzigel, 2012); improved ability to manage interference resulting from differences in sound-symbol mappings across languages (Kaushanskaya & Marian, 2009b); and differences in lexical learning strategies (Kaushanskaya & Rechtzigel, 2012; Kaushanskaya et al., 2013).

It is beyond the scope of this article to detail the various explanations that have been offered (for a review of the literature, see Hirosh & Degani, 2018), however, it is important to comment on the variety of theories proposed. This variety is partly the result of the complex nature of bilingualism, the circumstances that lead individuals to learn languages to varying degrees of proficiency, and the choices that researchers make in classifying individuals and accounting for confounds (de Bruin, 2019; Grosjean, 2003; Kaushanskaya & Prior, 2014). For example, whether an individual learns their second language as a child or an adult, or in the classroom rather than at home, may affect ultimate L2 proficiency as well as whether an individual develops explicit strategies to support language learning. A separate but contributing issue is that the lexical learning and phonological STM tasks that have been used across studies also vary considerably. While this can aid in delineating boundary conditions, it can make it difficult to infer general principles, as researchers are unable to identify whether an effect materialized (or did not) because of task characteristics or some other feature of the study. This issue is compounded by the inherently diverse nature of bilingualism, the small sample sizes typically used in research examining phonological STM and lexical learning in bilinguals, and because typically, only a few indices of any given construct are measured within a single study (for similar arguments, see de Bruin et al., 2015; Paap et al., 2015; Paap & Sawi, 2014). Of course, the variety in theories does not speak to their veracity: more than one of these explanations can be valid. Nevertheless, the state of the field is such that it is not clear which theory and under what conditions a specific theory or theories satisfactorily account for bilingual advantages in lexical learning and phonological STM. Moreover, there is serious doubt whether bilingualism has *any* effect on domain-general processes (e.g., Lehtonen et al., 2018).

The Present Study

The present study was part of a larger study that was undertaken to investigate the relationships between sign learning and word learning (Martinez & Singleton, 2018a). Given the diversity of our university and of the Atlanta community (where the study took place), we expected a large number of our participants would be bilingual, thus, a priori, we planned to use data from the primary study to replicate prior research on word learning and spoken phonological STM and to extend it to the sign domain. Investigating whether bilingual advantages extend to the sign domain allows us to explore the domain specificity of the advantage, as sign language processing relies on many of the same processes as spoken language processing (e.g., working memory, attention, long-term memory retrieval), with those differences that are identified generally being restricted to perceptuomotor processes, as evidenced by behavioural and imaging studies (Bavelier et al., 2008; Campbell, MacSweeney, & Waters, 2008; Emmorey, 2002; Rönnberg, Rudner, & Ingvar, 2004; Rudner, Andin, & Rönnberg, 2009). As such, bilingual advantages that generalise across tasks and language modality imply a more general effect. Additionally, there is theorizing (Baddeley, 2012, 2015) and some evidence from imaging studies that spoken- and signed phonological STM rely on similar processes (Bavelier et al., 2008; Newman-Norlund, Frey, Petitto, & Grafton, 2006; Rudner et al., 2009; J. T. Williams, Darcy, & Newman, 2015), thus if there is a bilingual advantage in spoken phonological STM there may be an advantage in signed phonological STM, providing further evidence for amodal phonological STM processes.

The investigation of effects across language modalities is not novel. For example, Emmorey et al. (2008) investigated whether bimodal (sign/spoken language) bilingual individuals, such as those born to deaf adults, also demonstrate bilingual advantages. The results of their study indicated that bimodal bilingualism does *not* confer the same advantages as unimodal bilingualism. Emmorey et al. (2008) argued that unimodal bilingualism enhances cognitive control because unimodal bilinguals "are constantly faced with more challenging production demands because their languages utilize the same articulatory system….Bimodal bilinguals do not face the same processing demands [as they can speak and sign somewhat simultaneously], and thus do not show the same enhanced performance on executive control…(p. 5).”

In addition to incorporating sign tasks as an extension on prior work and as a check on the generality of any conclusions that are drawn from the present study, some of the issues raised above were also addressed within this study. First, bilingual and monolingual participants were matched on a number of covariates and a related variable was investigated as well. Specifically, participants were matched on age, sex, education level, parental income, and fluid and crystallized intelligence. Additionally, we included measures of working memory capacity to aid in localizing the effect of bilingualism. Working memory capacity tasks rely on many of the same processes as STM tasks, however, they tend to be much more domain-general, relying on attention control to a greater extent than STM tasks which tend to be more domain-specific (Cowan, 2008; Kane et al., 2004; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Second, the sample size was larger than is typically observed in studies of this nature. For example, the range in total sample size in a selection of phonological STM studies was between 20 and 54 participants (Kaushanskaya, 2012; Papagno & Vallar, 1995; Yoo & Kaushanskaya, 2012); for lexical learning it was between 20 and 76 (Bartolotti & Marian, 2012; Kaushanskaya & Marian, 2009a; Kaushanskaya & Rechtzigel, 2012; Kaushanskaya et al., 2013; Papagno & Vallar, 1995; van Hell & Candia Mahn, 1997). In this study, there were a total of 190 participants, with 78 classified as bilingual. Finally, each construct was measured with at least three tasks, allowing us to investigate patterns of relationships within and across constructs. Of course, this is not to say that other problems do not arise from our approach—see the section on limitations below.

Method

Participants

Participants were recruited from the Georgia Tech School of Psychology subject pool as well as from the surrounding community. Georgia Tech participants received course credit and an additional $15 if they completed the entire study; community participants received a total of $65.

Young adults between the ages of 17-35, fluent in English, and who had normal or corrected-to-normal hearing and vision were recruited for this study. Due to the nature of the tasks and the aims of this study, participants were excluded if they indicated fluency in American Sign Language or Turkish (two languages used in the present study), were diagnosed with a language disorder, or if they possessed an upper-body injury or movement disorder affecting their arms or hands, as this could affect their ability to rehearse signs.

In total, 286 individuals consented to participate in the study. Of these individuals, some did not complete the entire study (n = 34), others indicated poor English fluency or did not answer questions pertaining to their language history (n=8), and three individuals were removed from analysis because they did not follow instructions, leaving 241 participants for consideration. As will be detailed below, participants were then grouped into bilingual, monolingual, and intermediate groups; analyses investigating bilingual advantages were only conducted with the monolingual (n=112) and bilingual (n = 78) groups. The data are available at osf.io/xmype.

Procedure

The study consisted of two sessions. Participants worked individually in a room with up to five participants. Nearly all tasks were completed on a PC running E-Prime software (Schneider, Eschman, & Zuccolotto, 2002); only a reading test and demographic and language experience questionnaire were completed on paper.

The first session took participants up to 2.5 hours to complete while the second session took approximately 2 hours. Task administration order was fixed and is presented in Table 1; task descriptions follow. Note, the OSIQ is not relevant to the present study and will not be discussed.

Tasks

All tasks developed in-house (those administered during Session 1) are available at osf.io/xmype. Tasks always began with instructions and at least one example with feedback. A research assistant was always present to observe participants and offer help.

Sign Learning

All sign learning tasks utilized a similar associative learning paradigm with study and test blocks. During study blocks, unfamiliar stimuli were presented first followed by their associated English translations. During test blocks, stimulus items were presented followed by a screen revealing all the English words encountered in the task; participants were to select the appropriate translation; once a selection was made, then the next item was presented and so on. Importantly, pilot data suggested that there were differences in strategy use. Subsequently, participants were instructed to use imagery or sentence generation to aid their learning and examples of these strategies were given.

The number of study-test blocks varied depending on the task and also on the participant's performance—a task ended when the maximum number of trials was reached or when a participant correctly responded to 100% of the items in a trial, whichever came first. For the purpose of analysis, an individual was given full credit for all remaining trials after correctly responding to all items in a test block.

The unfamiliar items were either contrived (i.e., pseudosigns) or drawn from unfamiliar languages. The pseudosigns used in this study were drawn from a prior study investigating sign learning and signed phonological STM (Martinez & Singleton, 2018b). The English translations utilized in these tasks were selected from the SUBTLEX-US corpus (Brysbaert & New, 2009; Brysbaert, Warriner, & Kuperman, 2014; New, Brysbaert, Veronis, & Pallier, 2007) and were familiar concrete nouns ranging between 1-3 syllables and 4-6 characters in length. Familiar words were used to reduce the role of verbal ability and to mimic what adults typically encounter when they first attempt to learn a new language.

In the *American Sign Language (ASL) sign learning (ASL-SL) task*, participants attempted to learn 24 ASL signs over two study-test blocks. ASL signs were drawn from the ASL-LEX database (Caselli, Sehyr, Cohen-Goldberg, & Emmorey, 2016) such that the ASL signs were low in iconicity, signs were visually distinct, and their English glosses conformed to the specifications listed above. A hearing native ASL signer performed all of the signs and the same video clips were used for both study and test trials. The maximum score was 48.

In the *pseudosign learning (PSL) task*, participants attempted to learn 15 pseudosigns over 4 trials. This task differed from the ASL-SL task in a number of ways. First, a dropout procedure was used—once a participant correctly responded to an item, it was no longer viewed in remaining study or test blocks. Second, different sign models performed the pseudosigns in the study and test blocks (see Fig. 1). Third, and most obviously, this task utilized pseudosigns rather than signs from a signed language, bestowing greater control over item characteristics. Because most individuals mastered all items prior to the fourth test block, scores were calculated as the total achieved by the third test block, thus the maximum possible score was 45.

The *delayed pseudosign learning (DPSL) task* consisted of a single trial of the PSL task, administered after four intervening tasks (see Table 1), approximately 30 min after the PSL task. Scores on this task were calculated as a proportion of the number of items correctly identified after four PSL study-test blocks. Had scores not been calculated in this way, then the relationship between the PSL (a measure of initial learning) and DPSL (a measure of retention) would have been artificially high.

The *three-term sign learning (3TSL) task* is a complex associative task adapted from B. A. Williams and Pearlberg (2006). In this task, pseudosigns are associated with three response words which are themselves associated with three number cues. During testing, participants are presented with a stimulus and number cue and must identify the correct response word. For example, stimulus pseudosign *P* is associated with *1-tree*, *2-bone*, and *3-fork*. At test, a participant would be asked to identify the correct response word given a particular pseudosign-number combination (e.g., for *P-2,* the correct response is *bone*). Participants had three study-test blocks to learn six pseudosigns and their corresponding response words, thus the maximum score was 54 (6 pseudosigns x 3 response words x 3 blocks).

Word learning tasks

The word learning tasks were very similar to the sign learning tasks with the exception that spoken language stimuli were used. Unfamiliar words were either drawn from Turkish, a Turkic language distinct from Indo-European languages (e.g., Spanish, English), or were contrived. All pseudowords conformed to English phonotactic constraints and were selected from the English Lexicon Project (Balota et al., 2007). All pseudowords and their characteristics are provided in Appendix A.

The Turkish word learning (TWL) task was analogous to the ASL-SL task—participants had three study-test blocks to learn 15 Turkish words. The Turkish words were spoken by a native Turkish speaker and presented over headphones.

The *pseudoword learning (PWL) and delayed pseudoword learning (DPWL) tasks* were analogous to the PSL and DPSL tasks, respectively. In the PWL task, participants had up to 6 study-test blocks to learn 15 pseudoword-English pairs, however, scores were calculated using the first 4 trials as participants tended to master the items by this point and further sets increased skew. A dropout procedure was used and two different females produced the study and test pseudowords. The DPWL consisted of a single test block of the PWL, administered after four intervening tasks, approximately 40 min after the PWL. DPWL scores were calculated as a proportion of the PWL pseudoword-word pairs learned over the six PWL study-test blocks.

Finally, the *three-term word learning (3TWL) task* was analogous to the 3TSL. Participants had up to three study-test blocks to learn six pseudowords and their three English words and associated number cues.

Signed phonological STM tasks

All phonological STM tasks were either span tasks or discrimination tasks. In the span tasks, participants saw a series of items and were required to reproduce the series by selecting items in the same order as presented. Span tasks were scored using the partial credit unit scoring procedure in which performance on a trial was calculated as a proportion of the total items correctly recalled in their serial position (for details, see Conway et al., 2005, pp. 775-777). In the discrimination tasks, participants judged whether a target item or sets of items were the same or different reproductions. Participants received credit for correct judgments and scores were calculated by summing across all trials.

In the *nonsign paired task (NSPT),* participants judged whether target pseudosigns differed from reproductions (see Fig. 2). Because participants were non-signers, prior to the critical trials, participants watched a brief instructional video introducing them to the task and the parameters (i.e., handshape, movement, or orientation of a sign) that may change in the reproduction; they then completed two blocks of three practice trials with feedback. Next, they completed two blocks of critical trials, each consisting of 28 trials. The same target pseudosigns were used across the two blocks, however, two different research assistants "attempted" to reproduce the target signs. Within each block, half of the pseudosigns were faithful reproductions and the other pseudosigns differed from the target in handshape, movement, or orientation. The maximum possible score was 56.

In the *probed sign (ProSign) task*, participants viewed 40 sets of three to six pseudosigns, each followed by a single probe pseudosign. Participants were to indicate whether the probe was the same as one of the pseudosigns presented in the set (see Fig. 3). If the probe was different, it differed from one of the pseudosigns presented in the set by one parameter: handshape, movement, or orientation. Within a set, items differed from each other on at least two of the aforementioned parameters. The maximum possible score was 40.

In the *sign configuration (SignCon) task*, participants attempted to recall one to four pseudosigns while simultaneously retaining a set of letters. Participants began each trial by viewing a set of letters—set size varied by participant and was dependent on their letter span, which was calculated from the LetSpan task described below (and completed earlier in the study session, see Table 1). Next, participants viewed one to four pseudosigns which, when set size was greater than one, differed from each other by at least two of the following: handshape, movement, and/or location. After this, participants' memory was tested; on 40% of trials, participants were tested on the letters, on the remaining 60%, participants were tested on the pseudosigns (see Fig. 4). The critical portion of the task was the pseudosign span portion, however, individuals tend to use verbal mediation when confronted with visuospatial information (e.g., giving a verbal label to a sign) and so asking participants to retain sets of letters was our way of suppressing verbal mediation and articulatory rehearsal. Like the other tasks, we could have chosen to use highly similar items which render verbal strategies ineffective, however, in span tasks, this has the effect of increasing interference and the role of attention control mechanisms (see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). There were a total of 20 trials, 12 of which were pseudosign trials, with three trials at each set length; using partial credit unit scoring, the maximum possible score was 12.

Spoken phonological STM tasks

The spoken phonological STM tasks were also either span tasks or discrimination tasks. Span tasks were scored using the partial credit unit scoring procedure. Two tasks used pseudowords; these pseudowords and, when available, their characteristics are listed in Appendix B.

In the *letter span (LetSpan) task*, participants attempted to recall four to nine letters in serial order. There were three trials at each set length and all letters in a set were presented on screen for a period equal to 500 ms per letter. The maximum possible score was 18.

In the *nonword recognition (NWRec) task*, participants judged whether a reproduction of a sequence of pseudowords, presented via headphones, was the same or different from an initial presentation; if the sequence was the same, then the same audio clip was used; if it was different, then two neighbouring pseudowords were transposed. The same 18 target sequences were used twice, once during a “same” trial and once during a “different” trial for a total of 36 trials, with four trials of set length three, six trials at set length four, and eight trials at set length five. Pseudowords were drawn from a pool of 28 items and were selected from Gathercole et al. (2001). The maximum score was 36.

The last spoken phonological STM task, the *nonword span (NWSpan) task* required participants hear a set of monosyllabic pseudowords and attempt to recall them in the order presented. The pool of items consisted of 12 pseudowords drawn from Gathercole et al. (2001). Pseudowords were presented in sets ranging between two and six and there were three trials at each set length for a total of 15 trials. Using partial credit unit scoring, the maximum score was 15.

Crystallized intelligence tasks

Crystallized intelligence (Gc) is defined as one’s stock of declarative knowledge (Cattell, 1943). There were four measures of Gc, *The information (Info) test*, the *extended range vocabulary (Vocab) test*, the *grammar and usage (Gram) test*, and the *reading comprehension (Reading) test*.

The Info test consisted of two parts. In the first part, participants had up to 7 min to answer 40 general knowledge multiple-choice questions from the Information subscale of the Multidimensional Aptitude Battery II (Jackson, 1998). In the second part, participants had up to 2 min to answer an additional 11 multiple-choice questions that were developed in-house and intended to broaden the domains of knowledge tested. The maximum possible score was 51.

In the Vocab test, participants matched a given word to one of five choices, whichever was a synonym or near-synonym. There were two parts, each with 24 questions and a time limit of 6 min. The maximum possible score was 48.

The Gram test consisted of 21 “improving sentences” items from SAT practice tests released between 2004 and 2013. The participant was to indicate whether a portion of a sentence could be improved and if so, they were to select the best answer choice out of four. Participants were given up to 10 min to complete the test and the maximum possible score was 21.

In the last Gc task, Reading, Participants had up to 20 min to answer 17 reading comprehension questions, corresponding to 5 passages ranging between 112-739 words in length. All passages were selected from released SAT and GRE tests. The maximum possible score was 17.

Fluid intelligence tests

Fluid intelligence (Gf) is defined as the ability to solve novel problems and reason (Cattell, 1943). In *Raven's Advanced Progressive Matrices, Set II* (*Ravens*; Raven, Raven, & Court, 1998), participants viewed 18 3x3 matrices. Each cell in a matrix but the last contained figures, organized in a specific pattern; the participant was to select which of eight answer choices best completed the pattern. The time limit for this task was 10 min and the maximum possible score was 18.

In *letter sets* (*LetSets*; Ekstrom, French, Harman, & Dermen, 1976), participants attempted to identify one set of letters that did not follow the same pattern as the four other sets (e.g., all but one set showed letters organized in ascending order). Participants had up to 7 min to complete 30 problems. The maximum possible score was 30.

In *number series* (*NumSeries*; Thurstone, 1938), participants were presented with series of numbers that obeyed a specific rule (e.g., 3,5,7,11…); their task was to select a number that continued the series. There were 15 problems and participants were given up to 5 min to complete the task. The maximum possible score was 15.

The last Gf task was a static version of the *Spatial Learning Ability Test* (*SLAT*; Embretson, 1992). In this task, participants mentally rotated and folded figures to match them to completed cubes (see Fig. 5). Although this task is best considered a measure of general visual ability, tasks of this type tend to load strongly on Gf factors (Lohman, 1996; Marshalek, Lohman, & Snow, 1983; Varriale, van der Molen, & De Pascalis, 2018). There were 20 items and participants had up to 15 min to complete them. The maximum possible score was 20.

Working memory capacity (WMC) tasks

All WMC tasks were shortened versions of the complex span tasks described by Foster et al. (2015). In a complex span task, participants complete a primary memory task and a secondary processing task. The dependent variable is the number of items from the primary task that the participant is able to remember in correct order. As with the phonological STM span tasks described above, WMC tasks were also scored using partial credit unit scoring.

In the *operation span (OSpan) task*, participants attempted to maintain sets of letters in memory while assessing the veracity of simple math equations. Participants were tested on one trial each at set lengths from three to seven; using partial credit unit scoring, the maximum possible score was 5.

In the *symmetry span (SymSpan) task*, participants attempted to recall the sequence of locations of a red square in a 4x4 matrix while also judging whether a figure was symmetrical.

The set size (i.e., number of different locations) varied from two to five, and each set size was tested once. The maximum possible score was 4.

The last WMC task was the *rotation span (RoSpan) task*. In this task, participants attempted to maintain a sequence of arrows in memory while also verifying whether a rotated letter was mirrored or not. Set size ranged between two and five. The maximum possible score was four.

The demographic and language questionnaire (Questionnaire)

The questionnaire was administered in paper format. Participants volunteered their age, sex, handedness, ethnicity, race, level of education, and parent's income. They were also asked to identify their first language and, if they knew another language, state which was their dominant language. If they did not know any other language then they stopped after responding about their L1. If they did know another language, then they were asked to identify up to four languages, the age when they began learning the language, and rate their proficiency in speaking, listening, reading, and writing, using a scale from 1 ("very low") to 10 ("perfect").

Statistical Analyses

Data preparation and group assignment

First, for the cognitive tests, outliers and missing values were imputed using the expectation-maximization (EM) algorithm (Little & Rubin, 2014; Rubin, 1976) and the full sample of 241 participants. The data was then re-inspected for outliers, defined as scores that were over 3.5 standard deviations from the mean. Two scores, both for the OSpan, were below the cut-off and so they were replaced with the cut-off score.

Next, participants were categorized as bilingual or monolingual (few [n = 12] participants reported more than two languages, thus that information was discarded). To do this, participants were first categorized as early, late, or intermediary L2 acquirers and separately as novice, advanced, or intermediate bilinguals. Individuals were classified as early acquirers if they reported an age of onset of acquisition (AoA) of 7 or younger; late acquirers if they reported an AoA of greater than 13; and intermediary acquirers if they reported an AoA greater than 7 but equal to or less than 13. These ages were chosen because the critical or sensitive period for second language acquisition is often said to close sometime between childhood and puberty (DeKeyser, 2000; Johnson & Newport, 1989). Individuals were classified as novice bilinguals if they reported an average L2 proficiency score of less than 4 out of 10 (10 being "perfect"); advanced bilinguals if they reported an average score of 7 or greater; and intermediate bilinguals if they reported a score between those two. Each category was then assigned a score from 0-2, with zero being assigned to late acquirers and novice bilinguals, one assigned to the intermediate groups, and two assigned to early acquirers and advanced bilinguals. Finally, the scores assigned to each category were summed and those individuals with a score of 0 or 1 were classified as monolingual (n = 112) while those with scores of 3 or 4 (n = 78) were classified as bilinguals; individuals with scores of 2 (n = 51) were not included in any further analyses. See Table 2, below, for mean AoA and proficiency scores by group and appendix D for a list and tally of languages spoken by bilingual individuals.

Statistical procedure

The data were submitted to two analyses. In the first analysis, we considered the full set of monolingual and bilingual participants and were primarily concerned with group differences in sign learning, word learning, signed phonological STM, spoken phonological STM, and WMC; however, we also investigated whether groups differed on demographic variables as well as Gc and Gf using independent groups t-tests. As will be observed below, there were some demographic differences and, while not statistically significant, the bilingual group tended to outperform the monolingual group on the Gc and Gf tasks. After seeing these differences, we feared that these minor differences could sum or interact in such a way as to affect our results, thus, in order to investigate further, a second (post-hoc) analysis was conducted with a subset of *matched* participants.

Analysis 1: Full sample of monolinguals and bilinguals

Results

Means, standard deviations, and test statistics for the demographic and cognitive covariates are provided in Table 2. In order to quantify the differences between monolingual and bilingual groups, we also offer Cohen's *d* (Cohen, 1988). Appendix C displays means, dispersion statistics, and internal consistency coefficients for the full set of 190 participants. Moreover, in a prior study (Martinez & Singleton, 2018a), we conducted a confirmatory factor analysis with the cognitive constructs under consideration and found evidence of convergent and discriminant validity.

Using an alpha level of .05, the two groups did not significantly differ on any of the demographic variables except gender, in which 57% of monolinguals identified as female compared to 72% of bilinguals, $t(188) = 2.107$, $p = .037$, $d = .315$. (Note: for consistency, we use a t-test and cohen's *d*, but the more appropriate X^2 test revealed the same result). The two groups did not differ on Gc or Gf. As expected given our classification scheme, there were significant differences in AoA and all proficiency measures, with bilinguals beginning to learn an L2 earlier, $t(188) = 16.578$, $p < .001$, $d = 3.169$, and to a greater level of proficiency in speaking, $t(188) = 24.834$, $p < .001$, $d = 3.629$, listening, $t(188) = 28.342$, $p < .001$, $d = 4.135$, reading, $t(188) = 15.325$, $p < .001$, d = 2.225, and writing, $t(188) = 12.893$, $p < .001$, $d = 1.983$.

Next, we formed *z*-score composites by converting raw scores to *z*-scores and summing relevant scores (e.g., the sign learning composite was formed by summing across the four sign learning tasks) and then we submitted these scores to individual *t*-tests. Using a corrected alpha level of $\alpha/5$ or .01, bilingual individuals outperformed monolinguals on one composite, spoken phonological STM, $t(188) = -3.383$, $p = .001$, $d = .504$. An exploration of the individual tasks revealed that bilinguals outperformed monolinguals on all of the spoken phonological STM tasks (see Table 3). No other significant group differences were found on any of the other cognitive variables.

Analysis 2: Matched groups

For this analysis, participants were matched using the MatchIt software package (Ho, Imai, King, & Stuart, 2011) for the R programming language. MatchIt allows for matching based on propensity scores, or the probability that an individual would be assigned to one group or another, given a vector of covariates (Austin, 2011a; Rosenbaum & Rubin, 1983). When participants are matched using propensity scores, the distribution of covariates is similar across groups (Austin, 2011a).

Participants were matched using a logit model with nearest neighbour matching without replacement and a caliper width of 0.2 (Austin, 2011b). The covariates were age, gender, education level, parental income, and Gc and Gf scores.

Results

A total of 23 participants from the full set of bilingual and monolingual participants failed to provide information for at least one of the demographic variables and so they were removed from this analysis. Additionally, five bilingual individuals could not be matched with any of the monolingual participants and so they too were removed from further analysis, leaving 65 matched pairs.

Table 4 displays information for all covariates. The matching procedure was successful, with group differences being reduced on all demographic (age, sex, education, parental income) and cognitive covariates. Average Gc scores for the monolingual and bilingual groups were .5 and .52, and average Gf scores were .59 and .60, respectively. Using an alpha level of .05, the groups did not significantly differ on any of the covariates.

The groups did significantly differ on AoA, $t(128) = -12.384$, $p < .001$, $d = -3.048$, and the L2 proficiency measures—speaking, $t(128) = 21.265$, $p < .001$, $d = 3.737$, listening $t(128) = 23.471$, $p < .001$, $d = 4.116$, reading, $t(128) = 12.492$, $p < .001$, $d = 2.193$, and writing, $t(128) = 11.824$, $p < .001$, $d = 2.119$—with bilinguals beginning to learn an L2 at a younger age and to a greater level of proficiency than monolinguals.

Table 5 displays the results of the analyses on the variables of interest. Using a corrected alpha level of .01 or $\alpha/5$, the groups did not differ on any of the composite measures. We note, however, that the largest and most consistent differences in performance (assessed via Cohen's *d*) were in the spoken phonological STM tasks and, accordingly, the spoken phonological STM composite.

Discussion

This study was undertaken to 1) replicate prior research exhibiting bilingual advantages in spoken phonological STM (Kaushanskaya, 2012; Papagno & Vallar, 1995; Yoo & Kaushanskaya, 2012) and word learning (Hirosh & Degani, 2018) in young adults and 2) investigate whether bilingual advantages exists in signed phonological STM and sign learning in hearing non-signing bilingual young adults. To that end, data from 190 participants (78 bilinguals) were analysed. An analysis of the full data indicated that bilinguals significantly outperformed monolinguals in spoken phonological STM and only those measures—we did not observe significant differences in word learning, sign learning, or WMC. A second analysis was conducted to ensure that our participants were precisely matched on confounding variables. This analysis did not reveal any significant differences, however, the largest differences were found on the spoken phonological STM tasks, with bilinguals tending to outperform monolinguals (though we reiterate, these differences were not statistically significant at a corrected alpha level of .01). Consequently, and in line with recent research (Duñabeitia et al., 2014; Lehtonen et al., 2018; Paap & Greenberg, 2013), we cannot infer an effect of bilingualism on domain-general processes but there was *some* evidence for a domain-specific bilingual advantage.

A Bilingual Advantage in Spoken phonological STM?

Assuming there is a bilingual advantage in spoken phonological STM and that it is not due to domain-general processes then what might it be due to? A very tentative speculation is that the bilingual advantage in spoken phonological STM was due to a general enhancement in *articulatory fluency*. In general, the faster an individual is able to rehearse items in a phonological STM task, the better one's performance (Baddeley, Thomson, & Buchanan, 1975; Ellis & Hennelly, 1980; Wilson & Emmorey, 1998; Woodward et al., 2008). One factor that affects rehearsal rate is co-articulatory skill. Co-articulation refers to the articulation of two successive phones, resulting in one phone assimilating some features of the other (Ohala, 1993). Within any given language, some sequences of phones co-occur more often than others, either within words or at boundaries between words—those that occur most often are more familiar and well-practiced, enhancing the rate at which they are articulated (Woodward et al., 2008). Phonological STM tasks, however, consist of arbitrarily sequenced phonemes that are less likely to co-occur (e.g., all of the items in the letter span used here were consonants). Bilingual individuals, by virtue of being familiar with two languages, are likely exposed to a broader range of phonetic sequences and thus may have a general enhancement in co-articulatory skill, resulting in faster articulation and better phonological STM performance compared to monolinguals.

This argument accounts for the fact that the largest group differences were observed for the letter span task, as it provided the most opportunities for co-articulation effects to accumulate. In the letter span task, set size ranged between four and nine items and the average number of items an individual could perfectly recall over three trials, i.e., their span, was 5.43. For the nonword recognition (NWRec) and nonword span (NWSpan) tasks, the range in set size was more restricted and the average span in the nonword span task was quite low (viz., 2.40). Consequently, there were more transitions between items in the letter span task and it is in these transitions that the effects of co-articulation are most prominent, at least when items are monosyllabic and obey the phonotactic constraints of the dominant language (Woodward et al., 2008), as they were here. We can also make predictions based on this interpretation. For example, we would expect that the greater distance between an individual's L1 and L2 phonology, the greater the effect; conversely, the greater the similarity between the L1 and L2, the smaller the effect[1].

Again, the above is post-hoc speculation and other explanations are possible. It is possible that the bilingual advantage observed here is a statistical artefact. Indeed, though our sample size was larger than typically observed for studies of this type, it was still fairly small considering the number of tests conducted and that the magnitude of bilingual effects are

[1] We thank an anonymous reviewer for this suggestion.

typically weak to moderate (Kaushanskaya, Gross, & Buac, 2014; Kaushanskaya & Marian, 2009a; Lehtonen et al., 2018). Another possibility is that our sample of bilinguals may have had a natural talent for memorizing verbal material. The vast majority of our bilingual participants listed English as their dominant language (see Appendix D) but, per our selection process, were also proficient in another language. Given the link between phonological STM and language learning (Baddeley et al., 1998), it is possible that these individuals were capable of learning two languages to a high degree *because* of their superior phonological STM ability. In fact, neuroimaging studies investigating differences between low and high proficiency bilinguals—that is, individuals exposed to two languages but who differ in their mastery of the second language—have found differences in areas of the brain related to phonological STM, with higher proficiency bilinguals showing greater efficiency in these areas (Chee, Soon, Lee, & Pallier, 2004; Majerus, Belayachi, et al., 2008). Whether greater phonological STM is a cause or result of L2 proficiency may be best answered by a large-scale longitudinal study.

The absence of a lexical learning advantage

We fully expected to replicate the bilingual advantage in word learning and were surprised when we failed to obtain significant differences on any of the word learning tasks, especially after finding a significant difference in spoken phonological STM, an established predictor of word learning (Baddeley et al., 1998). A re-review of the literature, however, revealed a significant methodological difference between this study and many others: we instructed participants to use elaborative rehearsal strategies while many others have either instructed participants to use rote rehearsal (e.g., Kaushanskaya, 2012; Kaushanskaya & Marian, 2009a, 2009b) or, to our knowledge, did not provide any instruction regarding encoding strategy (e.g., Kaushanskaya & Rechtzigel, 2012; Kaushanskaya et al., 2013).This is significant because rote rehearsal is presumed to depend heavily on phonological STM processes (Baddeley, Papagno, & Vallar, 1988; Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992) while elaborative techniques primarily depend on semantic processes (Kyllonen & Tirre, 1988; Kyllonen, Tirre, & Christal, 1991). Given that many individuals believe that rote rehearsal is just as effective as elaborative rehearsal (Dunlosky & Hertzog, 2000; Shaughnessy, 1981) and, additionally, that elaborative rehearsal is more effortful than rote rehearsal (Craik & Lockhart, 1972; Kee, 1994; Nairne, 1983), it is likely that *monolingual* participants will opt to use rote rehearsal if not instructed otherwise. In contrast, *bilingual* individuals, especially those who learned an L2 in a more explicit fashion (e.g., in a classroom), may have sufficient experience with the learning paradigm used in this and many other studies—L2 associative lexical learning—to recognize the benefits of elaborative rehearsal strategies and to be adept at employing them (for a similar argument, see Kaushanskaya, 2012, p. 485). As a result, when not explicitly instructed to use rote rehearsal, bilinguals may be more likely to exploit elaborative rehearsal strategies than monolinguals. Further research investigating strategy use amongst bilinguals and the effect of various strategies on word learning and spoken phonological STM performance should be conducted.

Limitations

While this study employed a larger sample and number of variables than is typical for this kind of work, there are limitations. First, we asked participants to self-report L2 proficiency rather than use an objective measure and thus some individuals likely underestimated their proficiency while others overestimated, resulting in increased error. Although there is evidence of a positive relationship between self-assessments and objective measures (Ackerman & Wolman, 2007; Mabe & West, 1982) and we used an extreme-group design that should have mitigated some of this error, it is still a limitation that may affect the results of this study. Indeed, it is becoming increasingly common for researchers to use objective measures of bilingualism, such as tests of L2 verbal knowledge, and to use these continuous variables in their analyses rather than categorical variables derived from questionnaires (e.g., Luk & Bialystok, 2013). Second, we conducted two sets of analyses with the same data and at least one of our conclusions—that there is a bilingual advantage in spoken phonological STM—primarily relies on the first analysis (with the full set of data). Our justification is that the second (matched) analysis reveals the same patterns of results but, because it is on a subset of data, it suffers from a lack of power. Thus we feel that the results of the first analysis are accurate and are supported by the results of the matched subsample. Still, a larger matched sample would offer stronger evidence. A third issue is that because this study uses data from a study that was primarily intended to answer a different question, the sample and task characteristics may not be properly controlled for this particular study. On the one hand, we see it as a strength that our sample included a diversity of L2s and tasks but, on the other hand, we agree that this diversity may come at the expense of experimental control. Again, further investigation is warranted.

Conclusion

In summary, we interpret our results as suggesting a possible bilingual advantage in short-term memory for spoken language material in young adults; however, we found no evidence of a bilingual advantage (or disadvantage) in short-term memory for signed material, sign learning, word learning, or working memory capacity. To the extent that there is a bilingual advantage in short-term memory for spoken material, it appears to be domain-specific.

Author Note

DM is now at the Applied Research Laboratory for Intelligence and Security (ARLIS), University of Maryland, College Park. JLS is now at the Department of Linguistics, University of Texas at Austin. This research was supported by National Science Foundation award #1651115. We would like to thank Maya Berinhout, Stephanie Blough, Lindsey Drummond, Morgan Foreman, Skyler Tordoya-Henckell, Kevaghn Hinckley, Sidni Vaughn Justus, Rachel Monahan, Elizabeth Pierotti, and Angelique Soulakos for their support on various aspects of this project. We would also like to thank the anonymous reviewers for their feedback.

Disclosure Statement

There was no potential conflict of interest.

REFERENCES

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied, 13*(2), 57-78. doi:10.1037/1076-898X.13.2.57

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research, 80*(2), 207-245.

Agirdag, O. (2014). The long-term effects of bilingualism on children of immigration: Student bilingualism and future earnings. *International Journal of Bilingual Education and Bilingualism, 17*(4), 449-464. doi:10.1080/13670050.2013.816264

Austin, P. C. (2011a). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi:10.1080/00273171.2011.568786

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150-161. doi:10.1002/pst.433

Baddeley, A. D. (1964). Immediate Memory and the "Perception" of Letter Sequences. *Quarterly Journal of Experimental Psychology, 16*(4), 364-367. doi:10.1080/17470216408416394

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D. (2015). Working Memory in Second Language Learning. In Z. Wen, M. B. Mota, A. McNeill, M. Bunting, & R. Engle (Eds.), *Working Memory in Second Language Acquisition and Processing* (pp. 17-28). Bristol, England: Multilingual Matters.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. doi:10.1037/0033-295X.105.1.158

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*(5), 586-595. doi:10.1016/0749-596X(88)90028-9

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior, 14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459. doi:10.3758/bf03193014

Bartolotti, J., & Marian, V. (2012). Language Learning and Control in Monolinguals and Bilinguals. *Cognitive Science, 36*(6), 1129-1147. doi:10.1111/j.1551-6709.2012.01243.x

Bartolotti, J., & Marian, V. (2017). Bilinguals' Existing Languages Benefit Vocabulary Learning in a Third Language. *Language Learning, 67*(1), 110-140. doi:10.1111/lang.12200

Bavelier, D., Newman, A. J., Mukherjee, M., Hauser, P., Kemeny, S., Braun, A., & Boutla, M. (2008). Encoding, rehearsal, and recall in signers and speakers: Shared network but differential engagement. *Cerebral Cortex, 18*(10), 2263-2274. doi:10.1093/cercor/bhm248

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11. doi:10.1017/S1366728908003477

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*(3), 233-262. doi:10.1037/bul0000099

Bialystok, E., & Grundy, J. G. (2018). Science does not disengage. *Cognition, 170*, 330-333. doi:https://doi.org/10.1016/j.cognition.2017.10.019

Botvinick, M. M. (2005). Effects of domain-specific knowledge on memory for serial order. *Cognition, 97*(2), 135-151. doi:10.1016/j.cognition.2004.09.007

Bower, G. H., & Glass, A. L. (1976). Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory, 2*(4), 456-466. doi:10.1037/0278-7393.2.4.456

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911.

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. doi:10.1016/j.jcomdis.2016.04.001

Campbell, R., MacSweeney, M., & Waters, D. (2008). Sign Language and the Brain: A Review. *Journal of Deaf Studies and Deaf Education, 13*(1), 3-20. doi:10.1093/deafed/enm035

Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M., & Emmorey, K. (2016). ASL-LEX: A lexical database of American Sign Language. *Behavior Research Methods*, 1-18. doi:10.3758/s13428-016-0742-0

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153-193. doi:10.1037/h0059973

Chee, M. W. L., Soon, C. S., Lee, H. L., & Pallier, C. (2004). Left insula activation: A marker for language attainment in bilinguals. *Proceedings of the National Academy of Sciences of the United States of America, 101*(42), 15265-15270. doi:10.1073/pnas.0403703101

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.. ed.). Hillsdale, N.J.: Hillsdale, N.J. : L. Erlbaum Associates.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*(1), 158-171. doi:10.3758/bf03193395

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cooper, T. C. (1987). Foreign Language Study and SAT-Verbal Scores. *The Modern Language Journal, 71*(4), 381-387. doi:10.2307/328467

Cooper, T. C., Daniel J., Y., Wisenbaker, J. M., Jahner, D., Webb, E., & Wilbur, M. L. (2008). Foreign Language Learning and SAT Verbal Scores Revisited. *Foreign Language Annals, 41*(2), 200-217. doi:doi:10.1111/j.1944-9720.2008.tb03289.x

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research, 169*, 323-338. doi:10.1016/S0079-6123(07)00020-9

Craik, F., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671-684. doi:10.1016/S0022-5371(72)80001-X

Dangiulli, A., Siegel, L. S., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics, 22*(4), 479-507. doi:10.1017/S0142716401004015

de Bruin, A. (2019). Not All Bilinguals Are the Same: A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behavioral Sciences, 9*(3), 33.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science (0956-7976), 26*(1), 99-107. doi:10.1177/0956797614557866

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition, 22*(4), 499-533.

Delcenserie, A., & Genesee, F. (2017). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism, 21*(5), 600-616. doi:10.1177/1367006916639158

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The Inhibitory Advantage in Bilingual Children Revisited. *Experimental Psychology, 61*(3), 234-251. doi:10.1027/1618-3169/a000243

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*(3), 462-474. doi:10.1037/0882-7974.15.3.462

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* . Princeton, NJ: Educational Testing Service.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word ‐ length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71*(1), 43-51. doi:10.1111/j.2044-8295.1980.tb02728.x

Embretson, S. E. (1992). Measuring and Validating Cognitive Modifiability as an Ability: A Study in the Spatial Domain. *Journal of Educational Measurement, 29*(1), 25-50.

Emmorey, K. (2002). *Language, Cognition, and the Brain : Insights From Sign Language Research*. Mahwah, N.J.: Psychology Press.

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: evidence from bimodal bilinguals. *Psychological Science, 19*(12), 1201-1206. doi:10.1111/j.1467-9280.2008.02224.x

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition, 43*(2), 226-236. doi:10.3758/s13421-014-0461-7

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*(1), 83-94. doi:10.3758/bf03210559

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543. doi:10.1017/S0142716406060383

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54A*(1), 1-30. doi:10.1080/02724980042000002

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81. doi:10.1017/S1366728998000133

Grosjean, F. (2003). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(2), 131-149. doi:10.1017/S136672899800025X

Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research, 33*(3), 325-340. doi:10.1177/0267658316678286

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 56A*(7), 1213-1236. doi:10.1080/02724980343000071

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61*(4), 481-502. doi:10.1016/j.jml.2009.08.001

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. doi:10.3758/s13423-011-0116-7

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review, 25*(3), 892-916. doi:10.3758/s13423-017-1315-7

Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software, 42*(8), 1-28. doi:http://dx.doi.org/10.18637/jss.v042.i08

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for Familiar and Unfamiliar Words: Evidence for a Long-Term Memory Contribution to Short-Term Memory Span. *Journal of Memory and Language, 30*(6), 685-701.

Hummel, K. M., & French, L. M. (2016). Phonological memory and aptitude components: Contributions to second language proficiency. *Learning and Individual Differences, 51*(Supplement C), 249-255. doi:https://doi.org/10.1016/j.lindif.2016.08.016

Jackson, D. N. (1998). *Multidimensional Aptitude Battery-II*. London, Ontario, Canada: Sigma Assessment Systems.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99. doi:https://doi.org/10.1016/0010-0285(89)90003-0

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217. doi:10.1037/0096-3445.133.2.189

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. doi:10.1017/S1366728911000472

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2011). The relationship between vocabulary and short-term memory measures in monolingual and bilingual speakers. *The international journal of bilingualism : cross-disciplinary, cross-linguistic studies of language behavior, 15*(4), 408-425. doi:10.1177/1367006911403201

Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science, 17*(4), 564-583. doi:10.1111/desc.12142

Kaushanskaya, M., & Marian, V. (2009a). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705-710. doi:10.3758/pbr.16.4.705

Kaushanskaya, M., & Marian, V. (2009b). Bilingualism reduces native-language interference during novel-word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 829-835. doi:10.1037/a0015275

Kaushanskaya, M., & Prior, A. (2014). Variability in the effects of bilingualism on cognition: It is not just about cognition, it is also about bilingualism. *Bilingualism: Language and Cognition, 18*(1), 27-28. doi:10.1017/S1366728914000510

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review, 19*(5), 935-941. doi:10.3758/s13423-012-0271-5

Kaushanskaya, M., Yoo, J., Van Hecke, S., & Oetting, J. (2013). Word Learning in Adults With Second-Language Experience: Effects of Phonological and Referent Familiarity. *Journal of Speech, Language & Hearing Research, 56*(2), 667-678. doi:10.1044/1092-4388(2012/11-0084)

Kee, D. W. (1994). Developmental differences in associative memory: Strategy use, mental effort, and knowledge access interactions. In H. W. Reese & H. W. Reese (Eds.), *Advances in child development and behavior, Vol. 25.* (pp. 7-32). San Diego, CA, US: Academic Press.

Klein, D., Watkins, K. E., Zatorre, R. J., & Milner, B. (2006). Word and nonword repetition in bilingual subjects: A PET study. *Human Brain Mapping, 27*(2), 153-161. doi:10.1002/hbm.20174

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381. doi:10.1017/S136672891000009X

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. *Intelligence, 12*(4), 393-421. doi: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. doi:10.1037/0096-3445.120.1.57

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. doi:10.1037/bul0000142
10.1037/bul0000142.supp (Supplemental)

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., . . . Doughty, C. J. (2013). Hi-LAB: A New Measure of Aptitude for High-Level Language Proficiency. *Language Learning, 63*(3), 530-566. doi:10.1111/lang.12011

Little, R. J. A., & Rubin, D. B. (2014). *Statistical analysis with missing data* (2nd ed.. ed.): Hoboken, New Jersey : John Wiley & Sons, Inc.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis, P. Tapsfield, I. Dennis, & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement.* (pp. 97-116). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280.

Majerus, S. (2008). 12 Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. *Interactions between short-term and long-term memory in the verbal domain*, 244.

Majerus, S. (2013). *Language repetition and short-term memory: An integrative framework* (Vol. 7).

Majerus, S., Belayachi, S., De Smedt, B., Leclercq, A. L., Martinez, T., Schmidt, C., . . . Maquet, P. (2008). Neural networks for short-term memory for order differentiate high and low proficiency bilinguals. *NeuroImage, 42*(4), 1698-1713. doi:https://doi.org/10.1016/j.neuroimage.2008.06.003

Majerus, S., Martinez Perez, T., & Oberauer, K. (2012). Two distinct origins of long-term learning effects in verbal short-term memory. *Journal of Memory and Language, 66*(1), 38-51. doi:10.1016/j.jml.2011.07.006

Majerus, S., Poncelet, M., Elsen, B., & van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology, 18*(6), 848-873. doi:10.1080/09541440500446476

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. doi:https://doi.org/10.1016/j.cognition.2007.10.003

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*(2), 107-127. doi:10.1016/0160-2896(83)90023-5

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in second language acquisition, 34*(3), 379-413. doi:10.1017/S0272263112000125

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language, 43*(2), 240-274. doi:10.1016/0093-934X(92)90130-7

Martinez, D., & Singleton, J. L. (2018a). *Individual differences in lexical learning across two language modalities: sign learning, word learning, and their relationship in hearing non-signers*. Manuscript submitted for publication.

Martinez, D., & Singleton, J. L. (2018b). Predicting sign learning in hearing adults: The role of perceptual-motor (and phonological?) processes. *Applied Psycholinguistics*, 1-27. doi:10.1017/S0142716418000048

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. doi:10.1037/a0017619

Morgan, C. (1993). Attitude Change and Foreign Language Culture Learning. *Language Teaching: The International Abstracting Journal for Language Teachers and Applied Linguists, 26*(2), 63-75.

Nair, V. K. K., Biedermann, B., & Nickels, L. (2017). Understanding Bilingual Word Learning: The Role of Phonotactic Probability and Phonological Neighborhood Density. *Journal of Speech, Language & Hearing Research, 60*(12), 3551-3560. doi:10.1044/2017_JSLHR-L-15-0376

Nairne, J. S. (1983). Associative processing during rote rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 3-20. doi:10.1037/0278-7393.9.1.3

Nelson, B. (1968). Teaching Culture in the Foreign Language Classroom. *Foreign Language Annals, 1*(3), 204-217. doi:doi:10.1111/j.1944-9720.1968.tb00135.x

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*(04), 661-677.

Newman-Norlund, R. D., Frey, S. H., Petitto, L.-A., & Grafton, S. T. (2006). Anatomical Substrates of Visual and Auditory Miniature Second-language Learning. *Journal Of Cognitive Neuroscience, 18*(12), 1984-1997.

Norris, D., Page, M. P. A., & Hall, J. (2018). Learning nonwords: the Hebb repetition effect as a model of word learning. *Memory, 26*(6), 852-857. doi:10.1080/09658211.2017.1416639

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(3), 377-402. doi:10.1017/S0142716406060322

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142*(7), 758-799. doi:10.1037/bul0000046

Ohala, J. J. (1993). Coarticulation and Phonology. *Language and Speech, 36*(2-3), 155-170. doi:10.1177/002383099303600303

Ordonez Magro, L., Attout, L., Majerus, S., & Szmalec, A. (2018). Short-and long-term memory determinants of novel word form learning. *Cognitive Development, 47*, 146-157. doi:https://doi.org/10.1016/j.cogdev.2018.06.002

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. doi: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, 69*, 265-278. doi:https://doi.org/10.1016/j.cortex.2015.04.014

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*. doi:10.3389/fpsyg.2014.00962

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions Of The Royal Society Of London. Series B, Biological Sciences, 364*(1536), 3737-3753. doi:10.1098/rstb.2009.0173

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*(3), 331-347. doi:10.1016/0749-596X(91)90040-Q

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 44A*(1), 47-67. doi:10.1080/14640749208401283

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. doi:10.1080/14640749508401378

Pierce, L. J., Genesee, F., Delcenserie, A., & Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics, 38*(6), 1265-1300. doi:10.1017/S0142716417000236

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*(1), 1-11. doi:10.1080/20445911.2014.976226

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Advanced Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Rönnberg, J., Rudner, M., & Ingvar, M. (2004). Neural correlates of working memory for sign language. *Cognitive Brain Research, 20*(2), 165-182. doi:10.1016/j.cogbrainres.2004.03.002

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika, 70*(1), 41-55. doi:10.2307/2335942

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63*(3), 581-592. doi:10.2307/2335739

Rudner, M., Andin, J., & Rönnberg, J. (2009). Working memory, deafness and sign language. *Scandinavian Journal of Psychology, 50*(5), 495-505. doi:10.1111/j.1467-9450.2009.00744.x

Rumbaut, R. G. (2014). English plus: Exploring the socioeconomic benefits of bilingualism in southern Californiat.

Saiz, A., & Zoido, E. (2005). Listening to What the World Says: Bilingualism and Earnings in the United States. *The Review of Economics and Statistics, 87*(3), 523-538.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*(2), 168-175. doi:10.3758/bf03202729

Shastry, G. K. (2012). Human Capital Response to Globalization: Education and Information Technology in India. *Journal of Human Resources, 47*(2), 287-330. doi:10.1353/jhr.2012.0016

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20*(2), 216-230. doi:10.1016/S0022-5371(81)90389-3

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal Of Speech, Language, And Hearing Research: JSLHR, 49*(6), 1175-1192.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-Term Knowledge Effects on Serial Recall of Nonwords Are Not Exclusively Lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 729-735. doi:10.1037/0278-7393.31.4.729

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific Knowledge and Short-term Memory in Bilingual and Non-bilingual Children. *The Quarterly Journal of Experimental Psychology Section A, 52*(2), 303-324. doi:10.1080/713755823

Thurstone, L. L. (1938). *Primary Mental Abilities*: Psychometric Monographs.

van Hell, J. G., & Candia Mahn, A. (1997). Keyword Mnemonics Versus Rote Rehearsal: Learning Concrete and Abstract Foreign Words by Experienced and Inexperienced Learners. *Language Learning, 47*(3), 507-546. doi:doi:10.1111/0023-8333.00018

Varriale, V., van der Molen, M. W., & De Pascalis, V. (2018). Mental rotation and fluid intelligence: A brain potential analysis. *Intelligence, 69*, 146-157. doi:https://doi.org/10.1016/j.intell.2018.05.007

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.* Paper presented at the 58th Annual Meeting of Psychonomic Society, Vancouver.

Williams, B. A., & Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with Raven scores, but not with measures of cognitive processing. *Intelligence, 34*(2), 177-191. doi:10.1016/j.intell.2005.03.007

Williams, J. T., Darcy, I., & Newman, S. D. (2015). Modality-independent neural mechanisms for novel phonetic processing. *Brain Research, 1620*, 107-115. doi:10.1016/j.brainres.2015.05.014

Wilson, M., & Emmorey, K. (1998). A 'word length effect' for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition, 26*(3), 584-590. doi:10.3758/BF03201164

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language, 58*(1), 48-65. doi:10.1016/j.jml.2007.07.002

Yoo, J., & Kaushanskaya, M. (2012). Phonological memory in bilinguals and monolinguals. *Memory & Cognition, 40*(8), 1314-1330. doi:10.3758/s13421-012-0237-x

Appendix A: Word learning pseudowords and their characteristics

Item characteristics were retrieved from http://elexicon.wustl.edu/default.asp

Task	Word	Length	Average Bigram Frequency	Mean Lexical Decision Latency (ms)	Mean Lexical Decision Latency Z-Score
PSL	attears	7	2,582.83	821.767	-0.232
	brogine	7	2,162.00	716.1	0.14
	candage	7	1,596.17	904.467	-0.389
	dob	3	473.5	844.7	-0.207
	eto	3	1,360.00	784.933	0.063
	haj	3	581.5	669.967	0.399
	illet	5	1,958.50	808.8	0.09
	jeeds	5	1,479.25	715.933	0.236
	lestroy	7	2,573.33	727.5	0.124
	mests	5	2,700.25	795.067	-0.349
	nased	5	2,348.75	846.1	-0.197
	valif	5	1,698.00	769.367	0.129
	varbine	7	2,248.33	732.933	0.133
	vig	3	797.5	686.467	0.209
	wid	3	657	769.633	-0.077
3TWL	curlang	7	1,952.50	779.033	-0.113
	forim	5	1,776.00	712.7	0.182
	gallade	7	1,827.67	914.833	-0.306
	hig	3	952	861.467	-0.306
	inlut	5	2,139.25	735.8	0.055
	lerge	5	2,773.75	689.533	0.256
	lut	3	761.5	773.867	0.002
	maz	3	892	759.267	0.095
	pellist	7	2,347.67	840.3	-0.11
	weith	5	1,073.50	820.367	-0.115
	Study Average	5	1668.51	779.236	-.0115
	Corpus Average[a]	8	1974.56	784.07	.00

[a]Average characteristics for words in the English Lexicon Project (Table 2; Balota et al., 2007).

Appendix B: Spoken phonological STM pseudowords and their characteristics

Note: All pseudowords were drawn from Gathercole et al. (2001). Item characteristics, when available, were retrieved from http://elexicon.wustl.edu/default.asp

Table B1. Pseudowords used in NWRec task

Word	Length	Average Bigram Frequency	Mean Lexical Decision Latency (ms)	Mean Lexical Decision Latency Z-Score
booge	---	---	---	---
borge	---	---	---	---
choom	---	---	---	---
chull[a]	5	1,114.75	741.545	0.081
dern	---	---	---	---
dorch	---	---	---	---
gan	3	2,048.50	766.407	0.033
ged	3	2,848.00	696.759	0.185
geed	4	2,188.67	742.367	0.017
jup	---	---	---	---
keech	---	---	---	---
leck	4	1,886.67	804.967	-0.053
lidge	---	---	---	---
lig	3	1,783.00	748.333	0.194
mahn	---	---	---	---
maht	---	---	---	---
marn[a]	4	1,706.00	806.179	0.099
mern	---	---	---	---
nahg	---	---	---	---
nart	---	---	---	---
newel	---	---	---	---
parn	4	1,512.00	837.452	-0.074
ped	3	3,074.00	869.826	-0.1
putch	---	---	---	---
tard	4	1,823.33	882.211	-0.288
tep	---	---	---	---
torm	4	1,567.67	843.852	-0.047
tudge	---	---	---	---

[a]These pseudowords were also used in the NWSpan task.

Table B2. Pseudowords used in NWSpan task

Word	Length	Average Bigram Frequency	Mean Lexical Decision Latency (ms)	Mean Lexical Decision Latency Z-Score
barp	---	---	---	---
cham	4	1,202.33	960.7	-0.535
chull[a]	5	1,114.75	741.545	0.081
derb	---	---	---	---
dordge	---	---	---	---
gerk	---	---	---	---
jerg	---	---	---	---
marn[a]	4	1,706.00	806.179	0.099
narch	---	---	---	---
narg	---	---	---	---
padge	---	---	---	---
terge	---	---	---	---

[a]These words were also used in the NWRec task.

Appendix C: Means, dispersion statistics, and internal consistency coefficients

Task	Mean (SD)	Skew	Kurtosis	α^a
ASL-SL	.70 (.23)	-.93	.11	.93
PSL	.70 (.20)	-.70	-.17	.87
DPSL	.69 (.23)	-.48	.02	.82
3TSL	.63 (.27)	-.54	-.82	.95
TWL	.54 (.24)	-.22	-.88	.91
PWL	.62 (.25)	-.43	-1.04	.91
DPWL	.59 (.25)	-.34	-.49	.83
3TWL	.48 (.33)	.08	-1.55	.98
NSPT	.79 (.07)	-.51	.25	.66
ProSign	.66 (.11)	-.14	-.42	.57
SignCon	.60 (.17)	-.72	.48	.75
LetSpan	.81 (.10)	-.68	.48	.78
NWRec	.80 (.12)	-.60	-.04	.75
NWSpan	.66 (.12)	-.09	-.04	.76
OSpan	.81 (.19)	-1.25	1.45	.74
SymSpan	.73 (.23)	-.95	.69	.67
RoSpan	.61 (.23)	-.73	.11	.64
Info	.58 (.14)	-.70	.57	.83
Vocab	.52 (.15)	.01	-.32	.84
Gram	.44 (.21)	.13	-.60	.78
Reading	.49 (.22)	.09	-.61	.79
NumSeries	.66 (.20)	-.42	-.66	.77
LetSets	.56 (.16)	-.29	-.52	.86
Ravens	.56 (.22)	-.35	-.54	.80
SLAT	.52 (.25)	.30	-1.03	.86

[a]Coefficient alphas were taken from the main study, N = 236. All other values were calculated using data from the present study (N = 190).

Appendix D: A Tally of the Languages Spoken by Bilingual Participants as Their First, Second, and Dominant Language

	L1	L2	Dominant
Arabic	1		
Armenian	1	1	1
Bengali		1	
Bulgarian	1	1	
Chinese	5	7	
English	38	13	63
French		7	1
German		1	
Gokana		1	
Gujarati	1	1	1
Haitian Creole	1	1	
Hindi	1	4	1
Igbo		1	
Japanese		1	
Korean	1	2	
Krio		1	
Marathi			1
Nepali	1		1
Romanian	1	1	
Slovak	1		
Spanish	8	21	1
Tamil	3	5	1
Telugu	5	2	3
Urdu	2	3	1
Vietnamese	2	2	
Yoruba		1	

Note: N = 78; additionally, five participants reported two language as their L1 and three participants reported two languages as their dominant language, always English and another language.

Table 1. Task administration order

	Session 1	**Session 2**
1	ASL-SL	Reading
2	PSL	Info
3	3TSL	Vocab
4	LetSpan	Gram
5	NWRec	OSpan
6	NWSpan	SymSpan
7	DPSL	RoSpan
	[Optional 5min Break]	[Optional 5min Break]
8	TWL	Ravens
9	PWL	LetSets
10	3TWL	NumSeries
11	NSPT	SLAT
12	ProSign	OSIQ
13	SignCon	Questionnaire
14	DPWL	

Note: ASL-SL = ASL Sign Learning; PSL = pseudosign learning; 3TSL = three-term sign learning; LetSpan = letter span; NWRec = Nonword Recognition; NWSpan = Nonword Span; DPSL = delayed pseudosign learning; TWL = Turkish word learning; PWL = pseudoword learning; 3TWL = three-term word learning; NSPT = nonsign paired task; ProSign = Probed Sign recognition task; SignCon = sign configuration task; DPWL = delayed word learning; Reading = test of reading comprehension; Info = information test; Vocab =extended range vocabulary test; Gram = grammar and usage test; Ravens = Raven's Advanced Progressive Matrices, Set II; LetSets = letter sets; NumSeries = number series; SLAT = spatial learning ability test; OSpan = operation span; SymSpan = symmetry span; RoSpan = rotation span; OSIQ = object-spatial imagery questionnaire; Questionnaire = language experience and demographic questionnaire.

Table 2: Demographic, Language Experience, and Covariates—Full Data (Monolinguals = 112, Bilinguals = 78)

Variable		Mean (*SD*) Monolinguals	Bilinguals	*t*-value	*p*-value	Cohen's *d*
Age		21.54 (4.09)	21.05 (3.01)	-.954[c]	.341	.136
Sex		.57 (.50)[a]	.72 (.45)[a]	2.107[d]	.037	.315
Education		3.10 (.57)	3.19 (.77)	.926[e]	.356	.133
Income		4.30 (2.75)	4.45 (2.60)	.381	.704	.056
AoA		14.27 (2.63)[b]	3.97 (3.77)	-16.578[f]	<.001	3.169
L2 Proficiency						
	Speaking	.97 (1.72)[b]	7.47 (1.86)	24.834	<.001	3.629
	Listening	.93 (1.63)[b]	7.88 (1.73)	28.342	<.001	4.135
	Reading	1.25 (2.17)[b]	6.53 (2.56)	15.325	<.001	2.225
	Writing	1.05 (1.83)[b]	5.87 (2.91)	12.893[h]	<.001	1.983
Gc						
	Info	.58 (.14)	.59 (.12)	.450	.653	.077
	Vocab	.52 (.15)	.52 (.15)	.010	.992	.001
	Gram	.42 (.19)	.47 (.24)	1.728	.086	.231
	Reading	.48 (.23)	.49 (.22)	.319	.750	.044
Gf						
	NumSeries	.65 (.21)	.67 (.19)	.841	.402	.100
	LetSets	.54 (.16)	.58 (.17)	1.712	.088	.242
	Ravens	.54 (.22)	.58 (.20)	1.306	.193	.190
	SLAT	.50 (.26)	.53 (.23)	.871	.385	.122

[a] = represents proportion of women.
[b] = The majority of monolinguals (78 out of 112) claimed no fluency in an L2—estimates were calculated by leaving their AoA cells blank and entering zeroes for their fluency scores—though it should be noted that in the United States, learning a foreign language is often compulsory in secondary and post-secondary school.
[c,d,e,f,g,h] = Levene's test was significant indicating that equal variance cannot be assumed; the degrees of freedom for these tests were 186.089, 174.966, 132.808, 88.145, and 116.861, respectively.

Table 3: Composite and Individual Test Results—Full Data, Monolinguals = 112, Bilinguals = 78

Variable		Mean (*SD*) Monolinguals	Bilinguals	t-value	*p*-value	Cohen's *d*
Sign Learning		**-0.23 (3.54)**	**0.33 (3.26)**	**1.114**	**0.267**	**0.166**
	ASL	0.69 (.23)	0.71 (0.22)	0.701	0.484	0.089
	PSL	0.69 (.21)	0.71 (0.20)	0.764	0.446	0.098
	DPSL	0.68 (.23)	0.71 (0.22)	0.806	0.421	0.133
	3TSL	0.60 (.28)	0.67 (0.25)	1.55	0.122	0.264
Word Learning		**-0.36 (3.65)**	**0.51 (3.42)**	**1.649**	**0.101**	**0.245**
	TWL	0.51 (.27)	0.57 (0.21)	1.755[a]	0.081	0.248
	PWL	0.60 (.27)	0.65 (0.23)	1.436[b]	0.153	0.199
	DPWL	0.57 (.24)	0.61 (0.26)	1.245	0.215	0.160
	3TWL	0.45 (.33)	0.53 (0.32)	1.553	0.122	0.246
Signed phonological STM		**-0.24 (2.67)**	**0.34 (2.35)**	**1.532**	**0.127**	**0.228**
	NSPT	0.79 (.08)	0.80 (0.07)	1.700	0.091	0.133
	ProSign	0.65 (.11)	0.67 (0.10)	1.110	0.268	0.190
	SignCon	0.59 (.17)	0.62 (0.16)	1.099	0.273	0.181
Spoken phonological STM		**-0.51 (2.60)**	**0.73 (2.30)**	**3.383**	**0.001**	**0.504**
	LetSpan	0.79 (.10)	0.83 (0.08)	3.528[c]	0.001	0.442
	NWRec	0.78 (.12)	0.82 (0.11)	2.480	0.014	0.347
	NWSpan	0.64 (.11)	0.68 (0.11)	2.718	0.007	0.364
WMC		**-0.22 (2.56)**	**0.32 (2.15)**	**1.536**	**0.126**	**0.230**
	OSpan	0.79 (.20)	0.83 (0.17)	1.728	0.086	0.216
	SymSpan	0.70 (.25)	0.76 (0.21)	1.697	0.091	0.260
	RoSpan	0.61 (.24)	0.62 (0.21)	0.278	0.781	0.044

[a,b,c] = Levene's test was significant indicating that equal variance cannot be assumed; the degrees of freedom for these tasks were 184.715, 177.583, and 186.481, respectively; all other degrees of freedom were equal to 188.

Table 4: Matched Data, Monolinguals = 65, Bilinguals = 65

		Mean (*SD*)		*t*-value	*p*-value	Cohen's *d*
		Monolinguals	Bilinguals			
Age		20.68 (3.42)	21.03 (3.05)	0.623	0.534	0.108
Sex[a]		1.66 (0.48)	1.71 (0.46)	0.563	0.575	0.106
Education		3.17 (0.55)	3.18 (0.64)	0.148	0.883	0.017
Income		5.22 (2.38)	4.88 (2.32)	-0.822	0.413	-0.145
AoA		13.94 (2.63)	4.05 (3.86)	-12.384[c]	<.001	-3.048
L2 Proficiency	Speaking	0.95 (1.79)	7.62 (1.78)	21.265	<.001	3.737
	Listening	0.89 (1.68)	7.97 (1.76)	23.471	<.001	4.116
	Reading	1.28 (2.34)	6.63 (2.54)	12.492	<.001	2.193
	Writing	1.06 (1.94)	6.03 (2.75)	11.824[d]	<.001	2.119
Gc	Info	0.59 (0.13)	0.59 (0.12)	0.295	0.768	0.051
	Vocab	0.53 (0.13)	0.52 (0.14)	-0.111	0.912	-0.074
	Gram	0.43 (0.17)	0.47 (0.23)	1.127	0.262	0.200
	Reading	0.50 (0.22)	0.48 (0.22)	-0.542	0.589	-0.091
Gf	NumSeries	0.68 (0.20)	0.68 (0.21)	0.021	0.983	0.004
	LetSets	0.55 (0.17)	0.58 (0.17)	0.987	0.325	0.176
	Ravens	0.58 (0.21)	0.58 (0.21)	0.046	0.964	0.008
	SLAT	0.54 (0.26)	0.54 (0.23)	0.100	0.920	-0.018

[a] = men were coded as 0, women as 1.
[b] = The majority of monolinguals (20 out of 65) claimed no fluency in an L2—estimates were calculated by leaving their AoA cells blank and entering zeroes for their fluency scores.
[c,d] = Levene's test was significant indicating that equal variance cannot be assumed; the degrees of freedom for these tasks were 36.213 and 113.145, respectively.

Table 5: Matched Data, Monolinguals = 65, Bilinguals = 65

		Mean (*SD*)		*t*-value	*p*-value	Cohen's *d*
		Monolinguals	Bilinguals			
Sign Learning		**0.14 (3.31)**	**0.22 (3.40)**	**0.150**	**0.881**	**0.026**
	ASL-SL	0.71 (0.22)	0.70 (0.23)	-0.184	0.854	-0.044
	PSL	0.71 (0.19)	0.72 (0.20)	0.315	0.753	0.051
	DPSL	0.69 (0.23)	0.70 (0.23)	0.320	0.750	0.043
	3TSL	0.65 (0.28)	0.65 (0.26)	0.067	0.947	0.012
Word Learning		**0.05 (3.35)**	**0.46 (3.43)**	**0.680**	**0.498**	**0.119**
	TWL	0.54 (0.26)	0.57 (0.21)	0.666	0.507	0.128
	PWL	0.63 (0.27)	0.65 (0.23)	0.629	0.530	0.080
	DPWL	0.59 (0.24)	0.61 (0.26)	0.408	0.684	0.080
	3TWL	0.48 (0.33)	0.52 (0.32)	0.708	0.480	0.123
Signed phonological STM		**0.11 (2.43)**	**0.29 (2.50)**	**0.425**	**0.671**	**0.075**
	NSPT	0.80 (0.07)	0.80 (0.07)	0.311	0.757	0.054
	ProSign	0.66 (0.11)	0.66 (0.11)	0.452	0.652	0.095
	SignCon	0.61 (0.16)	0.62 (0.17)	0.302	0.763	0.061
Spoken phonological STM		**-0.14 (2.38)**	**0.72 (2.39)**	**2.072**	**0.040**	**0.363**
	LetSpan	0.80 (0.10)	0.83 (0.08)	2.267	0.025	0.333
	NWRec	0.80 (0.12)	0.82 (0.11)	1.086	0.280	0.174
	NWSpan	0.65 (0.10)	0.68 (0.13)	1.895	0.060	0.261
WMC		**0.11 (2.17)**	**0.21 (2.29)**	**0.251**	**0.802**	**0.044**
	OSpan	0.81 (0.16)	0.83 (0.17)	0.728	0.468	0.121
	SymSpan	0.73 (0.22)	0.75 (0.22)	0.603	0.548	0.091
	RoSpan	0.63 (0.23)	0.60 (0.22)	-0.682	0.496	-0.133

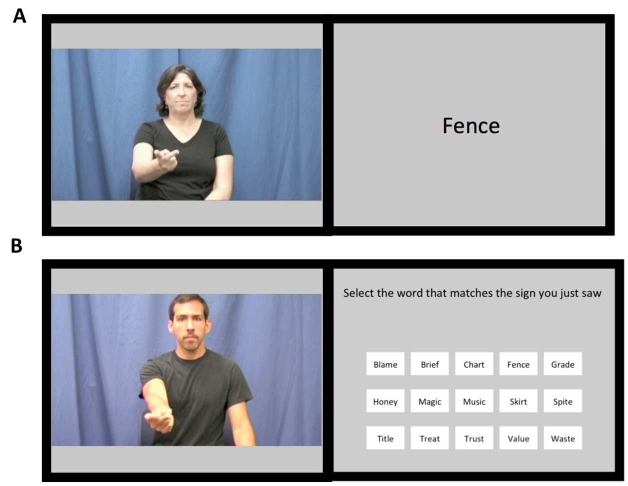

Figure 1. Depiction of the PSL task.

Figure 2. Depiction of the NSPT.

Figure 3. Depiction of a ProSign item.

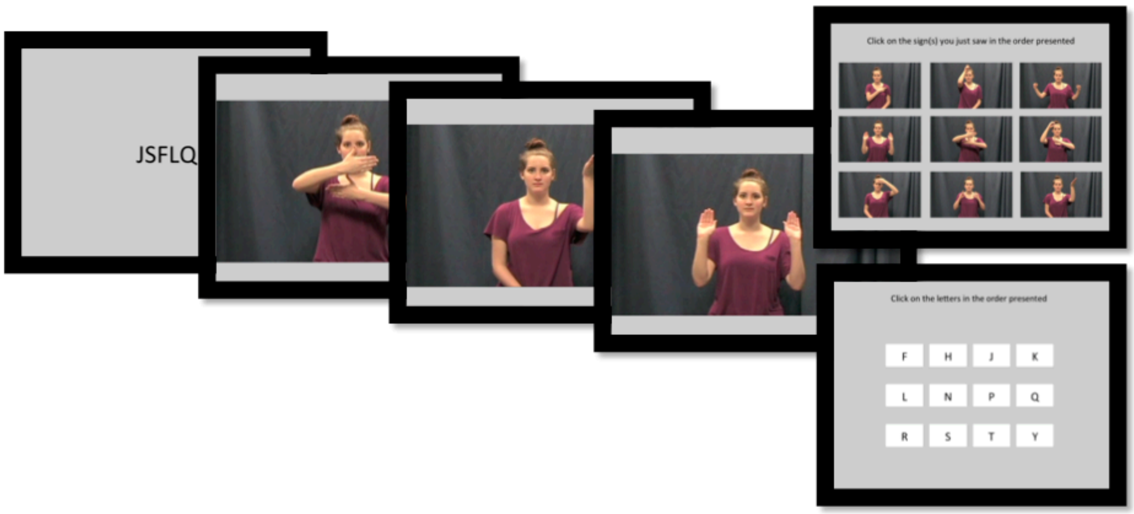

Figure 4. Depiction of the SignCon.

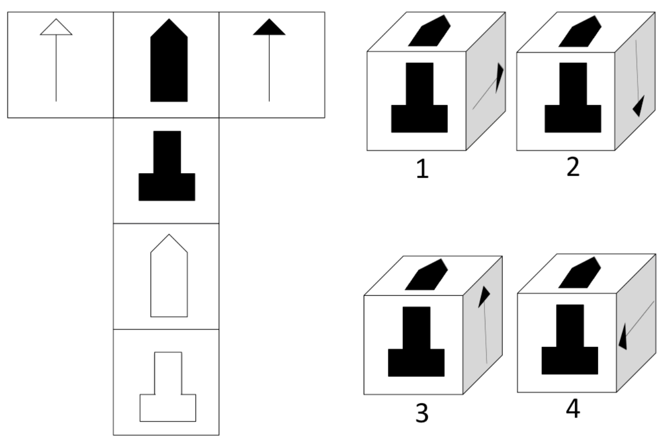

Figure 5. Depiction of a SLAT item.

Figure 1. Depiction of the PSL task. Panel A shows a study trial. Panel B shows a test trial.

Figure 2. Depiction of the NSPT. In the NSPT, there were 28 target signs each with two reproductions, produced by different individuals. In this illustration, both Reproductions should obtain a "SAME" response.

Figure 3. Depiction of a ProSign item. The probe differs from the second pseudosign in the set as it uses a different handshape, but keeps the same movement and orientation.

Figure 4. Depiction of the SignCon. Participants were either tested on the number of pseudosigns (60% of trials) or letters (40% of trials) they could recall in order.

Figure 5. Depiction of a SLAT item.